A Dual-Adversarial Model for Cross-Time and Cross-Subject Cognitive Workload Decoding

Yang Shao , Yueying Zhou , *Member, IEEE*, Peiliang Gong , Qianru Sun,
and Daoqiang Zhang , *Senior Member, IEEE*

Abstract—**Electroencephalogram (EEG) signals are widely utilized in the field of cognitive workload decoding (CWD). However, when the recognition scenario is shifted from subject-dependent to subject-independent or spans a long period, the accuracy of CWD deteriorates significantly. Current solutions are either dependent on extensive training datasets or fail to maintain clear distinctions between categories, additionally lacking a robust feature extraction mechanism. In this paper, we tackle these issues by proposing a Bi-Classifier Joint Domain Adaptation (BCJDA) model for EEG-based cross-time and cross-subject CWD. Specifically, the model consists of a feature extractor, a domain discriminator, and a Bi-Classifier, containing two sets of adversarial processes for domain-wise alignment and class-wise alignment. In the adversarial domain adaptation, the feature extractor is forced to learn the common domain features deliberately. The Bi-Classifier also fosters the feature extractor to retain the category discrepancies of the unlabeled domain, so that its classification boundary is consistent with the labeled domain. Furthermore, different adversarial distance functions of the Bi-Classifier are adopted and evaluated in this model. We conduct classification experiments on a publicly available BCI competition dataset for recognizing low, medium, and high cognitive workload levels. The experimental results demonstrate that our proposed BCJDA model based on cross-gradient difference maximization achieves the best performance.**

Index Terms—**Cognitive workload decoding, electroencephalogram (EEG), joint domain adaptation, adversarial learning, cross-time, cross-subject.**

I. INTRODUCTION

COGNITIVE workload is a critical indicator of the mental effort that humans spontaneously exert in a specific

task [1], [2] and has gained vast attention. There is, however, still no powerful consensus [3] about cognitive workload, and it can be the juxtaposition of the following definitions: Welford [4] defined cognitive workload as the resources available to meet task demands; Wickens argued that it is the relationship between the mental resources required by the task and the resources available to the operator [5]. Young and Stanton suggested that cognitive workload reflects the attention resources to meet both subjective and objective performance criteria [6]. Generally, the cognitive workload is distinguished from low and high levels [7], [8] and some work [9], [10] also classifies it into three more precise and practical categories, such as underload, normal, and overload. Due to the limited mental resources of humans, maintaining a moderate cognitive workload helps operators perform their work safely and efficiently [11]. Thus, cognitive workload decoding (CWD) is proposed to automatically judge the operator's mental state based on physiological signals. In recent years, CWD has been widely used in various industries, including education (online course evaluation [12]), medical (sequelae of rehabilitation [13]), transportation (car driving [8]), and aerospace [2], [14].

Multiple methodologies that utilize physiological signals have been proposed for qualitative cognitive workload levels [15]. Among many physiological signals, EEG is directly collected from the scalp and is proven to correlate with the CWD [16]. EEG reflects the activity of brain neurons and has the unique advantages of high temporal resolution, low cost, and easy acquisition [7], [17], which foster its wide use in CWD. Moreover, EEG signals are not easily camouflaged and can provide reliable results [18]. As a key neurofeedback application, EEG-based CWD aids in managing mental states effectively with its non-invasive, harmless, and enduring qualities [19]. Thus, we emphasize an EEG-centric perspective, focusing on the identification of reliable biomarkers that correlate with cognitive workloads.

However, the low signal-to-noise ratio makes EEG susceptible to noise interference. With the discrepancy between different times, subjects, and tasks, the EEG signal representation is unstable [20]. Therefore, in the field of EEG-based CWD, it remains a great challenge to obtain a general model that can be shared across different times and subjects (also called cross-time and cross-subject CWD problems). One solution to re-collect the corresponding data due to subject and time calibration before each test is impractical because
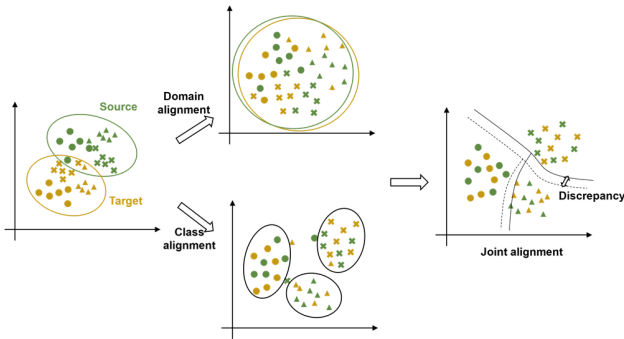
Fig. 1. Illustration of joint alignment in our proposed model. Domain-wise alignment and Class-wise alignment are conducted on two flows of adversarial learning process. The three shapes represent the three states of cognitive workload in our experiments, respectively.

the pre-acquisition of EEG data is time-consuming [17], [21], which undoubtedly brings great trouble and fatigue to users. Consequently, the challenge lies in optimizing the utilization of constrained datasets, a pivotal issue that necessitates resolution [22]. This issue arouses public attention to the cross-time and cross-subject CWD problem, which facilitates the migration of existing data, and this is the difficulty our paper focuses on.

In subject-dependent and time-dependent CWD, some machine learning [8], [10], [23] and deep learning methods [11], [24] have been proposed and achieved satisfactory results. However, they cannot adapt well to the cross-subject and cross-time CWD scenarios since they did not take the distribution differences between domains into account. Furthermore, domain generalization and domain adaptation methods are exploited to find shared label-common features (discussed in section II) by domain alignment. But they fail to preserve the class difference information simultaneously, as deep domain adaptation tends to smooth out the feature gullies, inadvertently resulting in a reduction in classification accuracy.

To address these issues, we propose a Bi-Classifier Joint Domain Adaptation (BCJDA) model for EEG-based cross-time and cross-subject CWD, containing not only adversarial domain adaptation but also adversarial inter-class distances of a Bi-Classifier, as shown in Fig. 1. The BCJDA is committed to extracting domain-invariant features through domain adaptation while maximizing the discriminant distance of two classifiers. This approach is aimed at maintaining the alignment of classes to revitalize the reliability of boundary samples. Different from the existing models, our proposed BCJDA takes raw EEG data as input to perform an end-to-end training process and considers both domain alignment and category alignment. This integration is achieved through adversarial training, which is further enhanced by a comparative analysis using three distinct disparity algorithms. Our contributions are as follows:

1) We design a novel CWD model based on Bi-Classifier and domain adaptation, alleviating the problem of class difference loss in deep domain adaptation models.
2) We investigate the effect of using a task-specific Bi-Classifier with different adversarial disparity algorithms

on the model performance. To our knowledge, we are the first to evaluate the influence of different determinacy disparity between classifiers in CWD.
3) We evaluate the BCJDA model on a public EEG dataset to recognize the practical low, medium, and high workload levels, demonstrating significant improvement over baselines.

The rest of this paper is structured as follows. Section II briefly reviews some work related to cross-time and cross-subject CWR. Section III introduces the structure and implementation of the proposed model in detail. Section IV presents the experimental setup and preliminary analysis of the experimental results. The resulting discussion is given in Section V. Finally, we conclude the paper and suggest some directions for future work.

II. RELATED WORKS

A. CWD With Machine Learning Methods

Traditional machine learning methods, such as support vector machine (SVM) [10], [23], K-nearest neighbor [23], [25], and linear discriminant analysis [26], need to manually extract features in advance and construct classification models through statistical methods. Although traditional machine learning methods are simple to implement and easy to train, they have reached the bottleneck due to the input of incomplete features and weak fitting ability for EEG samples that are susceptible to noise. Researchers are more inclined to exploit deep learning methods for EEG-based classification tasks because deep learning methods have more powerful representation capabilities [24] and realize end-to-end training, eliminating the trouble of manual feature extraction. Assuming that different sessions have a set of domain-invariant features, Jin et al. [27] achieved subject-specific CWD across time with a deep separable convolutional neural network (CNN) based on transfer learning; Kuanar et al. [28] presented a deep recurrent neural network (RNN) to learn robust cognitive workload features from intersubjective differences. Ni et al. [29] used an adversarial EEG generation method combined with hierarchical RNN to alleviate the performance degradation problem of event-related-potential-based BCI on cross-subject applications. As a review in [30], the CNN module is one of the most commonly used modules in the field of CWD, and we also adopt it in our proposed method.

