

Side-Aware Meta-Learning for Cross-Dataset Listener Diagnosis With Subjective Tinnitus

Zhe Liu¹, Yun Li², Lina Yao³, *Senior Member, IEEE*, Molly Lucas, Jessica J. M. Monaghan, and Yu Zhang

Abstract—With the development of digital technology, machine learning has paved the way for the next generation of tinnitus diagnoses. Although machine learning has been widely applied in EEG-based tinnitus analysis, most current models are dataset-specific. Each dataset may be limited to a specific range of symptoms, overall disease severity, and demographic attributes; further, dataset formats may differ, impacting model performance. This paper proposes a side-aware meta-learning for cross-dataset tinnitus diagnosis, which can effectively classify tinnitus in subjects of divergent ages and genders from different data collection processes. Owing to the superiority of meta-learning, our method does not rely on large-scale datasets like conventional deep learning models. Moreover, we design a subject-specific training process to assist the model in fitting the data pattern of different patients or healthy people. Our method achieves a high accuracy of 73.8% in the cross-dataset classification. We conduct an extensive analysis to show the effectiveness of side information of ears in enhancing model performance and side-aware meta-learning in improving the quality of the learned features.

Index Terms—Electroencephalography, subject-independent, cross-dataset, meta-learning, tinnitus.

I. INTRODUCTION

WITH the development of digital technologies, digital processing has been the core point of the next generation of tinnitus therapy [1]. State-of-the-art digital processing relies on the improvement of artificial intelligence

Manuscript received 26 April 2022; revised 6 July 2022; accepted 17 August 2022. Date of publication 23 August 2022; date of current version 29 August 2022. (Corresponding author: Yun Li.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted in part by the University College London Ethics Committee under Approval No. 2039/002 and in part by the National Research Ethics Service Greater Manchester West Ethics Committee under Approval No. 15/NW/0133 and IRAS Project No. 168221, and performed in line with the Declaration of Helsinki.

Zhe Liu is with the Jiangsu Provincial Engineering Laboratory of Pattern Recognition and Computational Intelligence, School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, China (e-mail: zheliu912@gmail.com).

Yun Li and Lina Yao are with the School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: yun.li5@unsw.edu.au; lina.yao@unsw.edu.au).

Molly Lucas is with the Department of Psychiatry and Behavioral Sciences, Stanford University, Palo Alto, CA 94305 USA (e-mail: molly.v.lucas@gmail.com).

Jessica J. M. Monaghan is with the National Acoustic Laboratories, The Australian Hearing Hub, Macquarie University, Sydney, NSW 2109, Australia (e-mail: jessica.monaghan@gmail.com).

Yu Zhang is with the Department of Bioengineering, Lehigh University, Bethlehem, PA 18015 USA (e-mail: yuzi20@lehigh.edu).

Digital Object Identifier 10.1109/TNSRE.2022.3201158

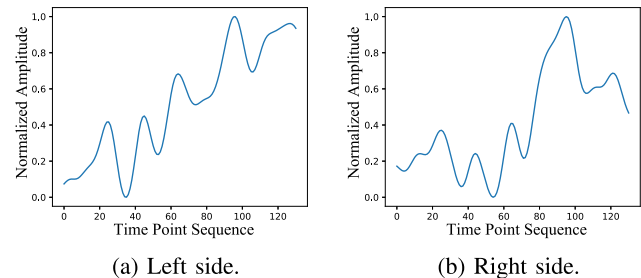


Fig. 1. Illustration of signal difference on (a) the left and (b) the right sides of the same subject. x- and y-axis denote time point sequence and normalized signal strength, respectively.

to physiological sensors, which enables the personalization of the patient-centered therapies [2], [3]. The most representative and commonly used artificial technology applied to improve tinnitus therapies is machine learning [4]. For example, machine learning has been widely applied in the analysis of electroencephalogram (EEG) [5], auditory brain-stem response (ABR) [6], and functional magnetic resonance imaging (fMRI) [7]. In particular, EEG can be an effective and inexpensive data source to analyze the neural feedback of tinnitus patients [8].

