

Cross-Task Cognitive Workload Recognition Based on EEG and Domain Adaptation

Yueying Zhou¹, Ziming Xu¹, Yifan Niu, Pengpai Wang¹, Xuyun Wen, Xia Wu¹, *Senior Member, IEEE*, and Daoqiang Zhang¹, *Senior Member, IEEE*

Abstract—Cognitive workload recognition is pivotal to maintain the operator’s health and prevent accidents in the human-robot interaction condition. So far, the focus of workload research is mostly restricted to a single task, yet cross-task cognitive workload recognition has remained a challenge. Furthermore, when extending to a new workload condition, the discrepancy of electroencephalogram (EEG) signals across various cognitive tasks limits the generalization of the existed model. To tackle this problem, we propose to construct the EEG-based cross-task cognitive workload recognition models using domain adaptation methods in a leave-one-task-out cross-validation setting, where we view any task of each subject as a domain. Specifically, we first design a fine-grained workload paradigm including working memory and mathematic addition tasks. Then, we explore four domain adaptation methods to bridge the discrepancy between the two different tasks. Finally, based on the supporting vector machine classifier, we conduct experiments to classify the low and high workload levels on a private EEG dataset. Experimental results demonstrate that our proposed task transfer framework outperforms the non-transfer classifier with improvements of 3% to 8% in terms of mean accuracy, and the transfer joint matching (TJM) consistently achieves the best performance.

Index Terms—Cognitive workload, cross-task recognition, electroencephalogram (EEG), domain adaptation.

I. INTRODUCTION

CURRENTLY, the cognitive workload of the operator has been studied widely in the fields of human-robot interac-

Manuscript received August 4, 2021; revised November 14, 2021 and December 16, 2021; accepted January 2, 2022. Date of publication January 5, 2022; date of current version January 28, 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 62136004, Grant 61876082, and Grant 61732006; in part by the National Key Research and Development Program of China under Grant 2018YFC2001600 and Grant 2018YFC2001602; and in part by the Chinese Association for Artificial Intelligence (CAAI)-Huawei MindSpore Open Fund. (Corresponding author: Daoqiang Zhang.)

Yueying Zhou, Ziming Xu, Pengpai Wang, and Xuyun Wen are with the MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China (e-mail: zhouyueying@nuaa.edu.cn; xuziming@nuaa.edu.cn; pengpaiwang@nuaa.edu.cn; wenxuyun@nuaa.edu.cn).

Yifan Niu and Xia Wu are with the School of Artificial Intelligence, Beijing Normal University, Engineering Research Center of Intelligent Technology and Educational Application, Ministry of Education, Beijing 100875, China (e-mail: niuyifan@mail.bnu.edu.cn; wuxia@bnu.edu.cn).

Daoqiang Zhang is with the MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China, and also with the Engineering Research Center of Traditional Chinese Medicine Intelligent Rehabilitation, Ministry of Education, Shanghai 201101, China (e-mail: dqzhang@nuaa.edu.cn).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TNSRE.2022.3140456>, provided by the authors.

Digital Object Identifier 10.1109/TNSRE.2022.3140456

tion environment and passive brain-computer interface [1], [2]. It is a special case of cognitive states, described as the ratio of the operator’s available cognitive resources (e.g., the attention resources and working memory capacity) over the task demanded resources [3]. Due to the limited cognitive resources of the brain, the heavy cognitive work in real-world environments will lead to cognitive overload, further affect task execution and harm the operator’s state [4]. As such, it is important to accurately recognize human cognitive workload to prevent accidents and maintain health.

To date, we can categorize the measurements for cognitive workload into subjective and objective measures [5]. Subjective measurements are mainly based on the operator’s perceived feeling and rating scales, e.g., the National Aeronautics and Space Administration-Task Load Index (NASA-TLX) [6]. Though these measurements implement easily, they are post-evaluation and cannot yield objective or real-time results [7]. Meanwhile, objective measurements are mainly relied on the recorded physiological signals during the task process, hence having less interference on the task [5]. Among various physiological signals, the electroencephalogram (EEG) signal has been widely used and studied, due to its high-temporal-resolution, security, and convenience. Besides, its effectiveness has been validated in detecting cognitive workload during the execution of workload-related cognitive tasks [8]. Hence, we concentrate on EEG-based cognitive workload recognition.

The EEG signals may have variabilities among different subjects and/or tasks, mainly including intra- and inter-subject, and inter-task variabilities. These variabilities are related to constructing subject-dependent, cross-subject (or subject-independent), and cross-task (or task-independent) models, respectively [7]. The subject-dependent model is trained and tested on data specific to each subject. Here, we can view the subject-dependent study as the standard cognitive workload recognition design [9] considering the large variability between subjects. Hitherto, many subject-dependent methods have been constructed and achieved acceptable recognition performance [10], [11]. Typically, the inter-subject and inter-task variabilities are more challenging and complex and would have inferior performance to the former. But if successful, they would enable and improve cognitive state monitoring and probing in real-world environments [7], [9]. So it is crucial to alleviate the cross-subject and cross-task issues. The cross-subject model is trained on data from one or a group of subjects and tested with data from a new subject. On the cross-subject research, based on the development of EEG optimal features extraction and classification models, acceptable results are obtained [12], [13].

The cross-task model is trained on one task and tested on another similar but different task. Ideally, although different cognitive tasks may elicit different cognitive resources, the cognitive workload pays more attention to the occupied amount of cognitive resources rather than the specific cognitive resources. Therefore, for a practical design, it is possible to construct a generalizable or cross-task cognitive workload recognition model that is capable to recognize workload across various tasks [14]. However, we find the cross-task cognitive workload recognition has degraded performance than the subject-dependent model [14]. During different task-elicited cognitive activities, the mismatched workload between the training and testing data [15], the highly dissimilar EEG patterns [16], the non-stationary characteristics of EEG data [15], [17], further the distribution variabilities between the EEG data, may cause considerable difficulties for cross-task cognitive workload recognition. Existing studies assumed a set of invariant features exists across tasks and proposed to find the common features and then constructed the task-independent model [8], [14]–[16], which might ignore the discrepancy between different tasks and thus have limited performance. To deal with the above problems, we propose to apply unsupervised domain adaptation for establishing the cross-task cognitive workload recognition model, aiming to reduce the distribution discrepancy as well as to improve the generalized classification accuracy across various tasks. Given efficient labeled source samples and unlabeled target samples, unsupervised domain adaptation transfers knowledge from the source domain to the target domain and tries to train a classifier that works well on a target domain. It aims to reduce the distribution discrepancy between the source and target data, thus making them similar [18], [19]. To our knowledge, domain adaptation has not been or rarely been applied for cross-task cognitive workload recognition systems.

In this paper, we propose a new framework for EEG-based cross-task cognitive workload recognition using domain adaptation. The proposed framework is implemented under three transfer schemes, which are the same/various one-to-one cross-task transfers, and the many-to-one cross-task transfer. We mainly explore four domain adaptation methods as a preliminary study for the new framework. These domain adaptation methods both assume the shared feature representations are existing between the source domain and the target domain simultaneously reducing the distribution gap. The difference in them lies in the ideas on the marginal distribution and conditional distribution between domains. Then, we compare the performance of these methods on a private EEG dataset with two different tasks to construct the workload recognition model in a binary classification way. Assuming the two tasks involved in the cross-task study should share have some brain mechanisms in common but also present distinctive activations, as suggested and applied in [8], [20], [21], we thus use the Sternberg Working Memory task (denoted as WM task) and Mathematics Addition task (denoted as MA task) to elicit cognitive workload states.

In [Figure 1](#), we display the general framework of cross-task cognitive workload recognition, including the cross-task design, EEG data acquisition, data preprocessing, feature

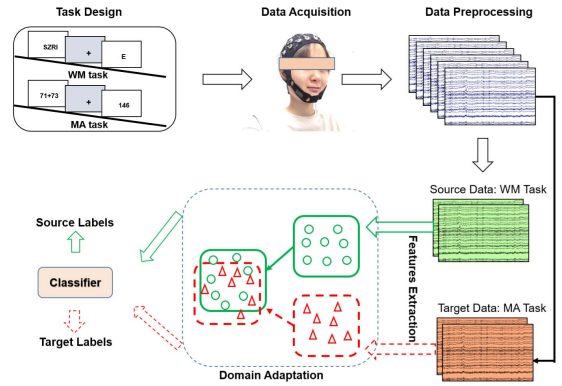


Fig. 1. The proposed framework of cross-task cognitive workload recognition using domain adaptation.

extraction, domain adaptation, and classification. Here, we take task A (e.g., the WM task) as source data to train the models, and task B (e.g., the MA task) as the target data to test the models.

The major contributions of this work are three-fold. First, we design a workload paradigm including working memory and mathematic addition tasks with fine-grained partition. Second, we propose to use domain adaptation to reduce the distribution discrepancy as well as to improve the classification accuracy. Third, we evaluate the proposed method on a real EEG dataset, with results demonstrating the superiority of our method over non-transfer methods.