B. CWD With Domain Adaptation Methods

Domain adaptation aims to transfer knowledge from the source domain to the unlabeled target domain and is suitable for knowledge transfer with a small number of samples [31], just like the EEG dataset. Domain adaptation can be a remedy for alleviating the nonstationarity of EEG. In cross-time CWD, the first session of a single subject is regarded as the source domain and the second session is regarded as the target domain. In cross-subject CWD, subjects whose EEG data have been acquired are classified as the source domain, and unknown subjects are classified as the target domain. Currently, domain adaptation has been applied to various other EEG-based state monitoring tasks to effectively

address the problem of excessive inter-domain differences. Furukawa et al. [32] used a small amount of estimated target data in emotion recognition to relieve the EEG measurement burden and proposed a model with multiple domain discriminators in their subsequent work [33]. For medical applications, a model utilizing hierarchical domain adaptation with projective dictionary pair learning [34] was used in epilepsy diagnosis for medical IoT for integrating pathological data from different nodes. Chambon et al. [35] and Heremans et al. [36] also explored the influence of domain adaptation on sleep state recognition. However, its application in the field of CWD is not sufficient. For example, Zhou et al. [15] proposed to use of domain adaptation approaches combined with machine learning models for binary-classification cross-task CWD, followed by a work [22] that constructs a CWD model with a deep neural network based on adversarial domain adaptation. These methods employ traditional machine learning techniques or fully connected layers for feature extraction. However, they inadequately exploit the deep temporal information of EEG signals and exhibit limited capacity in feature extraction. Moreover, none of these methods address the issue of category discrepancy loss arising from domain adaptation.

To address the aforementioned challenges, we propose BCJDA for cross-time and cross-subject CWD to boost general performance. Several CNN and pooling modules, which have a higher tolerance to noise, compose the feature extractor to enhance the automatic feature extraction ability that can obtain spatio-temporal features, reduce the dimensionality, and simplify the computational complexity. The BCJDA leverages the adversarial learning between the feature extractor and the domain discriminator to acquire the domain-invariant features, and the class difference is maintained by the disparity maximization of the Bi-Classifier. Through jointly aligning the domains and the classes, the domain distribution difference is diminished while more category feature differences are kept, and the BCJDA model is compelled to produce more robust output results.

III. METHODS

Since domain distribution and class discrepancy knowledge are equally important for EEG classification, we adopt joint alignment including domain-wise alignment and class-wise alignment to reduce inter-domain differences while retaining the class differences to the greatest extent. Thus, the classification performance of BCJDA is improved. This part mainly includes the following parts to illustrate our proposed model: 1) we will explain the overall structure of our proposed model; 2) the detailed structure and loss function of each module of the model are given; 3) two groups of adversarial methods and the overall optimization function are introduced, and three classifier adversarial losses used by our model are given.

A. Our Model Structure

Our model design draws inspiration from some existing domain adaptation methods, such as GANS, MCD [37], and BCDA [38]. There are three main parts in BCJDA: one feature extractor, one Bi-Classifier consisting of two task-independent

TABLE I
SUMMARY OF THE LAYER STRUCTURE USED IN OUR PROPOSED BCJDA MODEL WITHOUT DROP & BATCH-NORM LAYERS

| Module parts	Layer	Quantity	Parameters
Feature extractor	1×10 CNN	375	263125
	61×1 CNN	25	38150
	1×3 Max-pooling	375	1500
Bi-Classifier (Input:200)	FC (with 128 units)	2	51456
	FC (with 128 units)	2	33024
	FC (with 3 units)	2	516
Domain discriminator (Input:200)	FC(with 1024 units)	1	205824
	FC(with 1024 units)	1	1049600
	FC(with 1 units)	1	1025
Total	--	784	1644220

CNN stands for convolutional layer and FC stands for fully connected layer.

label classifiers, and one domain discriminator, as shown in Fig. 2. Both the domain discriminator and the Bi-Classifier obtain features from the feature extractor. The Bi-Classifier further extracts deep features and then predicts the cognitive workload level. It aims to minimize prediction errors with ground truth on source domain samples while maximizing prediction discrepancies on target domain samples to preserve category differences (inconsistency loss in Fig. 2). The domain discriminator, simultaneously, tries to distinguish features from source domain or target domain by minimizing domain classification errors. By contrast, the feature extractor attempts to deceive both the Bi-Classifier and the domain discriminator, thus forming two sets of minimax games which will be detailed discussed in part D. Therefore, the feature extractor can gradually generate domain-invariant features that maintain the differences between categories. The Bi-Classifier performs deep extraction on the features and learns label-specific features during the process of adversarial learning. Finally, our model's prediction results are determined by the superposition of the two classifiers.

The detailed layer structure and parameters are summarized in Table I. It describes the layers used by each module of the model and the number of parameters for different layers excluding dropout and batch-norm layers. Table SI in supplementary materials compares our proposed model's modules with other common CNN methods in section IV part D. Compared with common CNN methods (i.e. EEGNet and DeepCNN) in EEG classification tasks, our model introduces only minimal increases in parameters and computational complexity with two additional classifiers.

B. Feature Extractor

EEG data is small in size and highly sensitive, making it unsuitable for general deep network models, unlike other categories such as images and speech. We explored various common EEG feature extraction models and ultimately devised the feature extractor depicted in Fig. 3, drawing inspiration from DeepCNN, ShallowCNN, and Multi-branch 3D
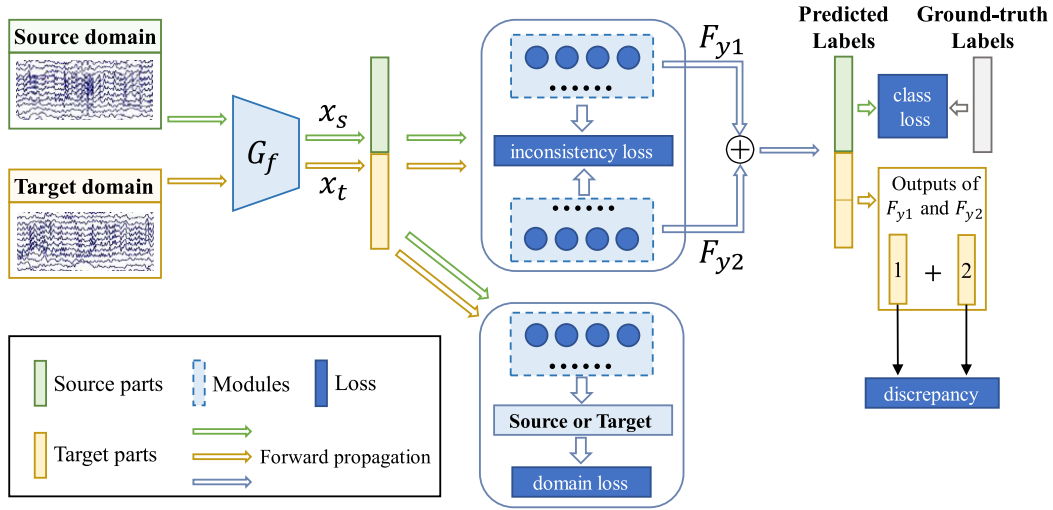
Fig. 2. The architecture overview of Bi-Classifier Joint Domain Adaptation (BCJDA), where green indicates the source domain and yellow indicates the target domain; G_f is the feature extractor; Bi-Classifier is composed of F_{y1} and F_{y2}, which are Classifier 1 and Classifier 2 respectively; x_s and x_t are the features of the source domain and target domain extracted by G_f. The inconsistency loss is specifically expressed as the degree of difference between the deep feature matrices generated by the two classifiers for the target domain samples. The corresponding numbers 1 and 2 on the target output label indicate which classifier the label is given by.
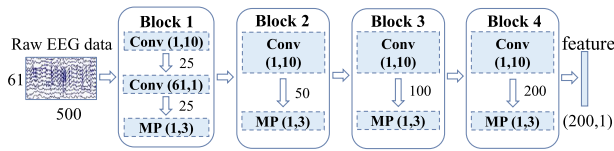
Fig. 3. Structure of feature extractor in BCJDA. The input is the raw 61×500 (Channel×Time) EEG data, and the output is the 200×1 feature map.

CNN. Some of the subject-dependent pre-experimental results are placed in Table SII in the supplementary materials to support our choice. Raw EEG data is fed into the feature extractor and output in the form of a 200×1 feature map after passing through four blocks. Block 1 includes a 1×10 temporal filter and a 61×1 spatial filter, where the latter fuses each channel with different weights, and the channel dimension is compressed to one. Block 2 - block 4 just utilize the 1×10 temporal filter respectively, capturing the local and global timing information of EEG signals. Max pooling layer is adopted in all 4 blocks to reduce the complexity of computation while effectively retaining EEG feature information and mitigating noises in consideration of the low signal-to-noise ratio and non-stationarity of EEG.

To minimize the misclassification and guarantee the effectiveness of the model's classification ability, it is necessary to minimize the training loss in annotated data in source domain. Here we take F_y as a classifier and take θ_y, θ_f as parameters of the classifier and feature extractor. The loss function is formulated as L_{class}:

$$L_{class}\left(\theta_f, \theta_y\right) = L_y\left(p\left(l_k | F_y\left(G_f\left(x_s; \theta_f\right), \theta_y\right)\right), y_s\right). \quad (1)$$

G_f and x_s represent the feature extractor and source domain samples respectively, as shown in Fig. 2. y_s denotes the label of the sample x_s. p denotes the softmax function that turns the classifier output into a probability for a given label l_k. L_y means the label prediction loss, also called supervised loss.