Extensive EEG-based machine learning research [6], [8], [9], [10], [11] has shown that machine learning is able to discover differences in signals between tinnitus patients and healthy people by learning patterns from given (i.e., training) datasets. Therefore, EEG dataset can be a reliable data source that enables machine learning models to assist in diagnosing tinnitus. While some research [8], [9], [12] focuses on learning the pattern difference of the existing tinnitus and control subjects, there exist significant variance between subjects [6]. When between-subject variance is high, these models may only work when the testing signals are sampled from known subjects or if their distributions closely mirror those seen within the training dataset. To enhance the model robustness, especially with regards to handling new subjects, some other research [6], [13] enables the model to be aware of the subject variance. By understanding the subject variance in signals, models can mitigate the corresponding negative influences in the prediction. However, in real-world scenarios, many EEG datasets [5], [6], [8] may collect data using different experimental environments and thus build data with diverse formats. The different data formats may cause an extremely significant domain shift. Therefore the majority of models are still limited to a single dataset and may not be able to be generalized across distinct datasets.

In the other related fields, it has been shown that the cross-dataset domain shift will heavily impair model performance in classifying EEG signals [14], [15], [16], [17], [18]. For example, unsupervised domain adaptation [16] and transfer learning [17] show good performance on cross-dataset EEG-based emotion classification. However, these methods rely on large-scale datasets like common deep learning methods. Meanwhile, some advanced technologies in machine learning can be applied to help tackle the cross-dataset domain shift in EEG. One such technique, model-agnostic meta-learning has been widely applied in computer vision [19], natural language processing [20], and recommender systems [21]. The model-agnostic meta-learning is compatible with any traditional machine learning method and does not rely on large-scale datasets. It can learn a set of generalized basic model parameters based on existing datasets and then be quickly adapted to new datasets by fitting a few samples [22], which effectively prevents domain shift among different datasets. However, directly using these methods may not fully capture the characteristic information of tinnitus signals. As shown in [Figs. 1 \(a-b\)](#), there exists a difference between the EEG signals corresponding to sounds presented to the left and right ears. Traditional methods may confuse this side difference with the difference in subject variance. This may lead models to learn false information that these signals are from two different subjects, which will impair model performance and cause inaccurate conclusions in the further analysis of the learned representation.

In this paper, we propose a domain-specific meta-learning, named Side-aware Meta-learning (SMeta), for tinnitus diagnosis. SMeta utilizes the subject difference of left/right ears and takes the difference as an auxiliary indicator to assist meta-learning. To conduct subject-specific meta-learning, we also propose subject-independent task training for SMeta, which allows episode-wise subject-specific training. We further implement SMeta based on autoencoder and design two model variants, i.e., Side-aware Meta-AutoEncoder (SMeta-AE) and Side-aware Meta-Siamese-AutoEncoder (SMeta-SAE). Both variants outperform state-of-the-art methods. Our major contributions in this paper can be summarized as follows:

- We propose a novel Side-aware Meta-learning for the domain-specific tinnitus diagnosis. In the cross-dataset tinnitus classification, our model outperforms state-of-the-art methods by 2.8%, 3.5%, and 5.0% on Negative-F1 score, Positive-F1 score, and accuracy, respectively.
- We propose a subject-specific task sampling and an auxiliary side predictor to conduct side-aware meta-learning. We also design a sliding window and down-sampling policy to align the cross-dataset signals. We implement both SMeta and conventional meta-learning on autoencoders. SMeta achieves an increase up to 1.4%, 6.7%, and 3.8% on Negative-F1 score, Positive-F1 score, and accuracy respectively, demonstrating the superiority of SMeta.
- Our method yields a 2.5%/5.0% improvement in left/right ear prediction accuracy. We conduct extensive experiments and ablation studies to show the robustness against the selection of hyper-parameters and good explainability.

