Disentangled and Side-Aware Unsupervised Domain Adaptation for Cross-Dataset Subjective Tinnitus Diagnosis

Yun Li, Zhe Liu, Lina Yao, Senior Member, IEEE, Jessica J. M. Monaghan, and David McAlpine

Abstract—EEG-based tinnitus classification is a valuable tool for tinnitus diagnosis, research, and treatments. Most current works are limited to a single dataset where data patterns are similar. But EEG signals are highly non-stationary, resulting in model’s poor generalization to new users, sessions or datasets. Thus, designing a model that can generalize to new datasets is beneficial and indispensable. To mitigate distribution discrepancy across datasets, we propose to achieve Disentangled and Side-aware Unsupervised Domain Adaptation (DSUDA) for cross-dataset tinnitus diagnosis. A disentangled auto-encoder is developed to decouple class-irrelevant information from the EEG signals to improve the classifying ability. The side-aware unsupervised domain adaptation module adapts the class-irrelevant information as domain variance to a new dataset and excludes the variance to obtain the class-distill features for the new dataset classification. It also aligns signals of left and right ears to overcome inherent EEG pattern difference. We compare DSUDA with state-of-the-art methods, and our model achieves significant improvements over competitors regarding comprehensive evaluation criteria. The results demonstrate our model can successfully generalize to a new dataset and effectively diagnose tinnitus.

Index Terms—EEG signals, tinnitus diagnosis, disentangled representation, unsupervised domain adaptation.

I. INTRODUCTION

Subjective tinnitus is the perception of sounds without real acoustic stimulus and can only be heard by the patients themselves [1], [2]. People with tinnitus may suffer from sleep problems, difficulty concentrating, and negative emotions that may progress to psychiatric disorders [3], [4], [5]. Subjective tinnitus is a worldwide problem, affecting around 15% of the population [6], [7]. Therefore, it is essential to study the diagnosis and treatments of subjective tinnitus.

Recently, there have been substantial achievements in tinnitus diagnosis. Among them, methods based on electroencephalography (EEG) signals attract much attention because 1) many works prove that there exists a strong correlation between abnormal EEG and tinnitus [8], [9]; 2) EEG signals have objective evaluation compared with traditional methods that are based on the Visual Analog Scale (VAS) or questionnaires [10], [11], [12], [13]. With the development of information technology, many efforts are made to utilize machine learning and deep learning algorithms to analyze EEG signals and classify patients’ tinnitus status [14], [15], [16], [17], [18], [19], [20]. For example, Liu et al. [15] design a neural network with Siamese structure to augment training EEG data.

Despite the success of these methods, few of them equip with generalization abilities, especially when applying the models to datasets where distributions are mismatched with the training datasets. The collection and annotation of EEG signals of tinnitus are labor-intensive and time-consuming, which makes labeled training data hard to acquire. However, a large amount of labeled data are fundamental to the success of deep learning models. Therefore, we hope to adapt models trained on labeled datasets to other unlabeled datasets. However, EEG signals of different datasets can be sufficiently different due to EEG signals’ high sensitivity to non-physiological factors, e.g., biological or physical conditions of brains [9] and experimental settings such as time duration, sampling frequency, electrode placement, and equipment. The distribution discrepancy prevents the model’s cross-dataset tinnitus diagnosis. In other EEG-based fields, a few works explore unsupervised domain adaptation or transfer learning to overcome the distribution mismatch across datasets [21], [22], [23], [24]. However, in tinnitus diagnosis, methods that can generalize to other datasets without labeled data are under-explored.

In addition, although unsupervised domain adaptation can relieve the distribution discrepancy and enable a cross-dataset diagnosis to some extent, the adaptation process inevitably causes loss of useful information and introduces noise that is irrelevant to or even harms tinnitus diagnosis. Also, the original EEG
signals have a low signal-to-noise ratio, i.e., the data contains noise even without adaptation to other datasets. Therefore, designing models that can perform adaptation to transfer knowledge across datasets and exclude tinnitus-unrelated information is fundamental to improving the cross-dataset diagnosis performance.

Apart from noise caused during signal collection and domain adaptation, the side information of the ears, i.e., which ear the evoking acoustic signal is presented to, may also impair the classification. The right and left ears display different EEG patterns even if they have the same condition as tinnitus or normal [25], [26]. This derives from functional and anatomical inter-hemispheric differences in auditory cortex [27], [28], [29], [30]. However, this is not noise that can be excluded. Thus, it is essential to adapt the information and patterns of the left and right ears at a subject level to eliminate the side influence.

We propose a novel model to perform Disentangled and Side-aware Unsupervised Domain Adaptation (DSUDA) for cross-dataset subjective tinnitus diagnosis to address the aforementioned problems. We propose Disentangled Auto-Encoder (DAE) to decouple tinnitus-related and tinnitus-unrelated information. The performance can be improved as noise is excluded and only pure information is fed into the classifier. In addition, we present the Side-aware Unsupervised Domain Adaptation (SUDA) to perform subject- and domain-level adaptation. The side-aware adaptation aligns features of left and right ears to avoid intrinsic left-right EEG differences harming the classification. The domain-level adaptation cooperates with DAE to extract domain variance, i.e., the tinnitus-unrelated information that must be excluded, and adapt the variance across domains. Then unsupervised domain adaptation can be achieved so that the classifier trained on the source domain can be applied in the target domain without labeled target data. This is essential for EEG-based tinnitus diagnosis as it may reduce the high demand for labeled EEG signals, save cost, and find universal models shared across subjects, datasets, and different experimental settings.

In summary, we make three-fold contributions as follows:

- We propose a novel model aiming at disentangled and side-aware unsupervised domain adaptation (DSUDA) for cross-dataset subjective tinnitus diagnosis. DSUDA combines disentangled auto-encoder and side-aware unsupervised domain adaptation, thus retaining the discrimination power when applied to the target domain.
- We design an adversarial training framework for the cooperation of the proposed disentangled auto-encoder and side-aware unsupervised domain adaptation. Training the two modules in an adversarial manner can distill class-related information and align data from different sides and domains.
- We conduct extensive experiments and ablation studies to evaluate and analyze our DSUDA. Our model yields 8.6%, 10%, and 7.5% improvements over state-of-the-art (SOTA) methods in both-ears, left-ear, and right-ear accuracy, and consistently outperforms SOTA in other criteria, e.g., Negative-F1 score, Positive-F1 score, and Negative Predictive Value, etc. The results demonstrate our model’s generalization ability in a new domain.