The rest of the work is organized as follows. Section II briefly introduces the concepts of domain adaptation and related methods. Section III introduces EEG data recordings during two different cognitive workload tasks. The results are compared and presented in Section IV to evaluate the performance of the proposed methods. Section V discusses the major findings. Finally, Section VI concludes the whole paper.

II. METHODS

We adopt unsupervised domain adaptation without using the labeled samples from the target subjects, to cope with the task-to-task variability for building the EEG-based cross-task cognitive workload recognition models.

The EEG signal collected in one task of each subject is viewed as a domain, which is defined as $\{(X^i, y^i)\}_{i=1}^N$, where N is the number of trials used. $X^i \in \mathbb{R}^{(E \times d)}$ is an EEG trial with E electrodes and d dimensional features, and $y^i \in \mathbb{R}^C$ is the corresponding label for X^i of C classes.

The source domain is the labeled EEG samples of one task and the target domain is the unlabeled EEG samples of the other task. Given a source domain $\mathcal{D}_S = \{(X_s^i, y_s^i), \dots, (X_s^{N_s}, y_s^{N_s})\}$ with N_s labeled samples, and a target domain $\mathcal{D}_T = \{X_T^1, \dots, X_T^{N_t}\}$ with N_t unlabeled samples, domain adaptation aims to predict the labels $y_t^1, \dots, y_t^{N_t}$ corresponding to the inputs $X_T^1, \dots, X_T^{N_t}$ in the target domain, using the learned knowledge from the source domain. Generally, domain adaptation assumes the source and target domains

have the same feature spaces and label spaces, i.e., $X_S, X_T \in \mathbb{R}^{(E \times d)}$ and $y_S, y_T \in \mathbb{R}^C$, but the marginal probability distribution and/or conditional probability distribution are different, i.e., $P(X_S) \neq P(X_T)$ and/or $P(y_S | X_S) \neq P(y_T | X_T)$ [22]. Specifically, in the paper, we follow the key assumption in most domain adaptation methods, that the source and target domains have different marginal probability distributions and different conditional probability distributions.

In the following, we will briefly introduce four domain adaptation methods used in the paper. These methods mainly focus on the shared feature representation and minimization of the distribution discrepancy.

A. Transfer Component Analysis

Transfer component analysis (TCA) tries to reduce the distribution discrepancy via embedding the source and target domains into a shared low-dimensional feature space and learning a set of transfer components [23]. TCA can be seen as a dimensionality reduction method. To achieve this goal, Pan *et al.* proposed to find a transformation function $\phi(\cdot)$ such that the $P(\phi(X_S)) \approx P(\phi(X_T))$ and $P(y_S | \phi(X_S)) \approx P(y_T | \phi(X_T))$. Such $\phi(\cdot)$ should be satisfied that, it can reduce the distance between the domain distribution (e.g., the empirical means) and preserve the variance of the original data at the same time. By adopting the empirical Maximum Mean Discrepancy (MMD) [24] as the distance measure, the final mathematical representation of TCA is

$$\begin{aligned} \min_W & \text{tr}(W^T K L K W) + \mu \text{tr}(W^T W), \\ \text{s.t. } & W^T K H K W = I_m, \end{aligned} \quad (1)$$

where the first term is the MMD distance, i.e., $\left\| \frac{1}{N_s} \sum_{i=1}^{N_s} W^T \phi(X_i) - \frac{1}{N_t} \sum_{j=1}^{N_t} W^T \phi(X_j) \right\|^2 = \text{tr}(W^T K L K W)$, $K = [\phi(X_i)^T \phi(X_j)]$ is the kernel matrix defined on all the data. $L_{ij} = 1/N_s^2$, if $X_i, X_j \in X_s$, else $L_{ij} = 1/N_t^2$, if $X_i, X_j \in X_t$, otherwise $L_{ij} = -1/(N_s N_t)$. $W \in \mathbb{R}^{(N_s + N_t) \times m}$ is the transformation matrix that transforms the kernel matrix K to an m -dimension space ($m \ll N_s + N_t$). A regularization term $\text{tr}(W^T W)$ is added to control the complexity of the W , $\mu > 0$ is the trade-off parameter. $W^T K H K W = I$ is the constraint to keep the structure of the original data, where $H = I_{N_s + N_t} - (1/(N_s + N_t)) \mathbf{1} \mathbf{1}^T$ is the centering matrix, $I_{N_s + N_t} \in \mathbb{R}^{(N_s + N_t) \times (N_s + N_t)}$ is the identity matrix, and $\mathbf{1} \in \mathbb{R}^{N_s + N_t}$ is the column vector with all 1's. Finally, W is composed of the m leading components of the matrix $(K L K + \mu I)^{-1} K H K$.

B. Joint Distribution Adaptation

Joint distribution adaptation (JDA) is proposed to adopt both the marginal and conditional distributions in a principled dimensionality reduction procedure [25],

$$\begin{aligned} D(\mathcal{D}_S, \mathcal{D}_T) & \approx D(P(X_S), P(X_T)) \\ & + D(P(y_S | X_S), P(y_T | X_T)), \end{aligned} \quad (2)$$

Similar to TCA, JDA tries to find a transformation matrix W , such that $P(W^T X_S) \approx P(W^T X_T)$ and

$P(y_S | W^T X_S) \approx P(y_T | W^T X_T)$. Marginal distributions can be constructed by the TCA method, however, it is nontrivial to adopt the conditional distributions since there are no labeled data in the target domain. Long *et al.* proposed to use the pseudo labels of the target data and class-conditional distributions ($P(X_S | y_S = c)$ and $P(X_T | y_T = c)$, $c \in \{1, \dots, C\}$ is in the label space) to approximate conditional distributions. Since the pseudo labels may be unreliable, we can iteratively update them. The JDA optimization problem is

$$\begin{aligned} \min_W & \text{tr}(W^T K L K W) + \sum_{c=1}^C \text{tr}(W^T K M_c K W) \\ & + \lambda \|W\|_F^2, \\ \text{s.t. } & W^T K H K W = I, \end{aligned} \quad (3)$$

where the first term is the MMD distance between marginal distribution, the second term $\left\| \frac{1}{n_s^{(c)}} \sum_{X_i \in \mathcal{D}_s^{(c)}} W^T \phi(X_i) - \frac{1}{n_t^{(c)}} \sum_{X_j \in \mathcal{D}_t^{(c)}} W^T \phi(X_j) \right\|^2 = \text{tr}(W^T K M_c K W)$ is the MMD distance between the class-conditional distribution, where $\mathcal{D}_s^{(c)} = \{X_i : X_i \in \mathcal{D}_s \wedge y(X_i) = c\}$ is the set of examples belonging to class c in the source data, $y(X_i)$ is the true label of X_i , and $n_s^{(c)} = |\mathcal{D}_s^{(c)}|$; $\mathcal{D}_t^{(c)} = \{X_j : X_j \in \mathcal{D}_t \wedge \hat{y}(X_j) = c\}$ is the set of examples belonging to class c in the target data, $\hat{y}(X_j)$ is the pseudo labels of X_j , and $n_t^{(c)} = |\mathcal{D}_t^{(c)}|$; $(M_c)_{ij} = 1/(n_s^{(c)} n_t^{(c)})$, if $X_i, X_j \in \mathcal{D}_s^{(c)}$; else $(M_c)_{ij} = 1/(n_t^{(c)} n_t^{(c)})$, if $X_i, X_j \in \mathcal{D}_t^{(c)}$; else $(M_c)_{ij} = -1/(n_s^{(c)} n_t^{(c)})$, if $X_i \in \mathcal{D}_s^{(c)}, X_j \in \mathcal{D}_t^{(c)}$ or $X_j \in \mathcal{D}_s^{(c)}, X_i \in \mathcal{D}_t^{(c)}$; otherwise $(M_c)_{ij} = 0$. The third term is the regularization term. The W can be obtained by the following generalized eigen-decomposition problem and composed of its d smallest eigenvectors, $(K(L + \sum_{c=1}^C M_c)K^T + \lambda I)W = K H K^T W \Phi$, where $\Phi = (\phi_1, \dots, \phi_d)$ are the Lagrange multipliers.