II. RELATED WORK

A. EEG-Based Tinnitus Diagnosis

About 10%-15% of humans suffer from tinnitus, making tinnitus a common disorder [23]. However, due to many possible causes, such as head injury, stress, etc., as well as different symptoms, such as hearing loss, noise trauma, etc., there is currently no universally effective clinical method for subjective tinnitus diagnosis and treatment [24]. Because tinnitus is often associated with changes in the brain, many researchers have proposed that the assessment of abnormal neural activity as assessed by EEG signals may aid in the diagnosis of tinnitus.

In the early stages, evaluation of EEG signals and further diagnosis were usually done by clinical specialists [25], [26]. Gosepath *et al.* [25] first attempted to use neurofeedback to assist tinnitus therapy. With the decreased activity of EEG observed, all patients claimed tinnitus relief. The researchers then further used statistical tools to quantitatively analyze EEG data and help appraise patients' recovery condition [12], [27]. For example, Weiler *et al.* [12] proved the correlation between tinnitus and alpha, delta, and theta bands, by comparing their *z*-scores with control subjects. Milner *et al.* [27] analyzed EEG signals quantitatively by sequentially performing high- and low-pass filtering, independent component analysis (ICA), fast Fourier transform, and mean absolute amplitudes calculation of different frequency bands. Although they have made significant progress in diagnosing tinnitus using EEG signals, their approach suffers from inconsistent results that may lead to different diagnoses or conclusions. The reason is that their research relies heavily on the judgment of human experts, the analysis is case-specific, and there is no uniform standard.

Recent research uses machine learning and deep learning to reduce reliance on experts and mitigate the influence of personal factors in the diagnosis process [8], [9]. Sun *et al.* [8] segmented EEG data to enrich training sets and use Support Vector Machine (SVM) for tinnitus classification. Li *et al.* [9] also used SVM as the classifier. They pre-processed EEG signals by first transforming the signals into the frequency domain and then performing nonlinear cosine mapping. The machine-learning-based methods can achieve better or comparable performance than traditional diagnoses accomplished by experts, demonstrating the effectiveness of introducing machine learning into EEG analysis.

However, the testing samples of this methods [8], [9] may belong to the same subjects as training samples. The subject-dependent sampling strategy can result in similar training and testing samples and thus lead to inflated performance. To address the problem, some efforts are made to conduct subject-independent experiments, which aims to distinguish tinnitus patients from control subjects. For example, Wang *et al.* [13] adopted Fast Fourier Transform (FFT) to obtain multiple views of features from EEG signals, utilized Multi-view Intact Space Learning (MISL) to obtain latent representations, and classified samples with the Least Squares Support Vector Machine (LS-SVM).

B. Cross-Dataset EEG Research in Related Fields

In addition to its application in tinnitus, EEG has also attracted much attention in other fields, where advanced deep learning methods are used [28], [29], [30], [31], [32]. For example, DeepSleepNet [28] used convolutional neural networks (CNN) and bidirectional-long short-term-memory (LSTM) to automatically score sleep stage based on EEG signals. Zhang *et al.* [29] and Hartmann *et al.* [31] both employed generative models, i.e., Auto-encoder and Generative adversarial networks (GANs), respectively, to synthesize EEG samples for downstream tasks, such as brain activity recognition, EEG super-sampling, or data recovery. However, their models are optimized on large-scale datasets, limiting their deployment in more practical scenarios.

To address this issue, some cross-dataset EEG studies are proposed [14], [15], [18]. They use the knowledge learned in other datasets to assist learning in the target dataset, naturally solving the problem of insufficient data. However, due to the non-stationary nature of EEG data and varying distributions across subjects and datasets, the performance of directly applying deep learning models to cross-dataset problems may decrease significantly. Cimtay *et al.* [14] adopted CNN to avoid manual feature extraction and find latent features, yielding impressive cross-subject and cross-dataset accuracy. Xu *et al.* [15] proposed online pre-alignment to eliminate cross-dataset variability. Some studies [16], [17] further adopted the domain adaptation or transfer learning to better reuse learned knowledge in new datasets. For example, Lan *et al.* [16] employed domain adaptation to reduce the inter-subject as well as inter-datasets variance and train and test the classifier on different datasets. In contrast, Lin *et al.* [17] integrated transfer learning and principal component analysis (PCA) to reduce the variance.