II. RELATED WORK

A. EEG-Based Tinnitus Diagnosis and Treatment

Subjective tinnitus is defined as the perception of sound that can only be perceived by patients in the absence of an external sound source [1], [2]. Tinnitus, also known as phantom noise, can occur in people with or without hearing problems. Tinnitus has always been a worldwide health problem, with an overall incidence rate of about 11% to 16% [6], [7]. Tinnitus can impair the life quality of a patient in many ways, which may cause sleep problems, difficulty concentrating, and may even develop into psychiatric disorders, such as lethargy, anxiety, and anger [3], [4], [5]. For the treatment and diagnosis of tinnitus, traditional investigators assess with a Visual Analog Scale (VAS) or a temporary questionnaire [10], [11], [12], [13]. Haefeli et al. [10] and Kersten et al. [11] designed VAS to subjectively quantify certain sensations of patients using a sliding scale from lack to extreme sensation, e.g., pain. Different to the sliding scale, Powers et al. [12] and Adamchic et al. [13] proposed to use questionnaires with only three responses, i.e., yes, sometimes, and no, to identify the tinnitus conditions of sufferers. With the development of Electroencephalogram (EEG), EEG has become a valuable tool to diagnose and treat tinnitus. EEG is capable of recording brain activity and analyzing neural oscillations in real time [9]. A lot of evidence shows a strong correlation between abnormal EEG and tinnitus. For example, Eggermont et al. [8], and Iburra-Zarate et al. [9] suggested that tinnitus may be caused by abnormal restructurings in the tonotopic maps, higher synchrony or spontaneous firing rates in the neuron or the central auditory system. Nguyen et al. [31] found that tinnitus sufferers have abnormal neural oscillations in the beta and gamma frequency bands.

Traditional EEG-based tinnitus diagnosis and treatment rely on the statistical models to quantify and analyze the provided neurofeedback [32], [33]. For example, Vanneste et al. [34] performed a quantitative EEG analysis of 154 patients with tinnitus and used a regression analysis to find that structural differences in gray matter may be associated with the distress and duration of tinnitus. Moazami-Goudarzi et al. [35] used power spectra to analyze tinnitus and control groups in a group-wise manner. They found that tinnitus will lead to a power enhancement in the delta and theta band of EEG. Although previous researchers have made significant progress in diagnosing and analyzing tinnitus-related EEG signals, their statistical analysis was case-specific and prone to being trapped in the characteristics of small study populations. To ease such problem, more and more work uses deep learning methods to improve the generalization ability of models [14], [15], [18], [19]. For example, Liu et al. [15] adopted a Siamese network to augment training samples with contrastive relationships and thus enhance the generalization ability of their model. Allgaier et al. [18] used noise reduction and down-sampling to generalize the learning process of their proposed deep learning method and achieved an accuracy of 75.6% in diagnosing tinnitus. However,
these methods are still limited to a single dataset. EEG signals are easy to be influenced by multiple non-physiological factors, e.g., biological or physical conditions of brains [9]. The data distributions may significantly differ from subjects. Moreover, the different settings of the experimental environment, e.g., time duration and sampling frequency, may lead to deep learning models failing to be applied in a new dataset. Therefore, it is necessary to develop a model that is robust across datasets to handle more diverse and general EEG signals.

B. Cross-Dataset EEG Research in Related Fields

Most current EEG research [36], [37], [38], [39] only takes into account the data distributions across subjects, which still ignores the distribution shift between datasets. Some modern models [22], [23], [24], [40], [41] have evolved to ease the distribution shifts across both subjects and datasets. In other words, these models can be applied to multiple datasets rather than a single dataset. For example, in the field of gesture classification, Bird et al. [40] proposed to transfer knowledge between an EEG dataset recording brainwaves and an Electromyographic (EMG) dataset recording muscular waves based on the pre-trained networks. They achieved a high accuracy of 85.12% and 93.82% by transferring EEG-to-EMG and EMG-to-EEG knowledge, respectively. Similarly, for emotion recognition, Cimitay et al. [22] proposed to use a pre-trained neural network from a source dataset to the target dataset and achieved an impressive enhancement in the model performance. These models use pre-trained models to provide extra knowledge that may be meaningful across datasets. To provide a more reliable solution to learn meaningful information across datasets, Xu et al. [23] proposed an online method to pre-align the data formats before training and classification. However, all these methods rely on the annotation of target dataset, while many real-world datasets may lack labels. Therefore, it is necessary to propose a work that can be applied in a more general condition, i.e., only a given source dataset is annotated, and the target dataset that needs classification remains unlabeled.

C. Unsupervised Domain Adaptation

A domain refers to feature space and the corresponding probability distributions related to a set of data, e.g., a dataset [42]. Unsupervised domain adaptation focuses on the situation where the training and testing data are collected from different distributions, e.g., datasets. The training dataset with labeled data and the testing dataset without any annotation are called the source and target datasets, respectively. Conventional deep learning models may suffer from a significant decrease in the model performance due to the domain difference between source and target datasets [43], [44], [45], [46], [47], [48], [49], [50]. Unsupervised domain adaptation aims to ease such performance decrease by adapting the different domains [51], [52], [53], [54], [55]. A lot of works [21], [56], [57], [58] have been done to adopt unsupervised domain adaptation in the field of EEG. Most of these works only view each subject as a domain and fail to ease the dataset-level domain difference. Limited research [59], [60] has been done to adopt unsupervised domain adaptation for cross-dataset settings. For example, Lan et al. [60] adopted domain adaptation to reduce both cross-subject and cross-dataset variance based on unsupervised domain adaptation and achieved an impressive improvement in emotion recognition. However, the previous work lacks domain-specific knowledge to handle domain differences related to tinnitus, e.g., the domain difference caused by the different sides of the ear of presentation of the evoking stimulus. Our work is tinnitus-specific domain adaptation that is capable of easing cross-ear and cross-dataset differences.