C. Bi-Classifier & Domain Discriminator

The Bi-Classifier consists of two task-specific classifiers (F_{y1}, F_{y2}) with the same structure, which is a network with three fully connected layers. The input layer and hidden layer each have 128 neurons and the output layer has 3 neurons (same as the number of classes). During the testing process, the classifiers cooperate to determine the final result. Yet in training, in addition to the supervised loss determined by Eq. (1), the confrontation between the classifiers is also considered to maximize the distance of the Bi-Classifier. Here, we modify Eq. (1) to give the supervised loss function in the form of the Bi-Classifier:

$$L_{class}\left(\theta_f, \theta_{y1}, \theta_{y2}\right)$$
$$= \frac{1}{2n_s}\sum_{j=1}^{2}\sum_{i=1}^{n_s} L_y\left(p\left(l_k | F_{yj}\left(G_f\left(x_{si}; \theta_f\right), \theta_{yj}\right)\right), y_{si}\right). \quad (2)$$

Here, n_s represents the number of source domain samples, x_{si} represents a source domain sample and y_{si} is its corresponding label. The domain discriminator is a special classifier consisting of three fully connected layers where the input layer, hidden layer, and output layer contain 1024, 1024, and 1 neuron respectively. As shown in Fig. 2, it takes features from the feature extractor and tries to accurately identify the domain each feature group belongs to. Instead, the feature generator is trained to mislead the domain discriminator (G_d). To achieve the above purpose, the domain discriminator and the feature generator are jointly optimized to perform domain alignment with parameters θ_d and θ_f. The Loss function of L_{domain} is given:

$$L_{domain}\left(\theta_f, \theta_d\right)$$
$$= -\frac{1}{n}\sum_{i=1}^{n} L_d\left(p\left(l_k | G_d\left(G_f\left(x_i; \theta_f\right), \theta_d\right)\right), d_i\right). \quad (3)$$
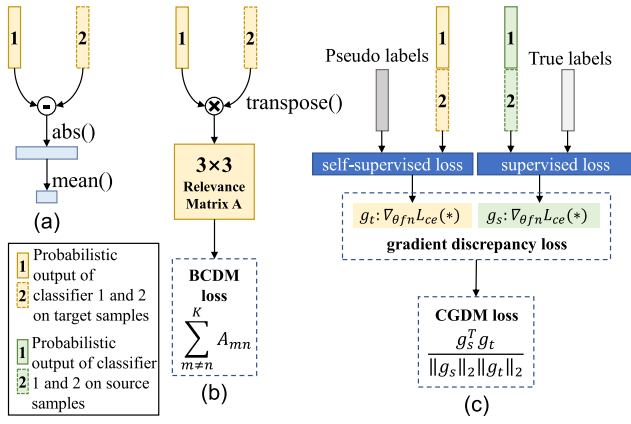
Fig. 4. Determinacy disparity of Bi-Classifier in BCJDA. Part (a) is the L1-norm calculation method, part (b) is the BCDM difference calculation method and part (c) is the CGDM difference calculation method.

In the formula, $n = n_s + n_t$, which is the total number of samples in a group. d_i corresponds to the label of the sample x_i. L_d means the label prediction loss in the domain discriminator.

D. Joint Domain Adaptation

This part mainly introduces two sets of adversarial processes in training, namely the game between the feature generator and the domain discriminator to implement domain-level alignment, and the game between the feature generator and the Bi-Classifier to implement class-level alignment.

For domain-level alignment, the loss function includes the supervised loss and the domain classification loss, as shown in Eq. (2) and Eq. (3). The overall loss function with weight coefficient λ at the domain level is denoted as:

$$L_0\left(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d\right) = L_{class}\left(\theta_f, \theta_{y1}, \theta_{y2}\right) + \lambda L_{domain}\left(\theta_f, \theta_d\right). \quad (4)$$

For class-level alignment, we selected three algorithms of determinacy disparity between two classifiers. They are L1-norm, Bi-Classifier Determinacy Maximization (BCDM) [38], and Cross-Domain Gradient Discrepancy Minimization (CGDM) [39]. Figure. 4 shows their calculation methods.

L1-norm is easily calculated by taking the absolute value of the difference between the outputs and averaging them over all the samples. BCDM method calculates the difference using the relevance matrix obtained by multiplying the two resulting vectors as shown in Fig. 4(b). Specifically, the CGDM method uses the recognition results of the source domain samples as in Fig. 4(c). It uses the ground truth and source domain samples to calculate the supervised loss and uses the model in training to pseudo-label the target samples to obtain the self-supervised loss. We then take the partial derivatives of the two losses with respect to the parameters; the partial derivatives g_t, g_s of the parameters θ_{y1}, θ_{y2} for the two losses are obtained. Finally, gradient discrepancy loss is derived by the above g_t, g_s. The formulas of the L1-norm, BCDM, and CGDM are given below in order, denoted as L_{dis0}, L_{dis1}, L_{dis2}:

$$L_{dis0} = \|p_t^{y1} - p_t^{y2}\|_1, \quad (5)$$

$$L_{dis1} = \sum_{m \neq n}^{3} \left(p_t^{y1} T\left(p_t^{y2}\right)\right)_{mn}, \quad (6)$$

$$L_{dis2} = 1 - \frac{g_s^T g_t}{\|g_s\|_2 \|g_t\|_2}. \quad (7)$$

Here, p_t^{yi} represents the prediction probability matrix of classifier $i \in \{1, 2\}$ for target domain samples, and $T(*)$ represents the transpose function.

In EEG recognition problems, the division of boundary samples is very important, especially when adopting cross-domain methods since the category discrepancy will be lost to some extent when the source domain knowledge is transferred to the target and the instability of EEG signals. And class-level alignment can mitigate the above issues so we hope the Bi-Classifier is able to discern target domain samples close to the classification boundary and force the feature generator to generate features that preserve significant category discrepancy. Therefore, the overall class-level loss function is as follows:

$$L_1\left(\theta_f, \theta_{y1}, \theta_{y2}\right)$$
$$= \frac{1}{n_t} \sum_{i=1}^{n_t} L_{dis}\left(F_{y1}\left(G_f\left(x_{ti}; \theta_f\right), \theta_{y1}\right), \right.$$
$$\left. F_{y2}\left(G_f\left(x_{ti}; \theta_f\right), \theta_{y2}\right)\right), \quad (8)$$

where L_{dis} is chosen from $\{L_{dis0}, L_{dis1}, L_{dis2}\}$.

Our model is trained with the back-propagation of two adversarial losses and the overall loss function is obtained as L:

$$L\left(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d\right) = L_0\left(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d\right) + \lambda L_1\left(\theta_f, \theta_{y1}, \theta_{y2}\right). \quad (9)$$

To reduce the discrepancy of the distribution across domains thus enabling the feature generator to learn domain-invariant features, the parameter θ_d is determined to minimize $-L_{domain}$, that is, to maximize L_0 and the parameters $\theta_f, \theta_{y1}, \theta_{y2}$ are determined to maximize L_{class}, that is, to maximize L_0 too. In this case, the domain discriminator will try its best to tell which domain the sample belongs to and the feature extractor will seek shared knowledge. We summarize the optimization objective function of the above process as:

$$\max_{\theta_d} L\left(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d; x\right),$$
$$\min_{\theta_f} L\left(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d; x\right), \quad (10)$$

in which x belongs to $\{x_s, x_t\}$. That means Eq. (10) is optimized for all samples.

To alleviate the smoothness of discriminativeness in EEG signals under domain adaptation and align sample partitions of the source and target domains, the purpose of the Bi-Classifier is to maximize its discriminative difference concerning the target domain and enforce the feature extractor to yield features with clear category discrepancy. Besides, minimizing the supervised loss on the source domain is also taken into account. Thus, the parameter θ_{y1} and θ_{y2} is optimized to maximize L_1 and minimize L_{class}, that is to maximize and minimize L respectively in source and target samples.

We summarize the optimization objective function of the above process as:

$$\min_{\theta_{y1}, \theta_{y2}} L\left(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d; x_s\right),$$
$$\max_{\theta_{y1}, \theta_{y2}} L\left(\theta_f, \theta_{y1}, \theta_{y2}, \theta_d; x_t\right). \quad (11)$$

It should be noted that although Eq. (11) may appear contradictory, it operates on separate samples, allowing for the use of the same loss function during training with a gradient reverse layer. Through Eq. (9), (10), and (11), once BCJDA is well trained, it can perform joint alignment of EEG features across domains.

IV. EXPERIMENTS AND RESULTS

A. Dataset

The public cognitive workload dataset released by the Neuroergonomics Conference is used in our experiments. We first illustrate the components of this dataset and then describe the preprocessing steps used. More details about the dataset can be found in *www.neuroergonomicsconference.um.ifi.lmu.de*.