C. Balanced Domain Adaptation

As mentioned above, the TCA method only considers the marginal distribution of source and target domains, and JDA considers both the marginal and conditional distributions. Though JDA may have more information used, it assumes the marginal and conditional distributions are dedicating identically to the domain divergence, which is not practical in real-world applications. To address this problem, Wang *et al.* [26] proposed a balanced domain adaptation (BDA) by using a balance factor μ to exploit the different importance of distributions,

$$\begin{aligned} D(\mathcal{D}_S, \mathcal{D}_T) & \approx (1 - \mu) D(P(X_S), P(X_T)) \\ & + \mu D(P(y_S | X_S), P(y_T | X_T)), \end{aligned} \quad (4)$$

where $\mu \in [0, 1]$, when $\mu \rightarrow 0$, the marginal distribution is more important to adapt, when $\mu \rightarrow 1$, the conditional distribution is more important. The optimization

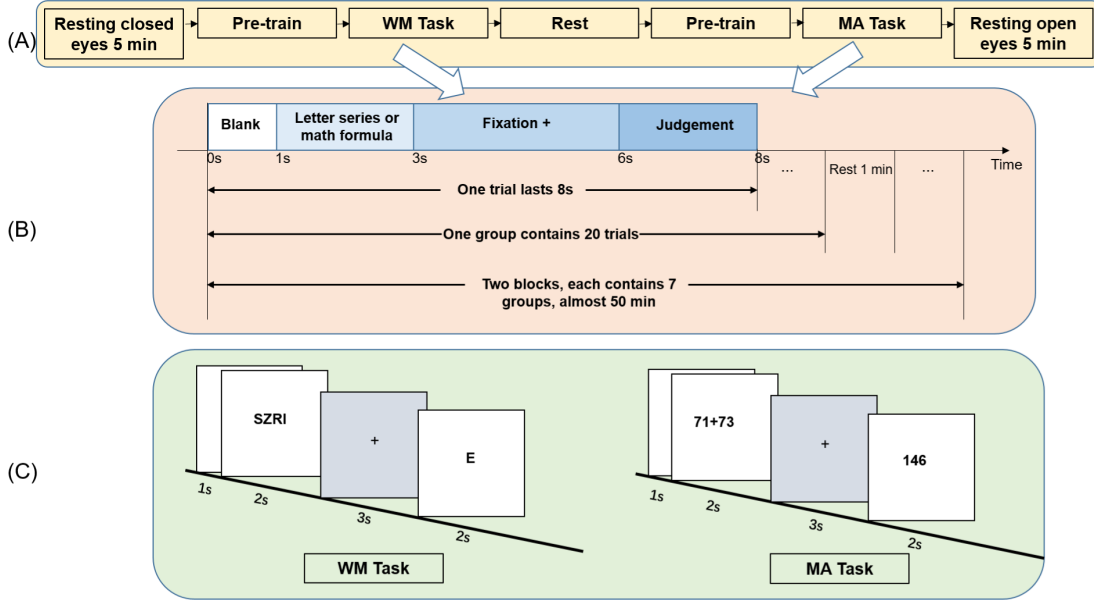


Fig. 2. The experimental design in this paper, including (A) the general task design, (B) the design of WM task and MA task, and (C) a detailed trial of both tasks at a difficulty level of L3.

problem of BDA is

$$\begin{aligned} \min_W \quad & tr \left(W^T K ((1 - \mu)) L + \mu \sum_{c=1}^C M_c K^T W \right) \\ & + \lambda \|W\|_F^2, \\ \text{s.t.} \quad & W^T K H K W = I, \quad \mu \in [0, 1]. \end{aligned} \quad (5)$$

The W is obtained by solving the following problem, $(K((1 - \mu))L + \mu \sum_{c=1}^C M_c K^T + \lambda I)W = K H K^T W \Phi$, and composed of its d smallest eigenvectors, where $\Phi = (\phi_1, \dots, \phi_d)$ are the Lagrange multipliers.

D. Transfer Joint Matching

When the source and target domains are different in both feature distribution and samples relevance, Long *et al.* [27] proposed Transfer Joint Matching (TJM) to cope with this setting. TJM aims to reduce the domain discrepancy by simultaneously matching the marginal feature distributions and reweighting the source samples across domains in a principled dimensionality reduction procedure, and construct a new feature representation that is stationary to both the marginal distribution discrepancy and the irrelevant samples,

$$\begin{aligned} \min_W \quad & tr \left(W^T K L K^T W \right) + \lambda (\|W_s\|_{2,1} + \|W_t\|_F^2), \\ \text{s.t.} \quad & W^T K H K^T W = I, \end{aligned} \quad (6)$$

here, the first term is the MMD distance between marginal distribution, the second is the regularizer of sample reweighting, including the row-sparsity for source sample in the sample space. Followed with TCA, the constraint is retaining the structure of the original data. So, the new feature representation is $Z = W^T K$. Finding the optimal adaptation matrix W is reduced to solving following generalized eigen-decomposition for the d smallest eigenvectors, $(K L K^T + \lambda G)W = K H K^T W$, $\phi = (\phi_1, \dots, \phi_d)$ are the Lagrange multipliers, and G is a diagonal sub-gradient matrix.

We recommend the readers refer to the detailed descriptions of TJM [27].

III. DATA AND EXPERIMENT

A. Subjects and Experiment Setup

In this study, we have invited 45 college students (28 males and 17 females, aged 20 to 30, mean age of 24.6 ± 6.6) to participate in EEG experiments at the iBRAIN laboratory of Nanjing University of Aeronautics and Astronautics. Except for one subject, all others were right-handed. All subjects have fulfilled the inclusion criterion that having a normal or corrected-to-normal vision. They did not suffer from any condition that may cause anxiety or fatigue. Participants were required to keep away from caffeine, medication, and alcohol, and had a normal amount of eight sleep hours before the experiment. After the explanation of the experimental protocol, we have obtained the signed written consent from all subjects.

Each subject undertook three tasks, including resting state with eye closed and eye open, WM task, and MA task. The general experimental design is displayed in Figure 2(A).

In the WM task, subjects needed to remember different numbers of stimuli groups (here we choose English letter sequences as stimuli), maintaining for 2 seconds, and determine whether the shown English letter had existed in the memorization sequences before [28]. The WM task consists of seven groups, and each group has 20 trials, 30% of them are target stimuli. We set the difficulty levels of seven, from the very low (L1), low (L2), medium (L3), medium-high (L4), high (L5), very high (L6) to extreme high (L7), with the length of sequences as 1, 2, 4, 6, 8, 10, 12, respectively. In each group, the order of levels was randomly presented. After a blank of 1s, the stimulus was presented for 2s followed by a fixation cross during the interval of 3s, then a judgment time of 2s.

In the MA task, the subjects needed to retain the results of the currently shown addition formula (e.g., $5+7$) and identify whether the given number (e.g., 11) matched the result they

TABLE I
DETAILS OF THE TASK DESIGN FOR ALL SEVEN LEVELS IN THE WM TASK AND MA TASK

Task levels	Working Memory Task		Mathematic Addition Task		
	#Letters	Example	#Digits	#Carries	Example
Very low (L1)	1	D	2&1	0	15+2
Low (L2)	2	NS	2&1	1	42+9
Medium (L3)	4	SZRI	2&2	1	71+73
Medium-high(L4)	6	NDVXUM	2&2	2	98+16
High (L5)	8	NEFODHLK	3&2	1	563+72
Very high (L6)	10	SWFZTIOXMP	3&2	2	515+87
Extremely high (L7)	12	RWFPIOYNEJLK	3&3	3	236+774

calculated before. This task involves temporal storage of intermediate results and information retrieval held in the cognitive workspace [11]. Similarly, the MA task consists of seven groups, and each group has 20 trials, 30% of them are target stimuli. We set the difficulty levels of seven, corresponding with the various digits and carries of addition. In each group, the order of levels is randomly presented. After a blank of 1s, each addition was presented for 2s followed by a fixation cross for 3s, while the answer is displayed for 2s.

In Figure 2(B), we display the detailed setup for each cognitive task, and Figure 2(C) shows a detailed trial of both tasks at a difficulty level of L3. In Table I, we list the level design and the examples of various difficulty levels.

All task stimuli were shown on a computer screen in white font on a black background. Participants were instructed to focus on both accuracy and speed and practice five trials before the EEG recordings were implemented for both tasks. Notably, mental fatigue may be induced during the experiment and it may contaminate the EEG quality. To avoid this, we first randomized the group order in each task. Then, we randomized the order of difficulty levels in each group. Also, we asked the subjects to rest before each task to ease the fatigue. The task paradigms were carried out in E-Prime 2.0 software [29].

B. Data Acquisition and Preprocessing

A portable wireless EEG amplifier (NeuSen. W64, Neuracle, China) was used for EEG data recording at a sampling rate of 1000 Hz. Fifty-nine electrodes were arranged according to the international 10-20 system, with reference at CPz and a forehead ground at AFz. Electrode impedances were kept below 5 k Ω for all electrodes throughout the experiment. In addition to the EEG signals, the reaction time and the answer accuracy of the subjects were also recorded as objective behavior data to evaluate the cognitive workload levels.