D. Disentangled Representation

To better serve specific tasks, many deep-learning-based generative adversarial networks [61], [62], [63], [64], [65], [66] disentangle the representations from the given data to learn the specific information. For example, Li et al. [67] used a disentangled auto-encoder to disentangle category-distilling and category-dispersing factors from images for better category classification. Similarly, there exists more information or noise than tinnitus-related factors in the collected EEG signals. For example, Cai et al. [68] adopted spatial smoothing to increase the signal-to-noise ratio and reduce the influence of signal noise. Joos et al. [69] and Meyer et al. [70] used questionnaires to disentangle and study the correlations of depression and distress to EEG from tinnitus patients. Therefore, we propose to disentangle the signals to exclude unrelated factors to obtain the pure classification information for tinnitus diagnosis.

III. METHOD

A. Problem Definition

Suppose that dataset $S = \{(x^s, y^s, a^s) | x^s \in R^{1 \times n^s}, y^s \in \{0, 1\}, a^s \in \{0, 1\}\}$ from source domain are given, where $x^s = \{o_j : j \in [1, n^s], o_j \in R\}$ is an EEG feature containing signals collected through $n^s$ time points, $y^s \in \{0, 1\}$ indicates the subject’s tinnitus status, and $a^s \in \{0, 1\}$ implies which ear the signals are collected from. More specifically, $y^s = 1$ denotes that the subject has tinnitus, and $y^s = 0$ represents control subject without tinnitus. For $a^s$, 0 or 1 denotes left or right ear, respectively. Then, given data from target domain $T = \{(x^t, y^t, a^t) | x^t \in R^{1 \times n^t}, y^t \in \{0, 1\}, a^t \in \{0, 1\}\}$, where $x^t$ contains $n^t$ time points, and $y^t$ ($a^t$, respectively) indicates same meanings as $y^s$ ($a^s$, respectively). Our goal is to obtain a general classifier that can diagnose tinnitus for the target domain without labeled target data based on the labeled source data. That is, during training, only $S$ from the source domain and $\{(x^s, a^s)\}$ of the target domain are given, and we need to implement unsupervised domain adaptation to achieve the cross-dataset diagnosis.

B. Pipeline Overview

As Fig. 1 shows, our proposed DSUDA integrates two modules: the Disentangled Auto-Encoder (DAE) and the Side-aware Unsupervised Domain Adaptation (SUDA). DAE (Section III-D) disentangles input features from source and target domains into pure class information that is domain-invariant and domain
Overview of DSUDA. DSUDA consists of two modules: DAE and SUDA. DAE is composed of pure-info encoder, domain-variance encoder, classifier, and decoder. The labeled data from the source domain are first preprocessed to align the unlabeled data from the target domain; then, the data are normalized to be fed into DAE. The two encoders of DAE disentangle the input data into domain-invariant features and domain variance, which are then combined as the input of the Decoder. The classifier takes domain-invariant features as input to predict tinnitus status. SUDA data are normalized to be fed into DAE. The two encoders of DAE disentangle the input data into domain-invariant features and domain variance.

C. Data Preprocessing

Due to the different acquisition equipment and settings of EEG signals, the data in the source and target domains may have significant discrepancies. The overall time duration and time steps of the collection may be different. And the ranges of signal strength or device noise differ across devices. Therefore, it is not feasible to directly apply a classifier suitable for the source domain to the target domain. We need to first preprocess the source data to align the data format and scope, and then use DAE and SUDA to eliminate domain variance.

As shown in Fig. 2, suppose that an EEG trial duration from source and target domains, i.e., \(x^s\) and \(x^t\), last for \(d^s\) and \(d^t\) milliseconds, respectively. We adopt the strategy of sliding window to segment \(x^s\) into slices \(\{x_1^s, x_2^s, \ldots\}\), where the window size equals \(d^s\), and the sliding step is a hyper parameter. Then, we use down sampling to compress \(x_1^s = \{\hat{o}_1^s, \hat{o}_2^s, \ldots\}\) to align the number of time points in \(x_t^s\) (i.e., \(n^w = \frac{d^s \cdot n_t}{d^t}\)) and \(x^t\) (i.e., \(n^t\)) as follows:

\[
\hat{o}_p^s = \begin{cases} \frac{1}{\tau} \sum_{q \in [(p-1) \cdot \tau + 1, p \cdot \tau]} o_q^s & p \in [1, m] \\ \frac{1}{\tau} \sum_{q \in [(p-1) \cdot \tau + 1, m \cdot \tau - p \cdot \tau + 1 - l]} o_q^s & p \in [m+1, n^t]\end{cases}
\]

\[
(1) \quad \text{s.t.} \quad l = \left\lfloor \frac{n^w}{n_t} \right\rfloor \quad m = n^w - 1 \cdot n_t
\]

where \(l = \left\lfloor \frac{n^w}{n_t} \right\rfloor\), the floor of \(\frac{n^w}{n_t}\), is the compressing step. In this way, we obtain a set of processed features \(\{\hat{x}_1^s, \hat{x}_2^s, \ldots\}\) from the original \(x^s\), where \(\hat{x}_1^s = \{\hat{o}_1^s, \hat{o}_2^s, \ldots\}\) is consistent with \(x^t\) in time duration and time point number. Note that we perform the data preprocessing based on the assumptions that \(d^s \geq d^t\) and \(n^w \geq n^t\), or we align target domain to source domain in reverse. Then, we normalize data from source and target domains with MinMaxNorm separately as follows:

\[
\hat{o}_p^s = \frac{o_p^s - \min o_p^s}{\max o_p^s - \min o_p^s}
\]

\[
(3) \quad \hat{o}_p^t = \frac{o_p^t - \min o_p^t}{\max o_p^t - \min o_p^t}
\]

where \(o_p^t\) is the \(p^{th}\) value in \(x^t\).