The dataset contains a total of 15 subjects, 6 females and 9 males, with an average age of 25 years. Each subject was invited to the lab for three independent sessions, one week apart (exactly seven days apart). Each experiment session had a short warm-up period. After that, the EEG data of the subject's resting state were recorded. Participants then completed a multi-attribute task battery (MATB) divided into three five-minute modules, each of which presented a different level of difficulty (that is, a different level of workload; see Table SIII of the supplementary materials for details) in a pseudo-random manner. 3 levels of workload were elicited by varying the number and complexity of subtasks which were validated by statistical analysis of subjective and objective behavioral and cardiac data. All EEG data were recorded under the EEG cap of the international 10-20 system with 64 channels. In the final dataset, each subject has 3 sessions, and each session contains 447 cognitive workload samples, which are evenly divided into 3 categories. Raw EEG signals include a reference electrode, a cardiac activity electrode, and a bad lead, which were removed. We use the first two sessions with provided labels for model training and evaluation. Each sample has 500 sampling points (2s epoch, sampling rate 250Hz).

Data preprocessing is done by EEGlab, and the specific steps are as follows:

1) Data were divided into 2-second non-overlapping epochs, using the right mastoid electrode as a reference.
2) Used a high-pass filter (FIR filter, pop-filtnew from EEGlab) at 1Hz.
3) Used electrode suppression (mean amplitude above 2 sd across channels) with spherical interpolation.
4) Used SOBI with automatic IC_Label rejection (muscle, heart, and eye components are rejected with a threshold of 95%).
5) Used a 40Hz low-pass filter (FIR filter).
6) A common average reference method was used and downsampled to 250Hz.

B. Comparative Methods

SVM: SVM is a supervised machine learning algorithm that maps the data into a high-dimensional space by using different kernel functions to find an optimal hyperplane to segment the data into different classes.

A common feature for CWD using traditional machine learning methods is Power Spectral Density (PSD). We use the multi-taper method proposed by Thomson [40] to extract the PSD features of each channel and concatenate the features of all channels to form a feature vector as the input of SVM. This method is a non-parametric method of PSD estimation, which does not require any prior information about the signal generation process and has a low estimation variance.

EEGNet [41]: EEGNet is a compact CNN widely used in the field of EEG classification, which uses deep-separated convolutions to build an EEG-specific model with the ability to classify across BCI paradigms. It has a strong generalization ability when the training data is limited.

Multi-branch 3D CNN [42]: To extract the spatial-temporal features of EEG signals, MB3D employs a multi-branch 3D convolutional neural network (MB3D), where each branch has a different receptive field size and can capture EEG features at different scales. This approach fully exploits the 3D structure of EEG signals and enhances the classification accuracy and robustness.

DeepCNN [43]: DeepCNN (DCNN) uses multiple CNN modules as the main model structure and mines the task-related deep features of EEG signals through multi-level temporal feature extraction. It has been proven useful for spatially mapping the learned features.

ShallowCNN [43]: ShallowCNN (SCNN) is a simplified CNN that uses only one feature layer to extract the intrinsic features of EEG sequences, and then uses training time and accuracy to evaluate the classification effect. It has a small number of parameters and high real-time performance.

Joint Distribution Adaptation [44]: Joint Distribution Adaptation (JDA) is a domain adaptation method that strives to obtain latent representations where the source domain is similar to the target domain by aligning marginal distributions and conditional distributions. JDA uses domain discriminators and associative reinforcement to deal with shallow and deep features. PSD features similar to SVM are fed into this model.

C. Implementation Details

The experimental part comprises cross-time and cross-subject CWD experiments. In the cross-time scenario, session 1 of a single subject serves as the source domain and session 2 as the target domain. Since time is irreversible, we do not consider the reverse direction. We apply all eight aforementioned methods to each subject, and a total of 360 experiments are conducted on 15 subjects. We then take the average of all subjects as the final result. In the cross-subject scenario, we employ leave-one-subject-out (LOSO) cross-validation to evaluate model performance. This involves using one subject as the target domain while utilizing all other subjects as the source domain in each session. A total of 240 experiments were performed for these two sessions.

TABLE II
CROSS-TIME COGNITIVE WORKLOAD DECODING RESULTS OF ACCURACY (%) IN 15 SUBJECTS

Models	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	AVG
SVM	44.07	41.61	48.32	48.99	40.49	45.86	53.47	42.73	40.04	53.69	48.99	43.18	38.03	43.62	47.43	45.37±4.55
EEGNet	40.72	41.39	52.35	37.36	52.13	55.03	49.66	37.58	57.49	46.67	53.91	36.24	46.89	51.68	49.22	47.22±6.73
MB3D	41.83	53.24	57.94	43.85	48.32	46.31	59.28	52.57	62.42	45.41	52.80	42.28	48.32	45.64	39.82	49.34±6.58
DCNN	39.37	42.06	48.32	36.24	55.03	60.40	47.20	62.19	66.89	48.77	55.48	39.60	39.37	63.31	44.07	49.89±9.69
SCNN	46.76	38.26	56.60	36.91	52.35	57.49	56.38	61.74	46.98	65.32	61.30	42.73	36.24	56.15	46.09	50.75±9.17
JDA	63.98	63.09	53.69	50.11	54.81	46.31	51.9	53.24	48.55	49.89	45.64	47.2	46.07	45.41	47.43	51.15±5.50
B. L1	57.49	68.01	67.11	70.25	64.65	65.10	51.90	78.97	79.19	67.79	55.48	51.90	54.81	62.42	58.61	63.58±8.39
B. BCDM	59.51	71.81	76.06	67.56	58.84	66.89	52.35	75.39	76.06	68.00	68.68	57.94	55.03	63.98	57.27	65.03±7.62
B. CGDM	61.52	77.63	74.94	69.35	60.18	69.57	54.14	81.21	80.09	66.22	59.51	56.38	58.84	67.11	59.51	**66.41±8.53**

The ones in italics are our methods. **B.** stands for BCJDA, followed by the Bi-Classifier adversarial loss used. For example, **B. L1** represents the BCJDA based on the L1 norm Bi Classifier loss function The optimal mean is bolded and the suboptimal mean is underscored

TABLE III
CROSS-SUBJECT COGNITIVE WORKLOAD DECODING RESULTS OF ACCURACY (%) IN 15 SUBJECTS IN SESSION 1

Models	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	AVG
SVM	60.18	49.22	42.28	52.80	33.33	52.35	50.11	47.65	49.66	37.81	56.60	47.87	41.61	49.44	46.53	47.83±6.68
EEGNet	54.36	54.81	44.07	56.60	45.19	57.05	53.02	45.86	42.06	46.98	34.68	49.44	42.73	41.16	45.86	47.59±6.26
MB3D	63.09	54.81	52.57	59.69	42.06	64.88	55.03	50.11	55.70	45.64	38.26	59.73	44.97	48.10	46.98	52.11±7.57
DCNN	63.61	44.07	42.28	49.22	45.41	66.22	53.91	43.40	43.40	42.06	35.12	65.77	46.98	41.16	42.73	48.36±9.31
SCNN	63.98	40.72	48.77	49.66	46.76	61.30	44.07	59.51	59.28	47.43	33.56	73.38	44.74	39.82	46.98	50.66±10.3
JDA	61.07	50.34	51.90	53.24	41.16	57.94	52.80	39.60	49.22	46.31	41.83	52.13	42.28	40.27	53.91	48.93±6.28
B. L1	70.25	66.67	76.51	68.68	58.39	65.77	64.43	69.35	74.72	63.53	59.28	63.76	58.84	56.60	58.84	**65.04±5.87**
B. BCDM	70.92	68.90	72.26	67.11	59.28	67.79	64.88	65.77	70.47	64.65	57.05	65.32	60.40	59.96	60.40	65.01±4.54
B. CGDM	70.89	69.48	70.11	65.05	59.20	67.66	64.02	65.67	70.44	63.50	57.05	64.49	59.78	59.68	59.07	64.41±4.79

The ones in italics are our methods. **B.** stands for BCJDA, followed by the Bi-Classifier adversarial loss used. For example, **B. L1** represents the BCJDA based on the L1-norm Bi-Classifier loss function. The optimal mean is bolded, and the suboptimal mean is underscored.