C. Meta-Learning Work

In recent years, deep learning has achieved great success in various scenarios, e.g., computer vision [30], [33], natural languages processing [34], applications in health domains [6], [35]. However, the advances in deep learning are based on large-scale datasets, which is not available in many scenarios of medical applications. Meta-learning [36], simulating humans learning how to learn, is proposed to address the problem. It aims to learn novel concepts quickly through accumulated knowledge learned from other tasks.

A typical solution of meta-learning, called *metric-based* methods, is to learn a metric or distance function to measure the similarity between new inputs and prior samples [37], [38], [39], [40], [41]. For example, Koch *et al.* [37] proposed to use Siamese Neural Networks to generate pairwise similarity between inputs and thus transfer prediction ability to data from unknown distributions. While Snell *et al.* [38] learned a new metric space and calculated the similarity between inputs and learned prototypes of classes in the space. Despite the advances achieved, the metric-based meta-learning fails to utilize knowledge from new tasks to update the meta-model.

Recently, some studies focus on *model-based* meta-learning to learn task representations to alter internal state of mod-

els [42], [43], [44], [45], [46]. The dynamic state can reflect task-specific knowledge and can thus be used to predict new inputs. For example, MANN [42] was a neural network with augmented external memory to encode and retrieve information of new data sequentially for prediction. Mishra [44] *et al.* integrated temporal convolutions and soft attention to accumulate information and pinpoint memories, respectively. A drawback of model-based meta-learning is that it requires extra memory to remember information from previous inputs.

More related to our models are *optimization-based* meta-learning [19], [47], [48], [49], [50], pioneered by Andrychowicz *et al.* [47]. These models adjust the optimization process to fulfill fast gradient propagation with few training samples. Andrychowicz *et al.* first proposed considering the optimization procedure as a learning problem and utilizing LSTMs to replace hand-designed optimizers. Subsequently, Finn [19] *et al.* proposed MAML, which attracts much attention as it is simple and effective. MAML samples multiple tasks and then adopts a meta-training, meta-testing, and fine-tuning procedure to learn initial parameters. These parameters can then be fine-tuned quickly to fit new tasks.

In this paper, we adapt MAML to train our model. The reasons are three-fold: 1) MAML can learn generalizable knowledge for better cross-dataset transfer; 2) the fine-tuning process in MAML can make full use of the side information; 3) the tinnitus datasets are tiny, MAML can enrich the datasets by constructing diverse tasks.

III. METHODOLOGY

A. Problem Definition

Given a source dataset S and a target dataset T with different data formats, $x_i \in S \cup T$ denotes an EEG signal. Specifically, x_i^s and x_i^g are signals sampled from the source and the target dataset, respectively. Each signal consists of n time points $x_i = \{o_j : j \in [1, n]\}$, has a corresponding label y_i^t for tinnitus diagnosis, and has a side label y_i^d to record the left/right ear where the signal is collected. $y_i^t = 1$ denotes positive subject (i.e., tinnitus patient) and $y_i^t = 0$ denotes negative subject (i.e., control subject). $y_i^d = 1$ denotes right ear and $y_i^d = 0$ denotes left ear. Our goal is to learn a unified encoder $f_{en}(x_i) \rightarrow e_i$ and a unified classifier $f_c(e_i) \rightarrow y_i^t$ that are effective for both source and target datasets, where e_i denotes the extracted feature of x_i . To learn tinnitus-specific information of left/right ear, we design a side predictor $f_p(e_i) \rightarrow y_i^d$ to predict the side information. To regularize the learned feature, we use a decoder $f_{de}(e_i) \rightarrow x_i$ to restrain e_i from over-fitting classification information. The model overview is shown in Fig. 2.