D. Disentangled Auto-Encoder

DAE aims to improve tinnitus diagnosis by extracting pure information that excludes categorically irrelevant factors. For this, we build a DAE using two parallel encoders. The purely informative encoder \(f_{pe}\) (Pure-info Encoder) learns classification-related features, and the classifier \(f_c\) supervises \(f_{pe}\) to contain
classification information. The domain variance encoder \( f_{ve} \) learns class-independent features. The decoder \( f_{de} \) constrains \( f_{pe} \) to avoid overfitting the classification information, and restricts \( f_{pe} \) and \( f_{ve} \) from overfitting the source domain.

We take \( S = \{ (\hat{x}^s, y^s, a^s) \} \) and \( T = \{ (\hat{x}^t, a^t) \} \) as the input of DAE, where \( \hat{x}^s = \{ x_1^s, x_2^s, \ldots \} \), \( \hat{x}_1^t = \{ \hat{o}_1^t, \hat{o}_2^t, \ldots \} \), and\( \hat{x}^t = \{ \hat{o}_1^t, \hat{o}_2^t, \ldots \} \). Given \( \hat{x} \in D = S \cup T \), DAE disentangles \( \hat{x} \) into class-related embedding \( e^c \) and class-unrelated embedding \( e^v \). Then \( e^c \) is fed into the classifier \( f_c \) for classification, and \( e^c \) and \( e^v \) are concatenated as input of the decoder \( f_{de} \) for reconstruction. DAE is optimized by the classification loss \( L_{cls} \) calculated based on all \( \hat{x} \in D \), while \( L_{cls} \) is computed based on only \( \hat{x}^s \in \hat{S} \).

However, only loss \( L_{cls} \) and \( L_{ae} \) are not sufficient. With \( L_{cls} \) and \( L_{ae} \), we can only ensure that \( f_c \) works for the source domain. To enable its feasibility in the target domain, we hope the extracted class-related information \( e^c \) is domain-invariant to generalize to the target domain. Intuitively, after excluding the domain variance from the source domain, the remained information is domain-invariant. Therefore, we can consider the class-unrelated features \( e^v \) that should be discarded as the domain variance, and adapt \( e^v \) from the source domain to the target domain by the side-aware unsupervised domain adaptation, which will be discussed in Section III-E.

E. Side-Aware Unsupervised Domain Adaptation

The features learned by DAE may not fit target domain. To adapt DAE to the target domain, we propose the SUDA focusing on adapting domain variance from source to target domain. To avoid the confusion caused by different signal patterns of right and left ears, we further equip the SUDA with the side-aware adaptation to align information from both ears.

As shown in Fig. 1, SUDA is composed of the domain \( f_{sd} \) and side discriminator \( f_{sd} \), predicting which domain the domain-variant features belong to and which ear the domain-variant features are collected from, respectively. Then, we design the loss functions of the domain and side discriminator as \( L_d \) and \( L_{side} \), which aims to optimize \( f_{sd} \) and \( f_{sd} \) to be able to distinguish source domain variance \( \hat{x}^s \) as 0, target domain variance \( \hat{x}^t \) as 1, left ear features as 0, and right ear features as 1:

\[
L_d = \mathbb{E}_{\hat{x} \sim T}[f_{dd}(f_{ve}(\hat{x}))] - \mathbb{E}_{\hat{x} \sim S}[f_{dd}(f_{ve}(\hat{x}))]
\]

\[
L_{side} = \mathbb{E}_{\hat{x} \sim D^r}[f_{sd}(f_{pe}(\hat{x}))] - \mathbb{E}_{\hat{x} \sim D^l}[f_{sd}(f_{pe}(\hat{x}))]
\]

\[
L_{DS} = L_d + \eta \cdot L_{side}
\]

where \( D^r \) and \( D^l \) denote sets of \( \hat{x} \) that are collected from right and left ears, respectively, \( L_{DS} \) is the overall discriminating loss, and \( \eta \) is a hyper-parameter.

With losses \( L_d \) and \( L_{side} \), we can obtain reliable discriminators. Then, we can optimize DAE by optimizing \( f_{ve} \) to confuse the domain discriminator to extract domain variance, and optimizing \( f_{pe} \) to confuse the side discriminator to align side information. Intuitively, if the domain discriminators cannot distinguish class-irrelevant features from different domains, the domain variance is adapted successfully and \( f_{pe} \) can disentangle class-related information from \( \hat{x} \) for both domains. Then, we can apply the classifier trained on source domain to target domain.

We optimize DAE and SUDA in an adversarial manner as follows: 1) maximizing \( L_{DS} \) to enable \( f_{dd} \) and \( f_{sd} \) to distinguish between different domains and ear sides; 2) optimizing \( f_{ve}, f_{pc}, f_{c}, \) and \( f_{de} \) to minimize \( L_{DAE} \) to improve the classifier and fool the discriminators. \( L_{DAE} \) is the combined loss based on (5)-(8) as follows:

\[
L_{DAE} = \alpha \cdot L_d + \beta \cdot L_{side} + L_{ae} + L_{cls}
\]

\[
= -\mathbb{E}_{\hat{x} \sim T}[f_{dd}(f_{ve}(\hat{x}))] + \mathbb{E}_{\hat{x} \sim D^r}[f_{sd}(f_{pe}(\hat{x}))]
\]

\[
+ L_{ae} + L_{cls}
\]

where \( \alpha \) and \( \beta \) are two hyper-parameters, \( L_d \) and \( L_{side} \) are simplified versions of \( L_d \) and \( L_{side} \), without requirements of recognizing data from source domain or right ear. This is because that optimizing DAE to fit target domain to the source domain and align information from left to right side is sufficient for adaptation between two domains.

F. Training and Inference

Training: Due to the instability of adversarial training, we first train Pure-info Encoder, Classfier, and Decoder to provide a basic classification ability, and use the trained parameters of Pure-info Encoder as parameters of Domain-variace Encoder. Then, we consider the trained DAE as the initial model, and optimize DAE and SUDA based on \( L_{DAE} \) and \( L_{DS} \) adversarially.