TABLE IV
CROSS-SUBJECT COGNITIVE WORKLOAD DECODING RESULTS OF ACCURACY (%) IN 15 SUBJECTS IN SESSION 2

Models	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13	S14	S15	AVG
SVM	51.45	47.65	38.93	47.20	45.86	49.44	51.01	41.61	51.68	33.11	55.70	44.74	38.70	45.41	48.32	46.06±5.73
EEGNet	44.52	52.80	47.43	53.47	49.22	44.97	54.14	47.87	44.97	46.98	48.32	52.13	41.16	39.82	48.55	47.76±4.12
MB3D	49.98	50.34	51.45	44.07	40.49	56.38	55.93	41.83	57.49	47.87	33.78	50.78	37.14	42.06	45.86	47.03±6.87
DCNN	45.86	41.83	56.15	51.01	48.55	45.86	48.32	45.41	42.95	50.78	38.70	48.55	52.35	52.57	61.07	48.66±5.50
SCNN	45.19	53.69	45.86	47.87	41.61	45.19	53.47	43.85	51.68	56.06	33.33	47.20	50.34	38.48	55.03	47.26±6.18
JDA	47.87	48.99	44.74	46.31	45.41	45.41	52.80	42.51	45.19	44.97	50.11	45.19	46.76	43.40	51.00	46.71±2.80
B. L1	62.19	61.97	61.75	66.67	63.31	62.86	66.89	69.13	69.13	66.67	54.14	62.19	60.85	59.96	62.86	63.37±3.76
B. BCDM	59.51	64.88	61.07	68.24	62.42	68.46	66.89	66.44	69.13	65.32	58.39	62.64	60.18	57.49	63.31	63.62±3.66
B. CGDM	60.85	65.32	61.97	69.13	65.77	68.90	67.79	71.14	72.26	70.25	57.94	63.53	59.73	57.27	63.09	**65.00±4.68**

The ones in italics are our methods. **B.** stands for BCJDA, followed by the Bi-Classifier adversarial loss used. For example, **B. L1** represents the BCJDA based on the L1-norm Bi-Classifier loss function. The optimal mean is bolded, and the suboptimal mean is underscored.

During training, epochs are set to 10. Batch size is set to 32 in the CNN methods and 48 in JDA and our proposed method. The learning rate is set to 0.003 in the CNN methods and 0.005 in JDA and our proposed methods, and a learning rate decay strategy is adopted. CrossEntropyLoss is used for training loss and Adam and SGD is selected as the optimizer in CNN methods and ours respectively. For the weight parameter λ in Eq. (4) and (9), if it is set too high, the model will have difficulty learning the class differences. If it is set too low, the source domain data will not be aligned well with the target domain, consequently reducing the classification performance. Therefore, in L1-based BCJDA, we set the weights of L_{domain}, L_{dis0} losses to 1.0, whereas in the other two BCJDAs, we set the weight of L_{dis} loss to 0.01.

In terms of evaluation metrics, we select accuracy, F1-score, sensitivity, and specificity as the model evaluation metrics. The macro-averaged F1-score results are given for 3 classes and sensitivity and specificity are given separately for each class.

We used the Linux operating system (Ubuntu 20.04) and an Nvidia RTX3090 graphics card with 24GB of video memory to train the model. Our model training speed is about 10.8 iters/s on average, and the training completion time (10 epochs) is around 138.6s.

D. Results

We implement our BCJDA model for validation on the dataset given in part A under the cross-time and the cross-subject scenarios. Tables II, III, and IV show the accuracy of recognition across time and subjects respectively. The highest mean value in the tables is shown in boldface, and the second highest mean value is underlined. The experimental settings are implemented as given in part C. The significance of this part will be discussed in section A part V.

For cross-time CWD scenario, we find that CGDM-based BCJDA achieved the best performance with an increase of about 1%-1.5% to L1-norm-based and BCDM-based BCJDA, and all our proposed methods outperform the compared methods. L1-norm-based BCJDA is about 18% higher than SVM and 13%-16% higher than deep learning methods and JDA, which demonstrates that our methods can effectively extract the domain-invariant features, and consider the alignment between categories while transferring the source domain knowledge to the target domain. Furthermore, we compare the accuracy, F1-score, sensitivity, specificity, and macro-averaged sensitivity and specificity of the methods across time in 15 subjects, as shown in Fig. S1 and Table SIV in supplementary materials. Take L1-norm-based BCJDA, for example, the method is 17%-22% ahead of the CNN methods in terms of metrics on F1. In addition, SEN is significantly improved in C2 (Normal) and C3 (Overload) categories, and SPE is significantly improved in C1 category. In general, our methods have few decreases in F1-score compared with ACC ($<$ 1%), whereas the other methods all have varying degrees of decrease ($>$ 3%). The well-known JDA method, which also belongs to domain adaptation, outperforms a bunch of CNN methods despite using only full connectivity. Our method improves about 14% relative to it.

For cross-subject CWD scenario, Table III shows the ACC results under session 1, and Table IV shows the ACC results under session 2. We observed that L1-norm-based BCJDA and CGDM-based BCJDA achieved the best performance in the two sessions with an increase about 17%-19% compared with CNN methods and JDA, respectively. Furthermore, we compare various metrics of the methods across subjects in 2 sessions, as shown in supplementary materials in Fig. S2 and Tables SIV and SVI. Fig. S2(a) and Table SIV give metrics in session 1 and Fig. S2(b) and Table SVI give metrics in session 2. In general, the F1-score of our methods is basically the same as ACC ($<$ 1%), whereas the other methods have varying degrees of decrease ($>$ 3%), further illustrating the advantage of our methods in CWD across subjects. From the comparison of SEN and SPE, the classification performance of our methods is significantly improved on C1 (Underloaded) and C2 (Normal).

V. DISCUSSION

A. Results Distribution

This section presents the statistical analysis of the ACC metrics and the significance test using the Friedman test, followed by the Nemenyi post-hoc test [18]. To further illustrate the difference of three determinacy disparity of Bi-Classifier in BCJDA, we conduct separate tests for the three methods we proposed, in addition to the unified test for all models.

Our models achieve state-of-the-art performance in all experimental scenarios, as shown by the significant differences ($p < 0.01$) from the other models. Specifically, in the cross-time scenario (Fig. 5(a)), the CGDM-based BCJDA differs significantly from the L1-norm-based BCJDA. In the cross-subject scenario of session 1(Fig. 5(b)), the CGDM-based BCJDA differs ($p < 0.05$) from the BCDM-based BCJDA. In the cross-subject scenario of session 2 (Fig. 5(c)), the
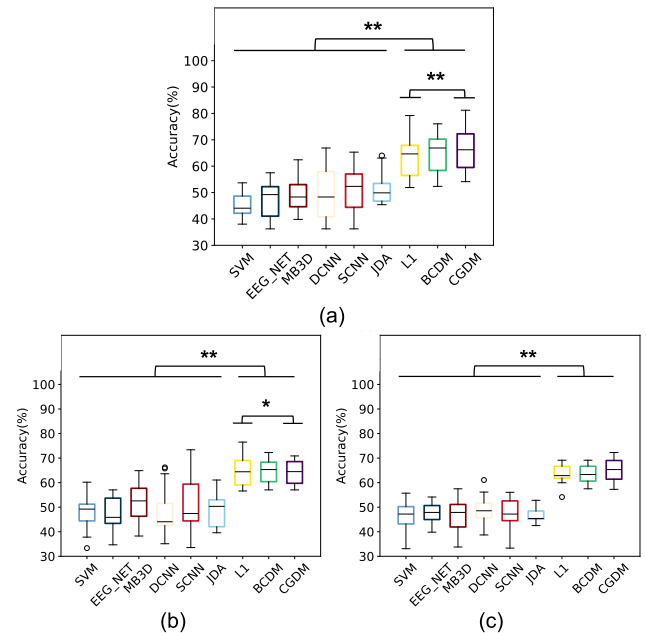

Fig. 5. Accuracy distributions and significance test between the proposed models and comparing methods. Friedman test is used; * represents $p < 0.05$ and ** represent $p < 0.01$ which means a great significance. (a) shows the results in cross-time experiment; (b) and (c) shows the results in session 1 and 2 of cross-subject experiments respectively.

CGDM-based BCJDA shows some advantages over the other two methods ($p < 0.1$), but they're not dramatically different. These results suggest that the type of inter-classifier determinacy disparity has a significant impact on the performance, and the CGDM-based BCJDA is superior to the other methods.

B. Confusion Matrix

To further analyze the misclassification rate of our proposed models on the three-class recognition of cognitive workload, we visualize the recognition accuracy using confusion matrices. Figure. 6 shows the confusion matrices of BCJDA under two experimental scenarios. For instance, in the cross-time scenario, the three BCJDAs achieve the highest recognition accuracy for the Underload state, which are 80.10%, 79.42%, and 82.51%, respectively.

The recognition accuracy of the Normal state is the lowest, and the Normal state is misclassified as Underload and Overload states with similar proportions, indicating that the Normal state is not well separated from the other two states. In the Overload state, about 33% of the samples are misclassified as Normal state. These results suggest that the Normal state has less distinctive features, partly because the subjects are fatigued when performing this part of the experiment, and the EEG signals generated are closer to the Overload state, even though the task difficulty is set to moderate. The samples classified as Overload by our model indicate that the subject is approaching the limit of cognitive resources while performing the task and needs special attention.