B. Dataset Pre-Processing

Different data formats may cause the problems that signals from source dataset S and target dataset T sample different numbers of sampled time points (i.e., sampling frequency), use different filter policies in the same time duration or have divergent strength range due to the different experimental equipment. Suppose that the signals that have the same time duration in source and target datasets are with n_s and n_g time

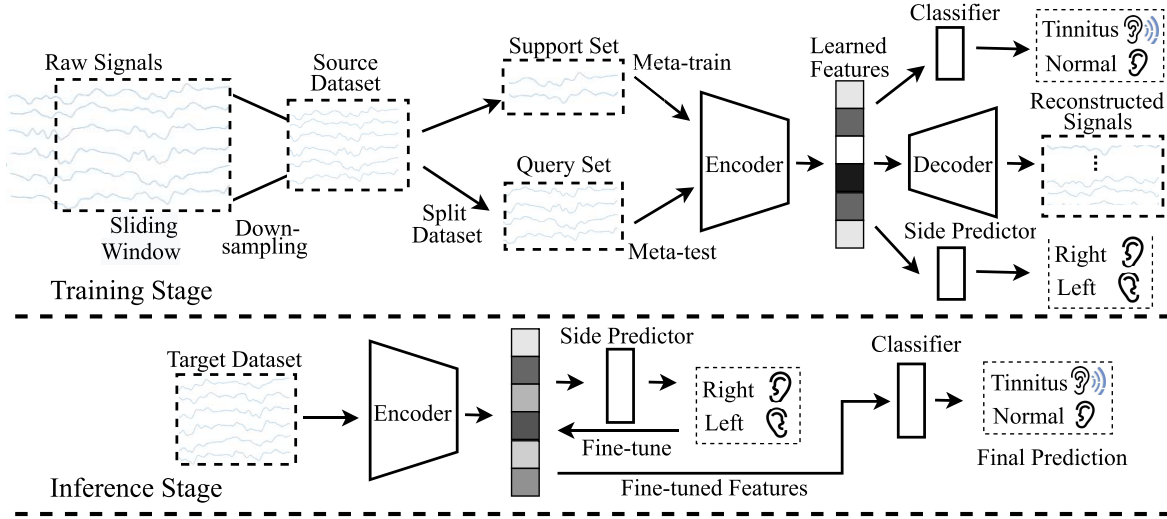


Fig. 2. Model overview. The upper part shows the training stage of SMeta, and the lower part draws the data flow of the inference stage. In our experiments, signals from the target dataset have fewer time points than those from source datasets. Therefore, the training stage first uses down-sampling to align the data format of the source dataset to the target dataset. Then, the down-sampled source dataset will be split into two sub-sets for meta-training and meta-testing, respectively. Both meta-training and meta-testing not only learn to reconstruct raw signals and classify signals but also learn tinnitus-specific information, i.e., the side information of left/right ears. The inference stage uses the trained encoder from the training stage and further fine-tunes the encoder based on side predictor fitting left/right ears. We take fine-tuned features for the final prediction.

points ($n_s \geq n_g$), respectively. We can use sliding window to cut source signals x_i^s into slices $\{x_1^o, x_2^o, \dots\}$ that have the same time duration as x_i^g . Then, we can use down-sampling to align the length of time points in the source dataset to those in the target dataset as follows:

$$\bar{o}_j^s = \begin{cases} \frac{1}{l} \sum_{k \in [jl-l+1, jl]} o_k^s & j \in [1, m] \\ \frac{1}{l+1} \sum_{k \in [jl+j-m-l, jl+j-m+1]} o_k^s & j \in [m+1, n_g] \end{cases} \quad (1)$$

$$s.t. \quad l = \lfloor n_s/n_g \rfloor \quad m = n_s - l * n_g \quad (2)$$

where o_k^s denotes k^{th} original time point in the sliced source signal x_i^s ; \bar{o}_j^s denotes the new averaged time point for the sliced source signal; the window size is n_s ; $\lfloor n_s/n_g \rfloor$ denotes the floor of n_s/n_g .