Inference: We use Pure-info Encoder to extract class-related features from target domain data and use the Classifier to predict tinnitus status directly without modifying its parameters.

IV. EXPERIMENTS

A. Dataset and Experimental Setting

We evaluate our model based on two Auditory brainstem responses (ABRs) datasets [72], [73] based on EEG sensors and consider the larger datasets [72] as the source domain and the smaller one [73] as the target domain. ABRs are signals that originate from the early stages of the auditory pathway and evoked potentials recorded with EEG sensors [72], [73]. Schaette and McAlpine [72] used presentation levels of 90 and 100 dB peSPL with an 11 click/s rate for the source domain.
approved by the University College London (UCL) ethics committee. The ethics ID number is 2039/002 and the declaration is the Helsinki declaration. Guest et al. [73] collected the target domain in response to 102 dB peSPL clicks presented with a rate of 7 clicks/s approved by the National Research Ethics Service Greater Manchester West Ethics Committee with REC reference 15/NW/0133 and IRAS project ID 168221. The signals are sampled with 50 kHz and bandpass-filtered before 100 Hz and 1500 Hz. The source dataset contains 408 EEG signals of 38 subjects recorded while an acoustic click train was presented over headphones to the affected ears for patients or either ear for control subjects, i.e., only one ear of each subject is recorded. Each trial lasts for 10 milliseconds resulting in 500-time points. The target dataset collects 80 EEG signals of 40 subjects with both of their ears recorded. The time duration is 8 milliseconds with 131-time points per trial. Labeling strategies in [72] and [73] are different. [72] provides labels for each signal, while [73] provides subject-level labels that signals from two ears of the same person are labeled with the same state regardless of which ear experiences tinnitus. Thus, we choose [72] with more precise labels for each signal to learn a more reliable source model. Note that subjects in the two datasets are different.

This target dataset was approved by the University College London (UCL) ethics committee. The ethics ID number of our study is 2039/002. The declaration is the Helsinki declaration.

The 2017 study was approved by the National Research Ethics Service Greater Manchester West Ethics Committee, REC reference 15/NW/0133, IRAS project ID 168221.

We use Fully Connected (FC) Layers to build our model. \( f_{pC} \) and \( f_{vC} \) have the same structures consisting of 2 FC layers with Tanh as the activation function. Other model components also contain 2 FC layers while adopting Sigmoid as the activation function. During data preprocessing, we set the sliding step as 20; thus an original trial from the source domain can be sliced into 6 segments, which are further down-sampled. We set learning rates of DAE and SUDA as 1.0e-3 with different updating steps per batch, 10 and 1, respectively. We set the loss ratios \( \eta \), \( \alpha \), and \( \beta \) as 1.0, 1.0, and 1.0e-6.

We choose 9 machine or deep learning methods as competitors: XGBoost [74], nu-SVC [75], nCSP [76], DeepNet [77], ShallowNet [77], AEXGB [78], EEGNet [79], AE [80], and SAE [15]. We also take our variants as baselines of transfer learning: UDA is a baseline of the unsupervised domain adaptation method; SUDA is a UDA enhanced by side-aware learning. For all methods, we first train the model on the source dataset and then preprocess the data in the target dataset to keep its format and scope consistent with the source data. After alignment, since the labels of the source and target dataset are the same, the target data can be directly input into the model trained on the source dataset for tinnitus diagnosis on the target dataset.

Unlike regular classification tasks, incorrect tinnitus diagnosis may cause severe consequences. Accuracy alone is not sufficient to evaluate the models. Therefore, we adopt Accuracy (Acc) along with Negative Predictive Value (NPV), True Negative Rate (TNR), Positive Predictive Value (PPV), True Positive Rate (TPR), F1-score for negative samples (N-F1), and F1-score for positive samples (P-F1) to evaluate the performance comprehensively.

### B. Experimental Results

1) **Comparison With State-of-The-Art Methods:** We present the results of our model and comparison in 3 different settings: both-side, left-side, and right-side, respectively, in Tables (I-III).

We can find that deep learning methods (DeepNet [77], ShallowNet [77], AEXGB [78], EEGNet [79], AE [80], SAE [15] and our model) generally perform better than traditional machine learning methods (XGBoost [74], nu-SVC [75], and nCSP [76]) on both ears. Among the compared methods, AE [80] and SAE [15] adopt an auto-encoder structure similar to our method, and both achieve the most competitive performance, proving that the auto-encoder structure is more appropriate to extract features from EEG signals. Besides, generative-based models, e.g., SAE, perform better than non-generative models, suggesting that the generative algorithms have a better generalization ability. DSUDA performs the best among all the algorithms, outperforming the second-best algorithm (including our variant) on N-F1, P-F1, accuracy by 3.5%/7.1%, 3.0%/5.0%, and 3.7%/7.5% in both-/left-side settings, respectively; 2.3%, 3.0%, and 5.0% higher than the best comparison algorithm on the right-side setting, respectively. Our model shows better improvement in the left ear than in other settings, which indicates that the left ear presentations have more transferable knowledge on these two datasets than the right ear. Incremental improvements of UDA over AE, SUDA over UDA, and DSUDA over SUDA on both left-side and both-side settings, respectively, indicating that domain adaptation, side-aware learning and disentangling tinnitus-related and -unrelated features are effective in enhancing the model performance. SAE [15] also introduces domain adaptation into the model training. More specifically, it adopts supervised subject-level adaptation inside dataset. Although its accuracy is not significantly improved compared with AE [80], the standard deviation is reduced from 0.056 to 0.013, proving the effectiveness of domain adaptation. UDA, SUDA, and DSUDA outperforming SAE indicates that dataset-level unsupervised domain adaptation is more effective than the subject-level supervised domain adaptation in the cross-dataset setting. On the right ear setting, the improvement on DSUDA is limited, which shows that the side-aware domain adaptation is more generalized and can fit more situations than the dataset domain adaptation.