C. Analysis of EEG Topography

In the field of EEG classification, Power Spectral Density (PSD) is a common feature to distinguish between different
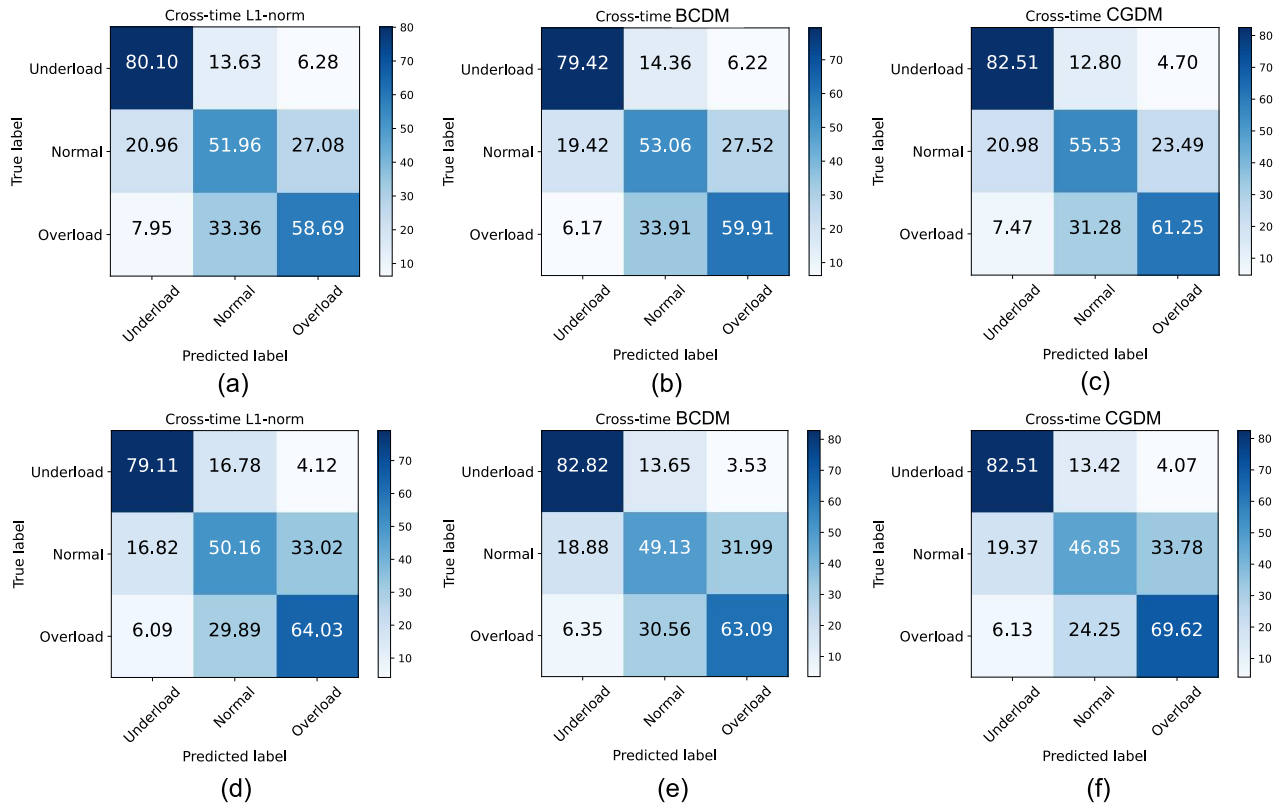
Fig. 6.   The accuracies (%) of BCJDA represented by confusion matrices. (a), (b), and (c) shows the results under the cross-time scenario; (d), (e), and (f) shows the results under the cross-subject scenario. The results displayed in the confusion matrix represent the aggregated classification outcomes across all subjects within a specific experimental scenario. For example, for (a), 80.1% of the Underload samples out of 15 subjects are predicted correctly, and 6.28% of the samples are misclassified to be in the Overload category.

cognitive workload levels. Although CNN and our methods do not directly employ PSD features and extract other features from raw EEG signals, PSD features are still implicit in these features. Therefore, it is meaningful to perform a visual analysis of PSD features. We visualized the PSD features of five common EEG frequency bands in two sessions of subject 7 (whose PSD exhibited the best classification ability) as topographic maps. Five frequency bands are: δ waves (1-3Hz), θ waves (4-7Hz), α waves (8-13Hz), β waves (14-30Hz), and γ waves (31-50Hz). Two-way ANOVA was used to test the correlation between PSD characteristics and Sessions and Workloads. The results are shown in Fig. 7.

We observe that the PSD features show similar properties, including the activated brain areas and the intensity of activation, in different domains under the same state. These demonstrate that PSD features are consistent in responding to the cognitive workload state, indicating that the joint alignment across domains is feasible. However, different Sessions exhibit a lower correlation with PSD features. Specifically, among the five frequency bands, only the α (alpha) band demonstrates a significant correlation ($p < 0.05$). This finding suggests that the variation between sessions is more pronounced than the differences related to Workload. Given these results, we emphasize the necessity and significance of pursuing cross-domain recognition.

In the same domain, we find that the PSD feature distributions of Underload state, Overload state, and Normal state are quite different and have stronger correlation with PSD;

Factors	δ	θ	α	β	γ
Session	0.063	0.127	0.042	0.203	0.527
Workload	0.272	0.181	0.024	0.016	0.054
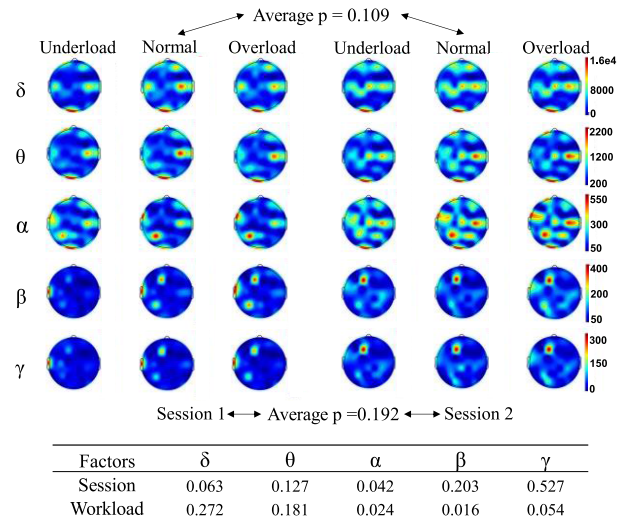
Correlation analysis results (p value) of Session and workload on PSD features. A higher p value represents a lower correlation, that is, a smaller difference.

Fig. 7.   PSD features and correlation analysis of subject 7 in five commonly used bands displayed in the form of EEG topography.