A widely-used EEG preprocessing pipeline was adopted in the current study, using EEGLAB software [30]. Specifically, the raw EEG data were firstly re-referenced to the average of the electrodes, then, band-pass filtered to 0.1-70 Hz to eliminate noise, and additionally a 50 Hz notch filtered to reduce the power line noise. The filtered data were down-sampled to 256 Hz. Signals were then baseline adjusted and segmented into 2s epochs after stimulus onset. Eye-blink and muscle-related artifacts were removed via Independent Components Analysis by rejecting the components that were highly correlated with those artifacts. Here we used ADJUST [31] and ICLabel [32] tools to mark the components. Due to high noise contamination, seven subjects were excluded, thus leaving

TABLE II
THE NUMBER OF SAMPLES OF DIFFERENT TASKS

Level	# WM Task	# MA Task
L1	648	668
L2	633	661
L3	657	682
L4	644	666
L5	647	664
L6	654	671
L7	671	669
Sum	4554	4681

38 subjects (25 males and 13 females, mean age of 24.4 ± 5.9 years) for the subsequent analysis. For these subjects, the bad epochs with noise were further removed by visual inspection, thus leaving 4554 and 4681 trials (each is 2s length) for the WM task and MA task, respectively. In Table II, we list the corresponding numbers of the workload levels with two tasks.

C. Feature Extraction

For feature extraction, we have extracted power spectral density (PSD) features from the spectral dimension [33] and coherence features from the brain connectivity network [34], which are both widely used in EEG analysis [35], and combined them as the final feature set.

To be specific, the PSD and coherence features are extracted from 5 clinical frequency bands (δ [1~3 Hz], θ [4~7 Hz], α [8~13 Hz], β [14~30 Hz], and γ [31~50 Hz]), and PSD is extracted based on short-time Fourier transform. Short-time Fourier transform is a time-frequency domain decomposed method, and it decomposes signals into small sequential data frames with shifting windows, and a fast Fourier transform is then used to each frame [36]. We use Hanning windows with 0.5s window length (total is 2s) and no overlap. The calculation formula of PSD is $PSD = \frac{1}{NF_s} \left| \sum_{n=1}^N x[n] w[n] e^{-j2\pi fn/F_s} \right|^2$, where $x[n]$ is the EEG trials, $n = 1, 2, \dots, N$, F_s is the sampling rate, $w[n]$ is the window function. The total dimension of the PSD feature of a 59-channel EEG segment is 59 channels \times 4 windows \times 5 bands = 1180.

Besides the PSD feature, functional connectivity network can be used for measuring the relationship between different EEG electrodes [35] or functional brain regions [37]–[39]. EEG-related functional connectivity networks can be constructed by coherence [40]. The formula for coherence is $Coh_{xy}(\omega) = \frac{|P_{xy}(\omega)|}{\sqrt{P_{xx}(\omega)P_{yy}(\omega)}}$, where $P_{xy}(\omega)$ is the cross-spectrum of channel signals x and y , and $P_{xx}(\omega)$, $P_{yy}(\omega)$ are the power

TABLE III
THE DESIGN OF TRANSFER SCHEMES

Transfer schemes	Source	Target
$\mathcal{O} \rightarrow \mathcal{O}$	Task A of S_i	Task B of S_i
$\mathcal{M} \rightarrow \mathcal{O}$	Task A of $\{S_i\}_{i=1}^{n_s}$	Task B of S_j
$\mathcal{VO} \rightarrow \mathcal{O}$	Task A of various $S_i, i \neq j$	Task B of S_j

spectrum of x and y , respectively. Since the connection matrix is symmetric, i.e., the upper triangle is the same as the lower triangle, we thus remove the lower triangle entries. Since the main diagonal entries are both self-connections, we also remove them. The total dimension of coherence feature of a 59-channel EEG segment is $59 \times (59-1)/2 \times 5 \text{ bands} = 8555$. Finally, we combine the PSD and coherence as the final features (total 9735).

D. Workload Classification

We aim to build EEG-based cross-task cognitive workload recognition models using domain adaptation techniques. Here, during training for cross-task transfer, the samples of the labeled source data from task A are used to predict the unlabeled target data in task B. Considering the different application scenarios, we define three kinds of transfer schemes with different focuses.

- (1) *One-to-one cross-task transfer* (denoted as $\mathcal{O} \rightarrow \mathcal{O}$): is the transfer considering the intra-subject task variability, where we use the data from each subject to complete the cross-task cognitive workload recognition. Given one subject, the source domain is the data of task A, and the target domain is the data of task B of the same person.
- (2) *Many-to-one cross-task transfer* (denoted as $\mathcal{M} \rightarrow \mathcal{O}$): is the transfer focusing on the source combination, where the source domain is the data of task A of all the subjects, and the target domain is the data of task B of each subject. It makes sense when multiple existing subjects are available and it might have more information than $\mathcal{O} \rightarrow \mathcal{O}$.
- (3) *Various one-to-one cross-task transfer* (denoted as $\mathcal{VO} \rightarrow \mathcal{O}$): which considers the subject variabilities, where the target domain is the task B of one subject, the source domain is the enumerated task A from the other subjects.

We thus focus on evaluating the domain adaptation methods for a common binary classification task. We take levels L1, L2, L3 as low workload, and levels L5, L6, L7 as high workload.

Here we adopt widely used classifiers for baseline comparison, including supporting vector machine (SVM) with radial basis function (RBF) kernel, K-nearest neighbor (KNN), linear discriminative analysis (LDA), and single hidden-layer artificial neural network (ANN). These methods also are non-transfer methods. SVM is constructed based on the RBF kernel with a soft margin parameter C . The parameter of KNN is the number of neighbors. Note that this paper only presents the best SVM and KNN results. The LDA classifier is used with the default setting provided by MATLAB. ANN is used with 10 hidden units. Based on the preliminary results, we then use SVM with RBF kernel as the base classifier for domain

TABLE IV
DETAILS OF HYPER-PARAMETER FOR THE USED METHODS

Method	Hyper-parameters
SVM	RBF kernel, $C = \{10^{-3}, 10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\}$
KNN	$k = \{1, 3, 5, 7, 9, 11, 13\}$
TCA\JDA\TJM	Kernel = {linear, RBF}, $\gamma = 1$, $\lambda = \{0.1, 1\}$, $d = \{5, 10, \dots, 100\}$
BDA	Kernel = {linear, RBF}, $\gamma = 1$, $\lambda = \{0.1, 1\}$, $d = \{5, 10, \dots, 100\}$, $\mu = \{0.1, 0.2, \dots, 1\}$

adaptation methods. For the TCA, JDA, and TJM methods, the hyper-parameters are the dimension of subspace d , the kernel function, RBF kernel-based γ , trade-off parameter λ ; for BDA, it also includes the balance factor μ . The related hyper-parameters are summarized in Table IV. We first use the MA task and WM task as source domain (training data) and target domain (testing data), respectively, as MA \rightarrow WM. Then, we select data in reverse order, as WM \rightarrow MA. So, for each method, we repeat the classification process 228 times (38 subjects \times 2 orders \times 3 conditions).

IV. EXPERIMENTAL RESULTS

The results for recognizing cognitive workload consist of two parts. First, behavioral data including response time and answer accuracy, and event-related spectral perturbation (ERSP) are analyzed to validate the different cognitive workload levels induced by WM and MA tasks. Second, the binary classification results of cross-task workload models are presented, using accuracy and F1 score as performance evaluation metrics.

A. Behavioral Results

We use the WM task and MA task to explore cross-task cognitive workload recognition. In the experiment, we have recorded the response time and answer accuracy as behavior data. Due to the non-block experimental design and random order of experimental stimulus representation [41], we did not collect the subjective measurements, such as the NASA-TLX questionnaire. Figure 3 presents the behavior data to validate the effectiveness of these two types of cognitive tasks and the corresponding one-way analysis of variance (ANOVA) results. This experiment confirms that easy and hard tasks are distinguishable.

For both the MA and WM tasks, when the difficulty levels increase, the subjects perform worse as in Figure 3(A) and (B), and take longer to provide the answers as in Figure 3(C) and (D). For the MA task, the accuracy difference between levels 1, 2, 3 and levels 4, 5, 6, 7 ($p < 0.01$), and the difference between levels 4, 5, 6, and level 7 ($p < 0.01$) are significant; the difference between levels 1, 2, 3 and the difference between levels 4, 5, 6 are not significant. For response time, the difference between levels 1, 2, 3 and levels 5, 6, 7 ($p < 0.01$), and the difference between level 4 and level 7 ($p < 0.01$) are significant; the difference between levels 1, 2, 3, 4, the difference between levels 4, 5, 6 and the difference between levels 5, 6, 7 are not significant.