We further analyze the wrongly-predicted samples to explore what limits our model in getting higher accuracy. The reasons may be different labeling strategies in the source and target datasets, diverse tinnitus severity, or distinct sound quality, etc. For example, Subject 19 has more severe tinnitus in the right ear, and our model can only recognize signals from the right ear as having tinnitus. Subject 36 misclassified by our model has tinnitus of high-frequency tone, different from all other patients. More precise labels, such as ringing tinnitus, may help address such limitations.

2) **ROC Curve Analysis:** In Fig. 3(a)-(c), we plot the ROC-AUC curves of DSUDA in both-ears, left-ear, and right-ear,
### Table I
**Overall Comparisons on Both Ears With SOTA Methods**

| Model      | NPV   | TNR   | N-F1  | PPV   | TPR   | P-F1  | Acc   |
|------------|-------|-------|-------|-------|-------|-------|-------|
| XGBoost    | 0.467 | 0.525 | 0.494 | 0.457 | 0.400 | 0.427 | 0.463 |
| Nu-SVC     | 0.467 | 0.350 | 0.400 | 0.480 | 0.600 | 0.533 | 0.475 |
| nCSP       | 0.568 | 0.625 | 0.595 | 0.583 | 0.525 | 0.553 | 0.575 |
| DeepNet    | 0.714 | 0.125 | 0.213 | 0.521 | 0.550 | 0.571 | 0.588 |
| ShallowNet | 0.581 | 0.625 | 0.602 | 0.595 | 0.550 | 0.571 | 0.588 |
| AE-XGB     | 0.696 | 0.400 | 0.508 | 0.579 | 0.825 | 0.680 | 0.613 |
| EEGBNET    | 0.605 | 0.650 | 0.627 | 0.622 | 0.575 | 0.597 | 0.613 |
| AE         | 0.660 | 0.775 | 0.713 | 0.727 | 0.600 | 0.658 | 0.688 |
| SAE        | 0.703 | 0.650 | 0.675 | 0.674 | 0.725 | 0.699 | 0.688 |

UDA       | 0.833 | 0.500 | 0.625 | 0.643 | 0.900 | 0.750 | 0.700 |
SUDA      | 0.744 | 0.725 | 0.734 | 0.732 | 0.750 | 0.741 | 0.738 |
DSUDA     | 0.789 | 0.750 | 0.769 | 0.762 | 0.800 | 0.780 | 0.775 |

Best results are in bold.

### Table II
**Results on the Left Ear**

| Model      | NPV   | TNR   | N-F1  | PPV   | TPR   | P-F1  | Acc   |
|------------|-------|-------|-------|-------|-------|-------|-------|
| XGBoost    | 0.455 | 0.500 | 0.476 | 0.444 | 0.400 | 0.421 | 0.450 |
| Nu-SVC     | 0.467 | 0.350 | 0.400 | 0.480 | 0.600 | 0.533 | 0.475 |
| nCSP       | 0.591 | 0.650 | 0.619 | 0.611 | 0.550 | 0.579 | 0.600 |
| DeepNet    | 0.500 | 0.100 | 0.167 | 0.500 | 0.900 | 0.643 | 0.575 |
| ShallowNet | 0.524 | 0.550 | 0.537 | 0.526 | 0.500 | 0.513 | 0.525 |
| AE-XGB     | 0.615 | 0.400 | 0.485 | 0.556 | 0.750 | 0.638 | 0.575 |
| EEGBNET    | 0.600 | 0.600 | 0.600 | 0.600 | 0.600 | 0.600 | 0.600 |
| AE         | 0.652 | 0.750 | 0.698 | 0.706 | 0.600 | 0.649 | 0.675 |
| SAE        | 0.750 | 0.600 | 0.667 | 0.667 | 0.800 | 0.727 | 0.700 |

UDA       | 0.900 | 0.450 | 0.600 | 0.633 | 0.950 | 0.760 | 0.700 |
SUDA      | 0.737 | 0.700 | 0.718 | 0.714 | 0.750 | 0.732 | 0.725 |
DSUDA     | 0.833 | 0.750 | 0.789 | 0.773 | 0.850 | 0.810 | 0.800 |

Best results are in bold.

### Table III
**Results on the Right Ear**

| Model      | NPV   | TNR   | N-F1  | PPV   | TPR   | P-F1  | Acc   |
|------------|-------|-------|-------|-------|-------|-------|-------|
| XGBoost    | 0.478 | 0.550 | 0.512 | 0.471 | 0.400 | 0.432 | 0.475 |
| Nu-SVC     | 0.467 | 0.350 | 0.400 | 0.480 | 0.600 | 0.533 | 0.475 |
| nCSP       | 0.545 | 0.600 | 0.571 | 0.556 | 0.500 | 0.526 | 0.550 |
| DeepNet    | 1.000 | 0.150 | 0.261 | 0.541 | 0.100 | 1.000 | 0.500 |
| ShallowNet | 0.636 | 0.700 | 0.667 | 0.667 | 0.600 | 0.632 | 0.650 |
| AE-XGB     | 0.800 | 0.400 | 0.533 | 0.600 | 0.900 | 0.720 | 0.650 |
| EEGBNET    | 0.609 | 0.700 | 0.651 | 0.647 | 0.550 | 0.595 | 0.625 |
| AE         | 0.667 | 0.800 | 0.727 | 0.750 | 0.600 | 0.667 | 0.700 |
| SAE        | 0.667 | 0.700 | 0.683 | 0.684 | 0.650 | 0.667 | 0.675 |

UDA       | 0.786 | 0.550 | 0.647 | 0.654 | 0.850 | 0.739 | 0.700 |
SUDA      | 0.750 | 0.750 | 0.750 | 0.750 | 0.750 | 0.750 | 0.750 |
DSUDA     | 0.750 | 0.750 | 0.750 | 0.750 | 0.750 | 0.750 | 0.750 |

Best results are in bold.
respectively. When discriminating the left ear, DSUDA achieved the highest scores in all criteria, i.e., 0/1 classification, Micro- and Macro-average AUC scores of 0.73/0.73, 0.74 and 0.76, respectively. This may be because after aligning the data of the left ear to the data of the right ear, the right ear data may be too aligned and close to each other. Then, the data of the left ear will be easier to distinguish compared to the right ear. In all three settings, our model scores higher on correctly classifying tinnitus subjects than control subjects. This may be due to the more similar patterns of tinnitus patients across different datasets than those of control subjects. It also shows that our model can effectively learn the data patterns of tinnitus patients across datasets.