however, the PSD distributions of Normal and Overload state are closer, and the difference of power between Normal and Overload state is not obvious. This explains why models (both CNNs and our methods) are prone to misclassify on the Normal state. As shown in the confusion matrix and the SEN and SPE bar charts of Fig. 6 and Fig. 2 of the additional document,
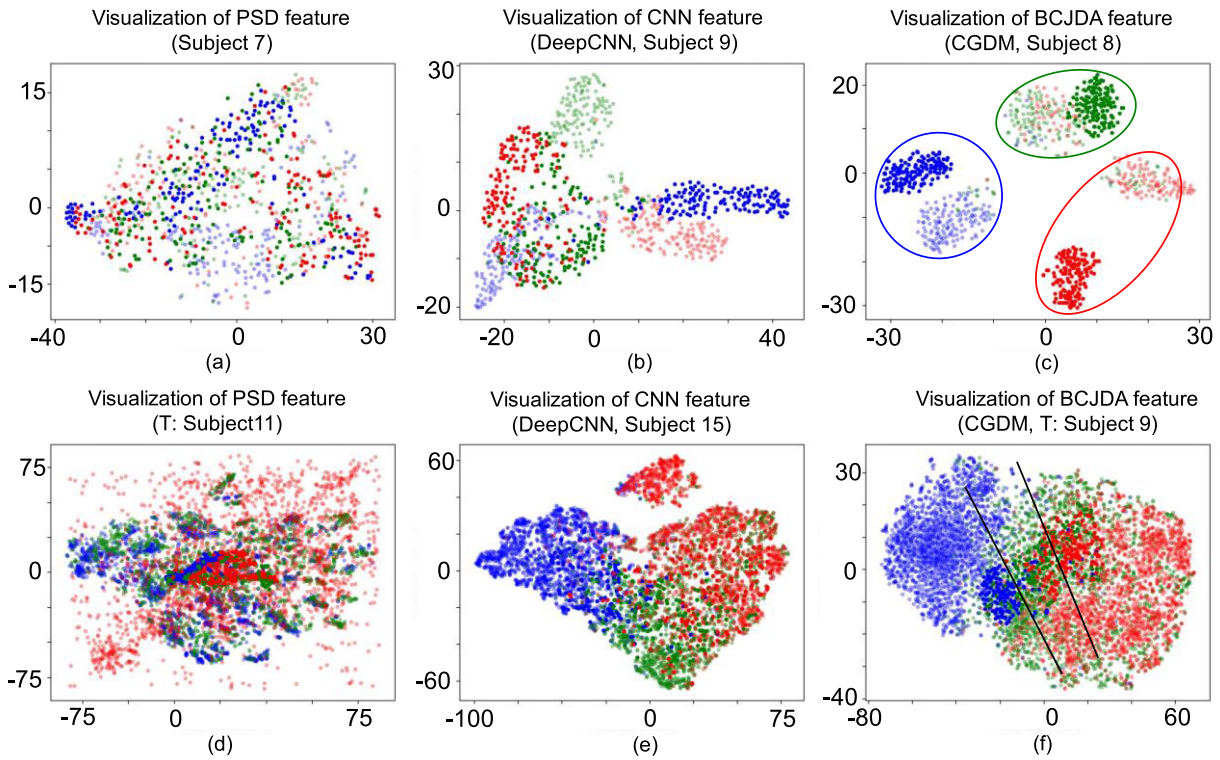
Fig. 8.   Visualization of EEG features extracted by models. Red dots represent overloaded samples, green dots represent normal samples, and blue dots represent underloaded samples. Semi-transparent points represent source domain samples, and opaque points represent target domain samples. (a), (b), and (c) are the cross-time visual features of recognition; (d), (e), and (f) are the cross-subject visual features of recognition in session 2. Model and subject details are annotated at the top of the subfigures (T means target domain).
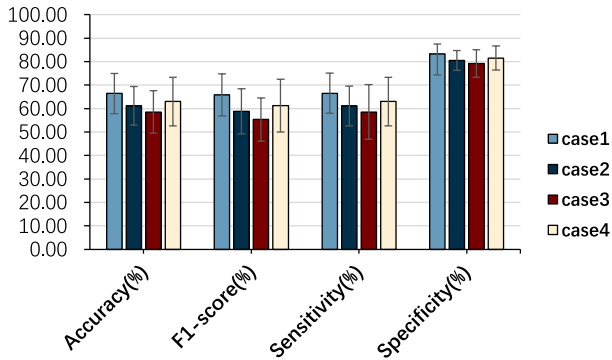

Fig. 9.   The impact of classifiers and domain discriminators on overall model performance under cross-time scenario.

the SEN and SPE of the classification on Underload are high, whereas the sensitivity and specificity of the classification on Normal are low, and they are mainly confused with Overload. In this case, although the subjects were performing moderately difficult tasks, the cognitive workload tended to be overloaded.

D. Feature Visualization

To further analyze the feature representation ability of SVM, CNNs, and BCJDAs, we use t-SNE [22] for visual analysis. The t-SNE is a statistical method for visualizing high-dimensional data by giving each data point a location in a two or three-dimensional map. Fig. 8 visualizes the distribution of features. For SVM, we use the PSD features

extracted beforehand. We find that the features extracted by the three types of models in the cross-time scenario have higher discrimination than in the cross-subject scenario.

Compared with the other two methods that deteriorate when the domain difference is large, CGDM-based BCJDA not only has a similar feature extraction ability as DeepCNN, but also considers domain alignment and category alignment, so it achieves good performance in cross-domain recognition. As shown in Fig. 8 (f) (the solid line is the common boundary of the two domains, which achieves the domain alignment), we find that the feature distribution between different domains is consistent, and the discriminability between categories is preserved. Figure. 8 (c) indicates that its category boundary is sharp, and the classification performance is not significantly reduced while shifting the knowledge from source to target.

E. Ablation Study

In this part, we perform ablation analysis to evaluate the impact of the domain discriminator and the Bi-Classifier module on the overall performance of the BCJDA based on the process in Fig. 2. We set up four cases for this study: case1 is the CGDM-based BCJDA model retaining the Bi-Classifier and the domain discriminator; case2 and case3 remove the domain discriminator, and uses CGDM and BCDM, two groups of determinacy disparity for inter-classifier adversarial distance respectively; case4 retains the domain discriminator, but only applies the single classifier (i.e., there is no adversarial distance).

The final results are presented in Fig. 9. These results show that both the domain discriminator and the Bi-Classifier can effectively improve the performance and robustness of the model. Although the Bi-Classifier can simulate the function of the domain discriminator to a certain extent, the domain discriminator can perform more targeted adversarial training, and then mine more effective domain invariant features. In addition, the added Bi-Classifier compared with a single classifier can preserve the discrimination of features while domain adaptation, and then optimize the partition of boundary features. In conclusion, the comparison between the model with different modules and the proposed model confirms that they all contribute to varying degrees to the BCJDA.

F. Analysis of Generalization Performance

To further evaluate the robustness and generalizability of the BCJDA model, we conducted cross-subject experiments on an emotion dataset called 'SEED'. In this section, we present the cross-subject recognition accuracy results of our model and the compared methods on this dataset, and provide detailed data for our model on each subject.

The SEED dataset consisted of 15 subjects, each of whom collected EEG data for 3 sessions. In each session, each subject was required to watch 15 movie clips, including 5 negative, 5 neutral, and 5 positive. Self-assessment ensured that each subject's emotion was consistent with that shown in the movie clips. The EEG data was collected at the 10-20 international standard and consisted of 62 electrode channels with a sampling rate of 200 Hz. More information can be found in paper [45]. In our experiments, we divided each subject's signal into 1692 samples, each with a time length of 2s. For cross-subject experiments, one of the subjects served as the target domain, and all the remaining subjects were the source domain. We used the most representative L1-norm-based BCJDA model and selected the session holding the best performance.

Table SVII in the supplementary materials gives the accuracy, F1-score, sensitivity and specificity metrics for the cross-subject recognition of our model on the SEED dataset, where the F1-score is computed in weighted form. Table SVIII in the Supplementary materials shows the average accuracies of the compared methods. Our method achieves the optimal result with an accuracy of 88.81%, which is an improvement of about 29% with respect to the CNN method, as well as slightly higher than the JDA method, and has a smaller standard deviation. The F1-score is also consistent with the accuracy, proving the better performance of our model in each category.

The experimental results show that our method is not only applicable to CWD, but can also be used in other areas of cross-domain decoding of EEG signals, such as emotion categorization.

VI. CONCLUSION

In this study, we introduce the BCJDA model tailored for decoding cognitive workload across time and subjects, incorporating domain-level and class-level alignment. The evaluation of our model encompassed three classification scenarios involving cross-time and cross-subject experiments, enhancing the state-of-the-art in adversarial domain adaptation methodologies for the determination of three cognitive states. Furthermore, we conducted a pioneering comparison of various determinacy disparity algorithms employed by Bi-Classifier on model performance, revealing that the CGDM-based BCJDA exhibited superior performance. Our model not only mitigates the loss of category distinction during knowledge transfer but also significantly enhances model performance, particularly demonstrating its efficacy in addressing sensitive EEG data classification challenges. Given the diversity in network structures beyond CNN, utilized for EEG classification tasks, we intend to integrate more powerful modules, such as transformer [46], adaptive graph convolution module [47], and multi-head self-attention layer [48] into BCJDA.

REFERENCES

[1] L. R. Krol, O. W. Klaproth, C. Vernaleken, N. Russwinkel, and T. O. Zander, "Towards neuroadaptive modeling: Assessing the cognitive states of pilots through passive brain–computer interfacing," in Current Research in Neuroadaptive Technology, S. H. Fairclough and T. O. Zander, Eds. New York, NY, USA: Academic, 2022, pp. 59–73.

[2] K. Van Benthem, S. Gard, and C. Herdman, "Monitored versus non-monitored stimuli in brain–computer-interface methods for classifying workload states during piloting tasks," in Proc. IEEE Sensors Appl. Symp. (SAS), Ottawa, ON, Canada, Jul. 2023, pp. 1–6.

[3] M. S. Young, K. A. Brookhuis, C. D. Wickens, and P. A. Hancock, "State of science: Mental workload in ergonomics," Ergonomics, vol. 58, no. 1, pp. 1–17, Jan. 2015.

[4] A. T. Welford, "Mental work-load as a function of demand, capacity, strategy and skill," Ergonomics, vol. 21, no. 3, pp. 151–167, Mar. 1978.

[5] C. D. Wickens, "Multiple resources and performance prediction," Theor. Issues Ergonom. Sci., vol. 3, no. 2, pp. 159–177, Jan. 2002.

[6] M. S. Young and N. A. Stanton, "What's skill got to do with it? Vehicle automation and driver mental workload," Ergonomics, vol. 50, no. 8, pp. 1324–1339, Aug. 2007.

[7] D. D. Chakladar, S. Datta, P. P. Roy, and V. A. Prasad, "Cognitive workload estimation using variational autoencoder and attention-based deep model," IEEE Trans. Cogn. Develop. Syst., vol. 15, no. 2, pp. 581–590, Jun. 2023.