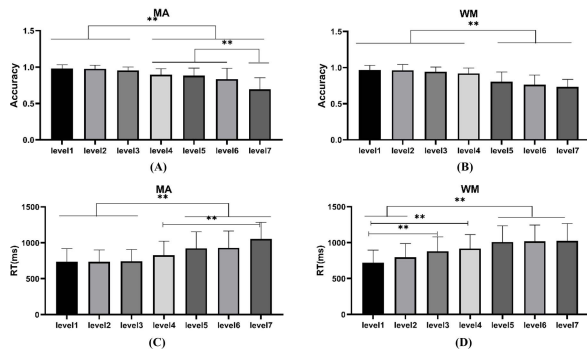


Fig. 3. Results of the one-way ANOVA of the behavioral data. Bars represent mean \pm standard error. WM task: working memory task; MA task: mathematic addition; RT: Response Time; ** indicates $p < 0.01$; Games-Howell's correction was used for multiple comparisons.

For the WM task, the accuracy difference between levels 1, 2, 3, 4, and levels 5, 6, 7 ($p < 0.01$) are significant; the difference between levels 1, 2, 3, 4, and the difference between levels 5, 6, 7 are not significant. For response time, the difference between levels 1, 2 and levels 5, 6, 7 ($p < 0.01$), and the difference between level 1 and levels 3, 4 ($p < 0.01$) are significant; the difference between levels 1, 2, the difference between levels 3, 4, and the difference between levels 4, 5, 6, 7 are not significant.

We analyze EEG using the event-related spectral perturbation (ERSP) [20], [42] and perform the one-way ANOVA with 2000 permutations for statistical testing. ERSP provides detailed information on event-related desynchronization/ synchronization and can visualize the mean power changes [43]. In Figure 4, we analyze the ERSP maps for each cognitive task, for each of the 7 stimulus conditions in all 59 electrodes, and in the four EEG frequency-bands (θ , α , β , and γ) during the 0-1000ms after task stimulus, using EEGLAB [30].

For the WM task, as shown in Figure 4(A), in θ band, levels 1 and 2, levels 4,5,6, and 7 have similar ERSP patterns, respectively. In α and β band, level 1 to level 7 have similar ERSP. Significant differences between level 1 to level 7 are confined in θ band with frontal, right temporal, and occipital brain regions ($p < 0.05$), whereas the difference in the α , β , and γ bands are not significant.

For the MA task, as shown in Figure 4(B), in θ band, levels 1 and 2, levels 4,5, and 6 have similar ERSP patterns, respectively. In α and β band, level 1 to level 7 have similar ERSP patterns. Significant differences between level 1 to level 7 are confined in θ band with frontal and left occipital brain regions ($p < 0.05$), and γ band with occipital, and central brain regions ($p < 0.05$), whereas the difference in the α and β bands are not significant. Furthermore, the ERSP distributions between the WM task and MA task have a similar pattern in the θ , α , and β bands.

B. Classification Results

Table V displays the classification accuracy of different methods and Supplementary Table S1 validates the differences between different methods. Here, SVM, LDA, KNN, and ANN are non-transfer methods, whereas TCA, JDA, BDA, and TJM are domain adaptation methods with SVM as the classifier.

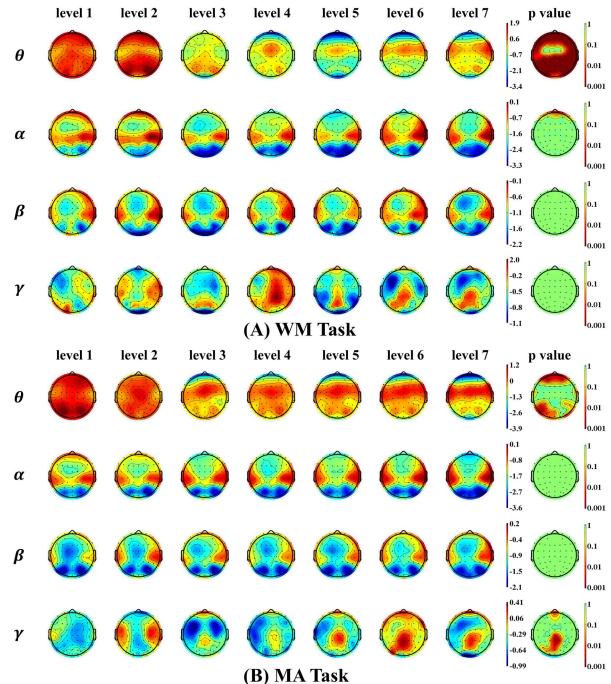


Fig. 4. Results of the one-way ANOVA for ERSP analysis. WM task: working memory task; MA task: mathematic addition. For different levels, blue indicates event-related desynchronization, red indicates event-related synchronization. The last column represents p values, with redder color indicating the stronger significance; FDR correction was used for multiple comparisons.

In $\mathcal{O} \rightarrow \mathcal{O}$, $\mathcal{M} \rightarrow \mathcal{O}$, and $V\mathcal{O} \rightarrow \mathcal{O}$ transfers, we find the proposed TJM method has the best average accuracy both for the $\mathcal{MA} \rightarrow \mathcal{WM}$ and $\mathcal{WM} \rightarrow \mathcal{MA}$. For non-transfer methods, SVM performs better than other methods. Specifically, in $\mathcal{O} \rightarrow \mathcal{O}$ task transfer, compared TJM with SVM, the average accuracy is increased by 8.36% in $\mathcal{MA} \rightarrow \mathcal{WM}$, 5.76% in $\mathcal{WM} \rightarrow \mathcal{MA}$, 7.06% on average. In $\mathcal{M} \rightarrow \mathcal{O}$ task transfer, compared with SVM, the average accuracy of TJM is increased by 7% in $\mathcal{MA} \rightarrow \mathcal{WM}$, 4.81% in $\mathcal{WM} \rightarrow \mathcal{MA}$, 5.91% on average. When comparing $\mathcal{O} \rightarrow \mathcal{O}$ to $\mathcal{M} \rightarrow \mathcal{O}$, we find SVM, ANN, and all domain adaptation methods have an increasing trend as having more source data samples. In $V\mathcal{O} \rightarrow \mathcal{O}$ task transfer, we have averaged the classification results of various source domains for each target domain, and have reported the average classification results of all the target domains as the final results. Compared with SVM, the average accuracy of TJM is increased by 3.59% in $\mathcal{MA} \rightarrow \mathcal{WM}$, 3.17% in $\mathcal{WM} \rightarrow \mathcal{MA}$, 3.38% on average. When comparing the $\mathcal{O} \rightarrow \mathcal{O}$ to $\mathcal{M} \rightarrow \mathcal{O}$, the improvement of $\mathcal{M} \rightarrow \mathcal{O}$ design might come from the more training data; for $\mathcal{M} \rightarrow \mathcal{O}$ design, the proposed transfer methods outperform the non-transfer methods, we argue the improvement of this result might come from the proposed domain adaptation frameworks. Comparing the $\mathcal{O} \rightarrow \mathcal{O}$, $V\mathcal{O} \rightarrow \mathcal{O}$, and $\mathcal{M} \rightarrow \mathcal{O}$ classification results, we have found that the $V\mathcal{O} \rightarrow \mathcal{O}$ setting has lower classification accuracy than the other two, mainly due to the huge subject variabilities existing in this way. It indicates the importance to reduce the effect of subject variabilities.

TABLE V
THE CLASSIFICATION ACCURACY (%) WITH MEAN AND STANDARD DEVIATION

Transfers	Non-transfer Methods				Transfer + SVM			
	SVM	KNN	LDA	ANN	TCA	JDA	BDA	TJM
$\mathcal{O} \rightarrow \mathcal{O}$								
MA→WM	62.6 \pm 9.71	58.63 \pm 7.99	62.5 \pm 9.33	57.67 \pm 8.53	68.43 \pm 10.26	70.56 \pm 10.67	68.1 \pm 9.96	70.96 \pm 9.31
WM→MA	60.94 \pm 8.4	55.43 \pm 5.8	60.93 \pm 8.33	54.86 \pm 5.38	65.9 \pm 7.66	66.43 \pm 8.44	65.5 \pm 8.49	66.7 \pm 7.39
Mean	61.77 \pm 1.17	57.03 \pm 2.27	61.72 \pm 1.12	56.27 \pm 1.99	67.17 \pm 2.5	68.5 \pm 2.92	66.8 \pm 1.84	68.83 \pm 3.01
$\mathcal{M} \rightarrow \mathcal{O}$								
MA→WM	66.53 \pm 9.73	54.48 \pm 6.03	57.22 \pm 6.5	62.62 \pm 8.88	70.23 \pm 8.29	72.54 \pm 7.41	70.09 \pm 8.46	73.53 \pm 8
WM→MA	64.37 \pm 7.97	54.09 \pm 4.39	56.92 \pm 6.42	61.18 \pm 7.39	65.64 \pm 8.74	68.22 \pm 8.08	65.55 \pm 8.61	69.18 \pm 7.85
Mean	65.45 \pm 1.53	54.29 \pm 0.28	57.07 \pm 0.21	61.9 \pm 1.02	67.94 \pm 3.25	70.38 \pm 3.05	67.82 \pm 3.21	71.36 \pm 3.08
$\mathcal{V}\mathcal{O} \rightarrow \mathcal{O}$								
MA→WM	55.74 \pm 3.91	52.38 \pm 2	55.59 \pm 3.65	53.62 \pm 2.85	55.61 \pm 4.29	57.45 \pm 4.88	55.15 \pm 4.35	59.33 \pm 4.51
WM→MA	54.75 \pm 4.23	51.67 \pm 1.68	54.66 \pm 4.06	52.79 \pm 2.57	54.47 \pm 4.36	56.16 \pm 4.76	54.03 \pm 4.05	57.92 \pm 4.43
Mean	55.25 \pm 0.7	52.03 \pm 0.5	55.13 \pm 0.66	53.21 \pm 0.59	55.04 \pm 0.81	56.81 \pm 0.91	54.59 \pm 0.79	58.63 \pm 1

THE BEST RESULTS ARE BOTH BOLD FOR NON-TRANSFER AND DOMAIN ADAPTATION METHODS.