C. Visualization of Learned Representations

1) Effectiveness of Side Adaptation: As shown in Fig. 4(a)–(b), we visualize the raw signals and representations after side-aware domain adaptation in the target dataset, respectively. Each subject contains two signals, i.e., one signal from the left and the right ear, which is in the same color. It is obvious that after adaptation, features of the right and left ears belonging to the same subjects are closer to each other than the raw signals. For example, two points of the same subject in light green are much closer in Fig. 4(b). This indicates that our side-aware domain adaptation effectively aligns signals from the left and right sides to a unified domain. SUDA can eliminate the side difference and make it easier for classifiers to detect whether the signals are from the same subject.

2) Effectiveness of Disentangled Auto-Encoder: We plot the raw signals and the tinnitus-related features learned by $f_{pe}$ in DSUDA in Fig. 4(c)–(d). The tinnitus subjects are colored pink, and the control subjects are blue. We can observe that in the raw features, the blue and pink nodes are entangled and hard to distinguish. In contrast, for features learned by $f_{pe}$, the pink nodes are more clustered together and further away from the blue node, making tinnitus and control subjects easier to classify. It proves that our model can effectively exclude noise and purify the data to improve classification, and $f_{pe}$ works for the target domain.

V. DISCUSSION

We set default experimental settings as 1) updating step per batch for DAE and SUDA: 1; 2) learning rates of DAE and SUDA: 1.0e-3; 3) loss ratios $\alpha$, $\beta$, $\eta$: 1.0. We do hyper-parameter studies to discuss how the updating steps, learning rates, and loss ratios can affect the performance. Each time, we fix other settings as the default setting, use a fixed random seed, and vary the factors we want to study.

A. Effect of Updating Steps

The results of the different updating steps for DAE and SUDA are shown in the Tables IV and V, respectively. We can find that the performance of SUDA is stable at different updating steps.
and shows a slightly increasing trend as the number of updating steps increases. In contrast, the results of varying updating steps of DAE vary widely, and the best results are achieved when the DAE is updated 10 times per batch. The reason may be that the optimization of SUDA can only affect the classification indirectly, while DAE directly extracts features to classify the signal. So minor changes can lead to non-negligible performance differences.

### B. Effect of Learning Rates

We show results with varied learning rates of two model components in Tables VI and VII. We can find that the model performance fluctuates up and down as the learning rate increases but is overall stable. Note that we set updating steps as the default value, i.e., 1, thus, according to Table IV, the both-ear accuracy would be around 0.725, and the best accuracy (0.775) can not be achieved. This explains why results in ablation studies (Tables V–X) are significantly lower than the best results.

### C. Effect of Loss Ratios

We summarize results with different loss ratios in Table VIII–X. We set the varying ranges based on the values and importance of $L_{d}$ and $L_{side}$: 0.01 to 1 for $\alpha$ ($L_{d}$); 1.0e-6 to 1 for $\beta$ and $\eta$ ($L_{side}$). Overall, the performance is stable when changing $\alpha$ and $\eta$, but decreases drastically when $\beta$ is larger than 0.1. The
reason lies in that during optimizing DAE, we mainly focusing on representation disentangling, and higher ratios of $L_{side}$ may misguide the optimization direction.

VI. CONCLUSION AND FUTURE DIRECTION

A. Conclusion

We propose a novel model, dubbed DSUDA, for EEG-based cross-datasets tinnitus diagnosis. DSUDA integrates disentangled auto-encoder and side-aware unsupervised domain adaptation to adapt the classifier trained on a source domain to the unlabeled target domain. The model can overcome domain discrepancy with domain variance adaptation, overcome inherent left-right ear pattern difference by subject-level side adaptation, and relieve the negative influence of EEG’s low signal-to-noise ratio by disentangling pure information from noises. Our model consistently outperforms state-of-the-art competitors with a large margin regarding classification accuracy and other criteria, demonstrating that the side-aware unsupervised domain adaptation facilitates the model’s generalization and the disentangled auto-encoder improves classifying performance.

B. Future Direction

Our model can be further extended to new scenarios in the future. There are several promising directions. For example, we will apply and improve our methods on large datasets to tackle other significant issues in signals, e.g., artifacts, outliers, and noises. We can also apply the DSUDA in other EEG-based tasks, such as activity or emotion recognition. The problems we addressed with our model, i.e., distribution discrepancy and low signal-to-noise ratio, are universal in EEG-based tasks. Thus,
our model can achieve good performance after being modified to fit other tasks. The potential of DSUDA is worth to dig. Or, we can implement DSUDA in real-world clinics or hospitals to assist doctors or therapists in diagnosing and treat tinnitus, which would be of great value to the community and healthcare. Also, we can improve DSUDA with interpretable components to help researchers in medical or bioinformatics reveal latent but essential findings, which may facilitate the cure of tinnitus. Taking other factors into consideration is another direction to improve our model. Although DSUDA already eliminates the influence of ear-side information, many other factors like ages or genders may also change the EEG signal patterns and thus impair classification accuracy.

ACKNOWLEDGMENT

The authors would like to thank Dr. Hannah Guest and Dr. Roland Schauette for sharing the datasets.

REFERENCES

[1] J. A. Henry, T. L. Zaugg, P. J. Myers, and M. A. Schechter, “Using therapeutic sound with progressive audiologic tinnitus management,” Trends Amplification, vol. 12, no. 3, pp. 188–209, 2008.