[8] M. A. Almogbel, A. H. Dang, and W. Kameyama, "EEG-signals based cognitive workload detection of vehicle driver using deep learning," in Proc. 20th Int. Conf. Adv. Commun. Technol. (ICACT), Chuncheon, South Korea, Feb. 2018, pp. 256–259.

[9] R. McKendrick, B. Feest, A. Harwood, and B. Falcone, "Theories and methods for labeling cognitive workload: Classification and transfer learning," Frontiers Hum. Neurosci., vol. 13, p. 295, Sep. 2019.

[10] Y. Liu et al., "EEG based mental workload assessment by power spectral density feature," in Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA), Aug. 2022, pp. 450–454.

[11] A. Gupta, G. Siddhad, V. Pandey, P. P. Roy, and B.-G. Kim, "Subject-specific cognitive workload classification using EEG-based functional connectivity and deep learning," Sensors, vol. 21, no. 20, p. 6710, Oct. 2021.

[12] M. Mazher, A. A. Aziz, A. S. Malik, and H. U. Amin, "An EEG-based cognitive load assessment in multimedia learning using feature extraction and partial directed coherence," IEEE Access, vol. 5, pp. 14819–14829, 2017.

[13] S. Mathan, A. Smart, T. Ververs, and M. Feuerstein, "Towards an index of cognitive efficacy EEG-based estimation of cognitive load among individuals experiencing cancer-related cognitive decline," in Proc. Annu. Int. Conf. IEEE Eng. Med. Biol., Buenos Aires, Argentina, Aug. 2010, pp. 6595–6598.

[14] Y. Wang et al., "LGNet: Learning local–global EEG representations for cognitive workload classification in simulated flights," Biomed. Signal Process. Control, vol. 92, Jun. 2024, Art. no. 106046.

[15] Y. Zhou et al., "Cross-task cognitive workload recognition based on EEG and domain adaptation," IEEE Trans. Neural Syst. Rehabil. Eng., vol. 30, pp. 50–60, 2022.

[16] R. Zanetti, A. Arza, A. Aminifar, and D. Atienza, "Real-time EEG-based cognitive workload monitoring on wearable devices," *IEEE Trans. Biomed. Eng.*, vol. 69, no. 1, pp. 265–277, Jan. 2022.

[17] 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. Develop. Syst.*, vol. 14, no. 3, pp. 799–818, Sep. 2022.

[18] Y. Peng, H. Liu, J. Li, J. Huang, B.-L. Lu, and W. Kong, "Cross-session emotion recognition by joint label-common and label-specific EEG features exploration," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 759–768, 2023.

[19] N. Jamil, A. N. Belkacem, S. Ouhbi, and C. Guger, "Cognitive and affective brain–computer interfaces for improving learning strategies and enhancing student capabilities: A systematic literature review," *IEEE Access*, vol. 9, pp. 134122–134147, 2021.

[20] C.-S. Wei, Y.-P. Lin, Y.-T. Wang, C.-T. Lin, and T.-P. Jung, "A subject-transfer framework for obviating inter- and intra-subject variability in EEG-based drowsiness detection," *NeuroImage*, vol. 174, pp. 407–419, 2018.

[21] C. Hao, J. Ming, and L. Zhunan, "MS-MDA: Multisource marginal distribution adaptation for cross-subject and cross-session EEG emotion recognition," *Frontiers Neurosci.*, vol. 15, Jan. 2021, Art. no. 778488.

[22] Y. Zhou et al., "Cross-subject cognitive workload recognition based on EEG and deep domain adaptation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–12, 2023.

[23] J. Yedukondalu and L. D. Sharma, "Cognitive load detection using binary salp swarm algorithm for feature selection," in *Proc. IEEE 6th Conf. Inf. Commun. Technol. (CICT)*, Gwalior, India, Nov. 2022, pp. 1–5.

[24] 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 Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1149–1159, Jun. 2019.

[25] W. L. Lim, O. Sourina, Y. Liu, and L. Wang, "EEG-based mental workload recognition related to multitasking," in *Proc. 10th Int. Conf. Inf., Commun. Signal Process. (ICICS)*, Singapore, Dec. 2015, pp. 1–4.

[26] C. Tremmel, C. Herff, T. Sato, K. Rechowicz, Y. Yamani, and D. J. Krusienski, "Estimating cognitive workload in an interactive virtual reality environment using EEG," *Frontiers Hum. Neurosci.*, vol. 13, p. 401, Nov. 2019.

[27] L. Jin, H. Qu, L. Pang, Z. Zhang, and Z. Lyu, "Identifying stable EEG patterns over time for mental workload recognition using transfer DS-CNN framework," *Biomed. Signal Process. Control*, vol. 89, Mar. 2024, Art. no. 105662.

[28] S. Kuanar, V. Athitsos, N. Pradhan, A. Mishra, and K. R. Rao, "Cognitive analysis of working memory load from eeg, by a deep recurrent neural network," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Calgary, AB, Canada, Apr. 2018, pp. 2576–2580.

[29] Z. Ni, J. Xu, Y. Wu, M. Li, G. Xu, and B. Xu, "Improving cross-state and cross-subject visual ERP-based BCI with temporal modeling and adversarial training," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 369–379, 2022.

[30] N. Z. Zenia and Y. Hu, "Deep learning architectures used in eeg-based estimation of cognitive workload: A review," in *Proc. IEEE Int. Conf. Auto. Syst. (ICAS)*, Aug. 2021, pp. 1–5.

[31] S. Zhang et al., "A review of rotation mechanical fault diagnosis research based on deep domain adaptation," in *Proc. IEEE 9th Int. Conf. Big Data Comput. Service Appl. (BigDataService)*, Athens, Greece, Jul. 2023, pp. 247–248.

[32] S. Furukawa, T. Sakuma, and S. Kato, "Emotion recognition with domain adaptation based on few-shot EEG learning," in *Proc. IEEE 10th Global Conf. Consum. Electron. (GCCE)*, Kyoto, Japan, Oct. 2021, pp. 1–2.

[33] S. Furukawa, T. Sakuma, and S. Kato, "Emotion recognition model with class-wise domain adaptation based on few-shot EEG learning," in *Proc. IEEE 11th Global Conf. Consum. Electron. (GCCE)*, Osaka, Japan, Oct. 2022, pp. 697–698.

[34] W. Cai et al., "Hierarchical domain adaptation projective dictionary pair learning model for EEG classification in IoMT systems," *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 4, pp. 1559–1567, Aug. 2022.

[35] S. Chambon, M. N. Galtier, and A. Gramfort, "Domain adaptation with optimal transport improves EEG sleep stage classifiers," in *Proc. Int. Workshop Pattern Recognit. Neuroimaging (PRNI)*, Singapore, Jun. 2018, pp. 1–4.

[36] E. R. M. Heremans, T. Osselaer, N. Seeuws, H. Phan, D. Testelmans, and M. De Vos, "Data augmentation in semi-supervised adversarial domain adaptation for EEG-based sleep staging," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat. (BHI)*, Ioannina, Greece, Sep. 2022, pp. 1–4.

[37] K. Saito, K. Watanabe, Y. Ushiku, and T. Harada, "Maximum classifier discrepancy for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3723–3732.

[38] S. Li, F. Lv, B. Xie, C. H. Liu, J. Liang, and C. Qin, "Bi-classifier determinacy maximization for unsupervised domain adaptation," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, 2021, pp. 8455–8464.

[39] Z. Du, J. Li, H. Su, L. Zhu, and K. Lu, "Cross-domain gradient discrepancy minimization for unsupervised domain adaptation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 3936–3945.

[40] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, Sep. 1982.

[41] 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, Oct. 2018, Art. no. 056013.

[42] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A multi-branch 3D convolutional neural network for EEG-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2164–2177, Oct. 2019.

[43] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," *Human Brain Mapp.*, vol. 38, no. 11, pp. 5391–5420, 2017.

[44] J. Li, S. Qiu, C. Du, Y. Wang, and H. He, "Domain adaptation for EEG emotion recognition based on latent representation similarity," *IEEE Trans. Cogn. Develop. Syst.*, vol. 12, no. 2, pp. 344–353, Jun. 2020.

[45] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Autonomous Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[46] G. Fu, Y. Zhou, P. Gong, P. Wang, W. Shao, and D. Zhang, "A temporal-spectral fused and attention-based deep model for automatic sleep staging," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1008–1018, 2023.

[47] P. Gong, P. Wang, Y. Zhou, and D. Zhang, "A spiking neural network with adaptive graph convolution and LSTM for EEG-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 1440–1450, 2023.

[48] P. Wang, Z. Li, P. Gong, Y. Zhou, F. Chen, and D. Zhang, "MTRT: Motion trajectory reconstruction transformer for EEG-based BCI decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2349–2358, 2023.