Supplementary Table S1 validates the differences between different methods using Dunn’s test with multiple comparison corrections. In both $\mathcal{O} \rightarrow \mathcal{O}$ and $\mathcal{M} \rightarrow \mathcal{O}$ designs, the difference between TJM and non-transfer methods (SVM, LDA, KNN, and ANN) are significant ($p < 0.05$); the difference between all transfer methods to KNN and LDA are significant ($p < 0.05$). For domain adaptation, TCA versus BDA, JDA versus TJM have no significant difference ($p > 0.05$). In $\mathcal{V}\mathcal{O} \rightarrow \mathcal{O}$ setting, the difference between non-transfer methods (SVM, LDA, KNN, and ANN) versus JDA and TJM are significant ($p < 0.05$). For domain adaptation, the difference between TCA versus TJM, BDA versus TJM are significant ($p < 0.05$), whereas JDA versus TJM, JDA versus BDA have no significant difference ($p > 0.05$). To further display the performance, in Figure 5, we show the accuracy distributions across different algorithms, using box-plot figures. Besides, Supplementary Table S2 shows the F1 score results, which have similar results in Table V. We also provide additional case 1 and case 2 settings for binary cross-task cognitive workload recognition, here, in case 1, we take level L1 as low workload and level L7 as high; in case 2, we take levels L1, L2 as low workload, levels L6, L7 as high. Supplementary Tables S3 and S4 show the classification accuracy and F1 score results for both case 1 and case 2. In sum, the proposed transfer framework has better performance and significant improvements than non-transfer methods, especially for TJM. As such, both cognitive tasks generate positive transfers in the proposed cross-task framework.

V. DISCUSSION

The discussion of cross-task cognitive workload recognition consists of five parts. First, the classification tasks with combinations of different source numbers and the dimensions of subspace for the TJM are conducted in Figure 6. Second, accuracy comparisons between PSD and coherence are employed in Figure 7. Third, PSD and coherence are employed separately to evaluate their impacts in Figures 8 and 9. Fourth, a nonlinear dimensionality reduction is used to visualize the feature representation learned after TJM in Figure 10. Finally, the limitations and future work are presented.

A. Parameter Sensitivity

To figure out the best source combination number in $\mathcal{M} \rightarrow \mathcal{O}$ design, we also conduct classification tasks using the

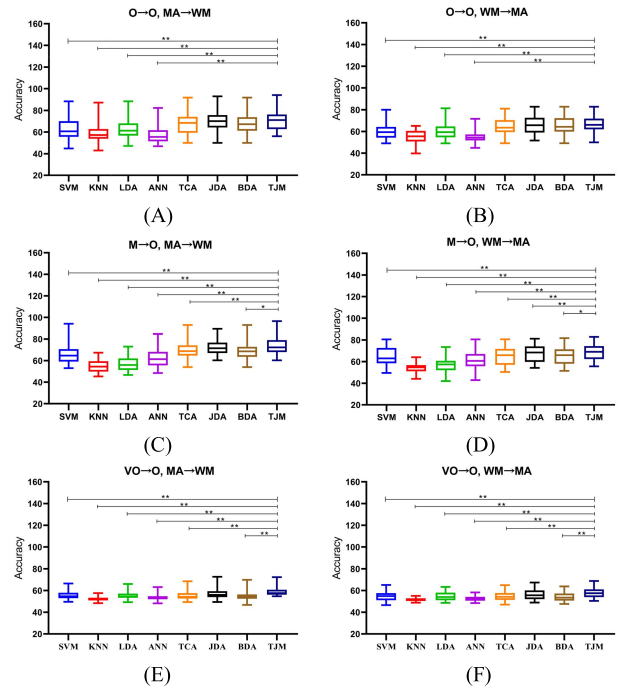


Fig. 5. Accuracy distributions for each algorithm and differences between the TJM and the other methods. Here, (A) and (B) are in $\mathcal{O} \rightarrow \mathcal{O}$ design, (C) and (D) are in $\mathcal{M} \rightarrow \mathcal{O}$ design, (E) and (F) are in $\mathcal{V}\mathcal{O} \rightarrow \mathcal{O}$ design. The first column is trained on the MA task and tested on the WM task, the second column is in reverse order. ** indicates $p < 0.01$; * indicates $p < 0.05$; Dunn’s test with multiple comparison correction was used.

TJM method with different source combination numbers from [4, 6, 8, \dots , 38] and the dimension of subspace from [40, 60, 80, 100], as shown in Figure 6. Here, (A) is trained on the MA task and tested on the WM task, (B) is in reverse, and d is the dimension number of TJM. For (A), the highest accuracy is obtained at source number $N_S = 28$, and $d=80$. For the MA→WM task, when the $N_S < 20$, the overall trend is increasing, and the low-dimensional d can obtain acceptable performance, such as $d = 40$ in (A); when $N_S > 20$, the bigger d shows more stable power. For (B), the highest accuracy is obtained at source number $N_S = 36$, and $d=80$. For the WM→MA task, the accuracy trend is first decreasing at $N_S = 16$, then increasing. Similarly, when $N_S > 16$, the higher d shows a stable increase. Since we direct combine

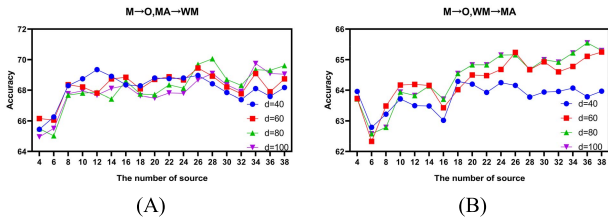


Fig. 6. Accuracy for TJM algorithm with different source combinations, where (A) is trained on MA task and tested on WM task, (B) is reverse, and d is the dimension number of TJM.

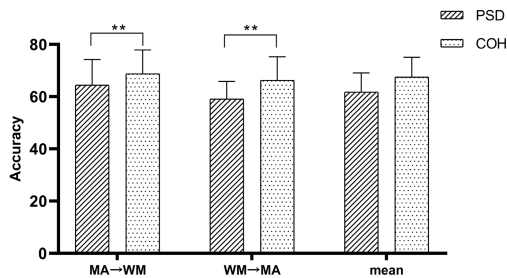


Fig. 7. Accuracy comparison between PSD and coherence (COH) features in a $\mathcal{M} \rightarrow \mathcal{O}$ design. Bars represent mean \pm standard error. Here, the mean is the average accuracy of MA \rightarrow WM and WM \rightarrow MA transfers. Paired t-test is used for feature comparisons, and ** indicates $p < 0.01$.

data in all the source domains as a new source domain, the training time of $\mathcal{M} \rightarrow \mathcal{O}$ design is longer than $\mathcal{O} \rightarrow \mathcal{O}$. The above results show that the multi-source selection might be useful for fast cognitive workload recognition.

B. Features Comparisons

To evaluate the different features on the results of the TJM method, we repeat the classification task using PSD and coherence features separately. We set the kernel function as linear, the dimension of subspace as 80, and the other experimental settings stay the same. Accuracy comparisons between PSD and coherence are displayed in Figure 7.

Coherence has an advantage over PSD for about 4.3% and 7.1% increments in the MA \rightarrow WM and WM \rightarrow MA transfers, respectively ($p < 0.01$, paired t-test). When compared with the average accuracy of the two transfers, the coherence has an advantage over PSD with about a 5.7% increment. In sum, in terms of the mean accuracy, the coherence performs better than PSD, although these two features have different feature numbers. Previous studies have shown that brain activity is usually completed by multiple brain regions [34], [35]. We argue that the functional connectivity feature (i.e., coherence used in the study) can better reflect the correlation between EEG electrodes, so coherence achieves better classification results than PSD.