[2] K. M. Reavis, V. S. Rotholtz, Q. Tang, J. A. Carroll, H. Djallalian, and F.-G. Zeng, “Temporary suppression of tinnitus by modulated sounds,” J. Assoc. Res. Otolaryngol., vol. 13, no. 4, pp. 561–571, 2012.

[3] M. P. Nova, E. E. Caicedo, J. A. Espinosa, L. D. L. Malagón, S. L. Guzmán, and C. G. González, “Tinitus: Avances en fisiopatología y tratamiento,” Archivos de Neurociencias, vol. 11, no. 2, pp. 112–122, 2006.

[4] J. M. Espinosa-Sánchez, T. Heitzmann-Hernández, and J. A. López-Escaméz, “Tratamiento farmacológico de los acúfenos: Mucho ruido y pocas nueces,” Revista de Neurología, vol. 59, no. 4, pp. 164–174, 2014.

[5] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[6] J. A. Henry, T. L. Zaugg, P. J. Myers, and M. A. Schechter, “Using therapeutic sound with progressive audiologic tinnitus management,” Trends Amplification, vol. 12, no. 3, pp. 188–209, 2008.

[7] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[8] J. A. Henry, T. L. Zaugg, P. J. Myers, and M. A. Schechter, “Using therapeutic sound with progressive audiologic tinnitus management,” Trends Amplification, vol. 12, no. 3, pp. 188–209, 2008.

[9] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[10] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[11] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[12] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[13] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[14] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[15] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[16] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[17] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[18] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[19] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[20] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[21] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[22] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[23] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.

[24] J. J. Eggermont and L. E. Roberts, “The neuroscience of tinnitus,” Neurosci. Front., vol. 11, no. 2, pp. 112–122, 2006.
[31] A. T. Nguyen, W. P. Hetrick, B. F. O’Donnell, and C. A. Brenner, “Abnormal beta and gamma frequency neural oscillations mediate auditory sensory gating deficit in schizophrenia,” *J. Psychiatr. Res.*, vol. 124, pp. 13–21, 2020.

[32] X. Gosepath, B. Nafe, E. Ziegler, and W. Mann, “Neurofeedback in der therapie des tinnitus,” *HNO*, vol. 49, no. 1, pp. 29–35, 2001.

[33] S. Schenk, K. Lamm, H. Gündel, and K.-H. Ladwig, “Neurofeedback-gestütztes EEG-a-und EEG-β-training,” *HNO*, vol. 53, no. 1, pp. 29–38, 2005.

[34] S. Vanneste, P. V. D. Heyning, and D. D. Ridder, “Tinnitus: A large VBM-EEG correlational study,” *PLoS One*, vol. 10, no. 3, 2015, Art. no. e011522.

[35] M. Moazami-Goudarzi, L. Michels, N. Weisz, and D. Jeannerod, “Temporo-insular enhancement of EEG low and high frequencies in patients with chronic tinnitus. QEEG study of chronic tinnitus patients,” *BMC Neurosci.*, vol. 11, no. 1, pp. 1–12, 2010.

[36] T. R. Yang, M. Huang, N. Zeng, and X. Liu, “A review on transfer learning in EEG signal analysis,” *Neurocomputing*, vol. 421, pp. 1–14, 2021.

[37] R. Santana, L. Marti, and M. Zhang, “GP-based methods for domain adaptation: Using brain decoding across subjects as a test-case,” *Genet. Program. Evolvable Machines*, vol. 20, no. 3, pp. 385–411, 2019.

[38] M. Dai, S. Wang, D. Zheng, R. Na, and S. Zhang, “Domain transfer multiple Kernel learning for classification of EEG motor imagery signals,” *IEEE Access*, vol. 7, pp. 49951–49960, 2019.

[39] J. Zhang, Y. Wang, and S. Li, “Converge: Subject mental workload classification using Kernel spectral regression and transfer learning techniques,” *Cogn. Technol. Work*, vol. 19, no. 4, pp. 587–605, 2017.

[40] J. J. Bird, J. Kobyliarz, D. R. Faria, A. Ekárt, and E. P. Ribeiro, “Cross-domain learning of EEG signal processing: EEG and EMG,” *IEEE Access*, vol. 8, pp. 54789–54801, 2020.

[41] Y.-P. Lin, “Constructing a personalized cross-day EEG-based emotion-classification model using transfer learning,” *IEEE J. Biomed. Health Inform.*, vol. 24, no. 5, pp. 1255–1264, May 2020.

[42] S. J. Pan and Q. Yang, “A survey on transfer learning,” *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[43] G. Wilson and D. J. Cook, “A survey of unsupervised deep domain adaptation,” *ACM Trans. Intell. Syst. Technol.*, vol. 11, no. 5, pp. 1–46, 2020.

[44] Z. Liu, Y. Li, L. Yao, J. McAuley, and S. Dixon, “Rethink, revisit, revise: A spiral reinforced self-revised network for zero-shot learning,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 38, pp. 13452–13457, 2011.

[45] Z. Liu et al., “Face to predict: Predicting consumer choices with structured facial and behavioral traits embedding,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 5, pp. 1345–1357, Apr. 2017.

[46] Z. Liu, L. Yao, and Z. He, “Non-local self-attentive autoencoder for cross-view zero-shot learning,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pp. 1415–1424, 2017.

[47] W. Han et al., “Cross-subject EEG signal recognition using deep domain adaptation network,” *IEEE Access*, vol. 7, pp. 128273–128282, 2019.

[48] Z. Liu, Y. Li, L. Yao, and F. Nie, “Agglomerative neural networks for multiview clustering,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 2842–2852, Jul. 2022.

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

[50] C. Y. Hsu et al., “A survey on deep transfer learning,” in *Proc. Int. Conf. Artif. Intell. Knowl. Eng.*, pp. 270–279, 2015.

[51] C. Weiss, T. M. Khoshgofha, and D. Wang, “A survey of transfer learning,” *J. Big Data*, vol. 3, no. 1, pp. 1–40, 2016.

[52] Z. Liu, Y. Li, L. Yao, and F. Nie, “Agglomerative neural networks for multiview clustering,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 2842–2852, Jul. 2022.