C. Feature Importance

We also investigate the importance of the original features for both two tasks. In Figure 8, we show the PSD distributions in low and high workload conditions with the mean PSD value of each channel across 5 frequency bands. Here, the whole

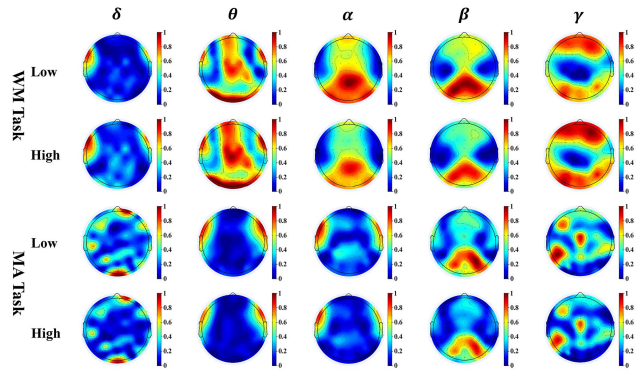


Fig. 8. The PSD distribution in low and high workload, where the upper is for WM task and the lower is for MA task. The color bar below indicates the mean PSD for each frequency band.

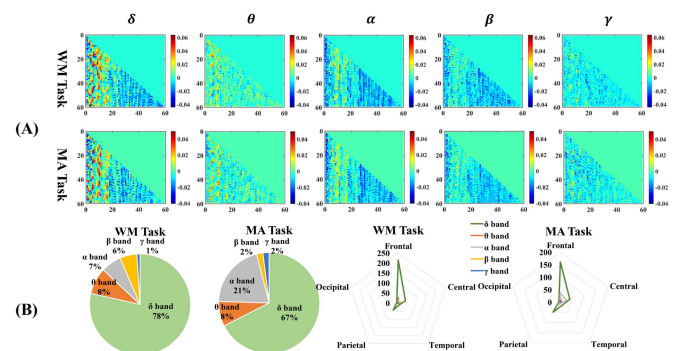


Fig. 9. The coherence variance distribution of low and high workload. (A) The upper is for the WM task and the lower is for the MA task. The color bar indicates the coherence variation, where the redder color represents an increasing trend and the bluer represents a decreasing trend. (B) We then show features distribution for the five frequency bands in the pie charts and the channel location in the radar charts for each task.

EEG brain is divided into the frontal, central, occipital, parietal, and temporal brain regions. We find the PSD distributions in the same task have a similar pattern but vary between the two tasks. For the WM task, with the workload increment, the θ frequency band in the frontal and central brain regions has an increased PSD; the α and β bands in the parietal and occipital brain regions both have a decreased PSD; the γ band in the frontal and occipital brain regions has an increased PSD. For the MA task, the θ and α bands in the temporal regions have similar PSD distribution; the β band in the occipital region has a decreased PSD with the workload increment.

In Figure 9, we show the coherence variance distribution of low and high workloads. To obtain the coherence variance distribution, we first use the mean coherence features of high workload to minus the low workload in corresponding frequency bands. In Figure 9(A), the upper is for the WM task and the lower is for the MA task. As we can see, the coherence variance distributions have a similar pattern for both tasks, where δ band in frontal regions has great importance and the coherence values have an increasing trend, whereas the coherence values in β and α bands from central and occipital regions have a decreasing trend. To further find the workload-related frequency bands and channels, we then use a t-test to select the most significant top 200 features ($p < 0.01$). In Figure 9(B),

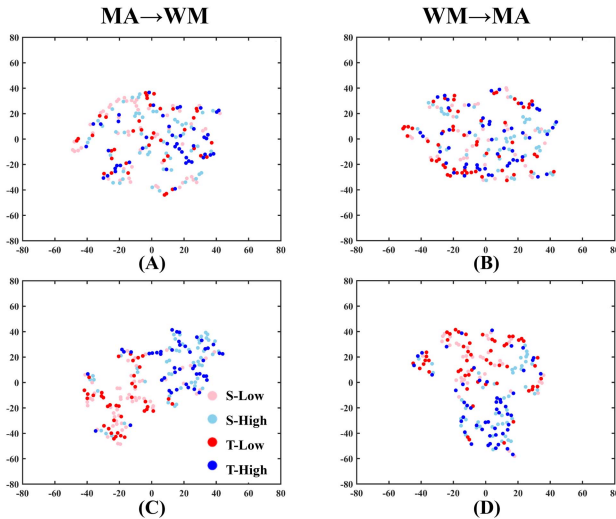


Fig. 10. Feature visualization by t-SNE. Here, (A) and (B) are the features of the original distribution, (C) and (D) are the features after TJM. The first column is trained on the MA task and tested on the WM task, the second column is in reverse order. Light colors denote features from the source domain and deep colors represent features from the target domain. For better visualization, we also highlight features from different categories with different colors (i.e., red and blue). S is Source, and T is Target domain.

we show features distribution for the five frequency bands in the pie charts and the channel locations in the radar charts for each task. The majority of coherence features in the δ band included frontal, central, and occipital regions, following with theta and alpha bands. The PSD features and coherence features describe the EEG characteristics from the spectral and spatial connectivity perspectives, respectively, thus combining them may offer meaningful information.

D. Feature Visualization

To visualize the distributions of feature representation learned by TJM, we try to project the latent feature representations on a two-dimensional (2D) plane using *t*-distributed stochastic neighbor embedding (t-SNE) in Figure 10. T-SNE is a nonlinear dimensionality-reduction method, aiming to sustain data structure in a low-dimensional space [44]. For brevity, we randomly select one target from the dataset in $\mathcal{O} \rightarrow \mathcal{O}$ design and draw the t-SNE maps on 2D space. Here, Figures 10 (A) and (B) are the features of the original distribution, (C) and (D) are the features after TJM. The first column is trained on the MA task and tested on the WM task, the second column is in reverse order. Light colors denote features from the source domain and deep colors represent features from the target domain. Red denotes low and blue denotes high workload class. Since we use a random order to display the experimental stimulus rather than the commonly-used block design, the classification task in our paper is harder to discriminate than the block design. As we can see from Figures 10 (A) and (B), originally, the samples belonging to different classes overlap substantially across domains, making it difficult to discriminate. In Figures 10 (C) and (D), the discrepancies between different tasks have been reduced, as the target samples are closer to the corresponding classes, which

also makes it easy to classify. Although it is easier to classify the low workload and high workload samples in (C) and (D) than (A) and (B), we also find the classification boundaries after the TJM still have some overlaps. We argue that some hard-to-distinguish samples might be misclassified and mislead the classification boundaries in source domains, indicating the need to construct the effective and strong classifier in source domains.

E. Limitations and Future Work

In this paper, the experimental results show that domain adaptation methods in a $\mathcal{M} \rightarrow \mathcal{O}$ design have better performance than $\mathcal{O} \rightarrow \mathcal{O}$ design, indicating the multi-source domain adaptation would enhance the classification performance than a single source. We argue the $\mathcal{M} \rightarrow \mathcal{O}$ design may provide more information and more data for constructing a powerful and robust model than $\mathcal{O} \rightarrow \mathcal{O}$ design. However, we use traditional domain adaptation methods in the experiment. Besides, we focus on the binary classification task in the paper, whereas the experimental design can be used to classify seven workload levels. These findings provide the requisite deep transfer models necessary for recognizing the multi-class workload levels for real-world use, due to their effectiveness and strong power [45], especially for multi-source deep transfer learning [46].

VI. CONCLUSION

In this paper, we have presented a preliminary study on EEG-based cross-task cognitive workload recognition using domain adaptation techniques, and a comparative study on a private EEG dataset. The cross-task workload recognition models have been constructed in $\mathcal{O} \rightarrow \mathcal{O}$, $\mathcal{M} \rightarrow \mathcal{O}$, and $V\mathcal{O} \rightarrow \mathcal{O}$ transfers. We have compared the performance of domain adaptation methods and conventional non-transfer classifiers. The experimental results have demonstrated the proposed task transfer framework outperforms the non-transfer classifier with an accuracy increase of 3% to 8%. Taking advantage of jointly considering the adaptation of data distributions and the weights of source samples, the TJM consistently achieves the best performance and significant improvements. The results provide the requisite deep and/or multi-source domain adaptation methods that might be necessary for recognizing the multi-class workload levels for real-world use.

REFERENCES

- [1] P. Aricò, G. Borghini, G. Di Flumeri, N. Sciaraffa, A. Colosimo, and F. Babiloni, "Passive BCI in operational environments: Insights, recent advances, and future trends," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1431–1436, Jul. 2017.
- [2] R. K. Mehta and R. Parasuraman, "Neuroergonomics: A review of applications to physical and cognitive work," *Frontiers Human Neurosci.*, vol. 7, p. 889, Dec. 2013.
- [3] C. D. Wickens, "Multiple resources and performance prediction," *Theor. Issues Ergonom. Sci.*, vol. 3, no. 2, pp. 159–177, Jan. 2002.
- [4] N. Sciaraffa, P. Aricò, G. Borghini, G. D. Flumeri, A. D. Florio, and F. Babiloni, "On the use of machine learning for EEG-based Workload assessment: Algorithms comparison in a realistic task," in *Human Mental Workload: Models and Applications*. Cham, Switzerland: Springer, 2019, pp. 170–185.
- [5] J. Heard, C. E. Harriott, and J. A. Adams, "A survey of workload assessment algorithms," *IEEE Trans. Human-Mach. Syst.*, vol. 48, no. 5, pp. 434–451, Oct. 2018.

- [6] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, Jan. 1988.
- [7] 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. Cognit. Develop. Syst.*, early access, Jun. 17, 2021, doi: 10.1109/TCDS.2021.3090217.
- [8] G. N. Dimitrakopoulos *et al.*, "Task-independent mental workload classification based upon common multiband eeg cortical connectivity," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1940–1949, Nov. 2017.
- [9] A. Appriou, A. Cichocki, and F. Lotte, "Modern machine-learning algorithms: For classifying cognitive and affective states from electroencephalography signals," *IEEE Syst., Man, Cybern. Mag.*, vol. 6, no. 3, pp. 29–38, Jul. 2020.
- [10] P. Zarjam, J. Epps, and F. Chen, "Characterizing working memory load using EEG delta activity," in *Proc. 19th Eur. Signal Process. Conf.*, Aug. 2011, pp. 1554–1558.
- [11] D. Guido, R. M. José del, H. Thilo, J. M. Dennis, and M. Klaus-Robert, "Improving human performance in a real operating environment through real-time mental workload detection," in *Toward Brain-Computer Interfacing*. Cambridge, MA, USA: MIT Press, 2007, pp. 409–422.
- [12] Z. Wang, R. M. Hope, Z. Wang, Q. Ji, and W. D. Gray, "Cross-subject workload classification with a hierarchical Bayes model," *NeuroImage*, vol. 59, no. 1, pp. 64–69, Jan. 2012.
- [13] P. Zarjam, J. Epps, and N. H. Lovell, "Beyond subjective self-rating: EEG signal classification of cognitive workload," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 4, pp. 301–310, Dec. 2015.
- [14] Y. Ke *et al.*, "An EEG-based mental workload estimator trained on working memory task can work well under simulated multi-attribute task," *Frontiers Hum. Neurosci.*, vol. 8, p. 703, Sep. 2014.
- [15] C. L. Baldwin and B. N. Penaranda, "Adaptive training using an artificial neural network and EEG metrics for within- and cross-task workload classification," *NeuroImage*, vol. 59, no. 1, pp. 48–56, Jan. 2012.
- [16] C. Walter, S. Schmidt, W. Rosenstiel, P. Gerjets, and M. Bogdan, "Using cross-task classification for classifying workload levels in complex learning tasks," in *Proc. Humaine Assoc. Conf. Affect. Comput. Intell. Interact.*, Sep. 2013, pp. 876–881.
- [17] Y. Ke *et al.*, "Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression," *Int. J. Psychophysiol.*, vol. 98, no. 2, pp. 157–166, Nov. 2015.
- [18] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain adaptation techniques for EEG-based emotion recognition: A comparative study on two public datasets," *IEEE Trans. Cogn. Develop. Syst.*, vol. 11, no. 1, pp. 85–94, Mar. 2019.
- [19] W. Zheng and B. Lu, "Personalizing EEG-based affective models with transfer learning," in *Proc. IJCAI*, 2016, pp. 2732–2738.
- [20] P. Zhang, X. Wang, W. Zhang, and J. Chen, "Learning spatial-spectral-temporal EEG features with recurrent 3D convolutional neural networks for cross-task mental workload assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 1, pp. 31–42, Jan. 2019.
- [21] I. Kakkos *et al.*, "EEG fingerprints of task-independent mental workload discrimination," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 10, pp. 3824–3833, Oct. 2021.
- [22] D. Wu, Y. Xu, and B.-L. Lu, "Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016," *IEEE Trans. Cognit. Develop. Syst.*, early access, Jul. 7, 2020, doi: 10.1109/TCDS.2020.3007453.
- [23] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, "Domain adaptation via transfer component analysis," *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.
- [24] A. Gretton, K. M. Borgwardt, M. J. Rasch, B. Scholkopf, and A. J. Smola, "A kernel method for the two-sample problem," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2006.
- [25] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207.
- [26] J. Wang, Y. Chen, S. Hao, W. Feng, and Z. Shen, "Balanced distribution adaptation for transfer learning," in *Proc. IEEE Int. Conf. Data Mining (ICDM)*, Nov. 2017, pp. 1129–1134.
- [27] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer joint matching for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1410–1417.
- [28] R. N. Roy, S. Charbonnier, A. Campagne, and S. Bonnet, "Efficient mental workload estimation using task-independent EEG features," *J. Neural Eng.*, vol. 13, no. 2, Apr. 2016, Art. no. 026019.
- [29] W. Schneider, A. Eschman, and A. Zuccolotto, *E-Prime Reference Guide*. Pittsburgh, PA, USA: Psychology Software Tools, 2002.
- [30] A. Delorme and S. Makeig, "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis," *J. Neurosci. Methods*, vol. 134, no. 1, pp. 9–21, Mar. 2004.
- [31] A. Mognon, J. Jovicich, L. Bruzzone, and M. Buiatti, "ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features," *Psychophysiology*, vol. 48, no. 2, pp. 229–240, Feb. 2011.
- [32] L. Pion-Tonachini, K. Kreutz-Delgado, and S. Makeig, "ICLabel: An automated electroencephalographic independent component classifier, dataset, and website," *NeuroImage*, vol. 198, pp. 181–197, Sep. 2019.
- [33] A.-M. Brouwer, M. A. Hogervorst, J. B. F. van Erp, T. Heffelaar, P. H. Zimmerman, and R. Oostenveld, "Estimating workload using EEG spectral power and ERPs in the n-back task," *J. Neural Eng.*, vol. 9, no. 4, Aug. 2012, Art. no. 045008.
- [34] Z. Yao *et al.*, "Altered dynamic functional connectivity in weakly-connected state in major depressive disorder," *Clin. Neurophysiol.*, vol. 130, no. 11, pp. 2096–2104, Nov. 2019.
- [35] B. He, L. Astolfi, P. A. Valdés-Sosa, D. Marinazzo, S. O. Palva, C.-G. Bénar, C. M. Michel, and T. Koenig, "Electrophysiological brain connectivity: Theory and implementation," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 7, pp. 2115–2137, Jul. 2019.
- [36] M. K. Kiymik, I. Güler, A. Dizibüyük, and M. Akin, "Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application," *Comput. Biol. Med.*, vol. 35, no. 7, pp. 603–616, Oct. 2005.
- [37] Z. Zhu, Z. Zhen, X. Wu, and S. Li, "Estimating functional connectivity by integration of inherent brain function activity pattern priors," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 18, no. 6, pp. 2420–2430, Nov. 2021.
- [38] Y. Zhou, L. Qiao, W. Li, L. Zhang, and D. Shen, "Simultaneous estimation of Low- and high-order functional connectivity for identifying mild cognitive impairment," *Frontiers Neuroinform.*, vol. 12, p. 3, Feb. 2018.
- [39] Y. Zhou, L. Zhang, S. Teng, L. Qiao, and D. Shen, "Improving sparsity and modularity of high-order functional connectivity networks for MCI and ASD identification," *Frontiers Neurosci.*, vol. 12, p. 959, Dec. 2018.
- [40] J.-P. Lachaux *et al.*, "Estimating the time-course of coherence between single-trial brain signals: An introduction to wavelet coherence," *Neurophysiologie Clinique/Clin. Neurophysiol.*, vol. 32, no. 3, pp. 157–174, Jun. 2002.
- [41] R. Li *et al.*, "The perils and pitfalls of block design for EEG classification experiments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 1, pp. 316–333, Jan. 2021.
- [42] S. Makeig, "Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones," *Electroencephalogr. Clin. Neurophysiol.*, vol. 86, no. 4, pp. 283–321, 1993.
- [43] W. Yi *et al.*, "Enhancing performance of a motor imagery based brain-computer interface by incorporating electrical stimulation-induced SSSEP," *J. Neural Eng.*, vol. 14, no. 2, Apr. 2017, Art. no. 026002.
- [44] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
- [45] Z. Wan, R. Yang, M. Huang, N. Zeng, and X. Liu, "A review on transfer learning in EEG signal analysis," *Neurocomputing*, vol. 421, pp. 1–14, Jan. 2021.
- [46] W. Wei, S. Qiu, X. Ma, D. Li, B. Wang, and H. He, "Reducing calibration efforts in RSVP tasks with multi-source adversarial domain adaptation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 11, pp. 2344–2355, Nov. 2020.