After sliding and down-sampling, \bar{o}_j^s will be the new time points in source signals, where the front m values in new signals of the source dataset will be the mean values of l time points; the last $(n_g - m)$ values will be the mean values of $(l + 1)$ time points.

Then, we use min-max normalization to separately limit the signal strengths to the same value range as follows:

$$\hat{o}_j^s = \frac{\bar{o}_j^s - \min \bar{o}_j^s}{\max \bar{o}_j^s - \min \bar{o}_j^s} \quad (3)$$

$$\hat{o}_j^g = \frac{o_j^g - \min o_j^g}{\max o_j^g - \min o_j^g} \quad (4)$$

where o_j^g denotes the time points of the signal from the target dataset (i.e., $x_i^g \in T$; \hat{o}_j^s and \hat{o}_j^g represent the normalized time points in the source signal and the target signal, respectively.

We let \hat{x}_i^s and \hat{x}_i^g be the new source and target signals constituted by \hat{o}_j^s and \hat{o}_j^g , respectively. \hat{x}_i^s will have the same

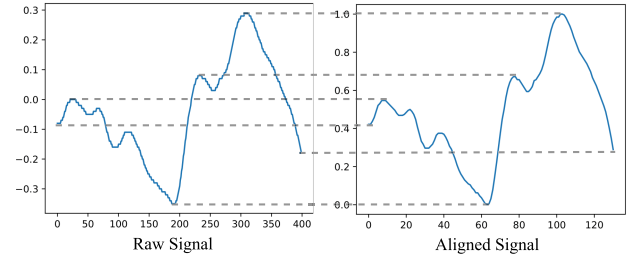


Fig. 3. Illustration of the same signal trend in (left) the raw signal and (right) the aligned signal, i.e., the normalized and down-sampled sliding window. x-axis denotes the time points and y-axis denotes the signal strength.

length (i.e., n_g) as the signal \hat{x}_i^g for the target dataset. Note that when $n_s < n_g$, we can use the same operations to align the target dataset to the source dataset by sliding window and down-sampling.

As shown in Figs. 3, we take the 1) raw signal and 2) the normalized down-sampled signal as an example. From the parallel dashes, we can observe that the down-sampling and min-max normalization will squeeze the signal strength to a unified value range (0, 1) but not change the original signal trend or pattern.

C. Side-Aware Meta-Autoencoder

We let $\hat{S} = \{\hat{x}_1^s, \hat{x}_2^s, \dots\}$ be the normalized and down-sampled source dataset. Following the meta-learning setting [19] and the subject-independent setting [6], we propose a subject-specific task sampling strategy for tinnitus analysis. Regarding the significant subject variance, we view each subject as a task and denote a subject k in the source dataset by $\tau_k \in \hat{S}$. Then, we conduct episode-wise meta-learning.

In each episode, we sample b tasks (i.e., independent subjects) from \mathcal{S} , where b denotes batch size. For each subject τ_k , we randomly split the signals into two disjoint sets: support set τ_k^{spt} and query set τ_k^{qry} for meta-training and meta-learning, respectively. Note that $\tau_k^{spt} \cap \tau_k^{qry} = \emptyset$. In the meta-learning phase, we design three loss functions to learn different aspects of signal information: classification loss L_{cls} , reconstruction loss L_{rec} , and ear side prediction loss L_{ear} .

Classification loss L_{cls} aims to supervise the learned features to carry the information to accurately classify tinnitus patients and control subjects:

$$L_{cls} = \text{CrossEntropy}(f_c(f_{en}(\hat{x}_i^s)), y_i^s) \quad (5)$$

where CrossEntropy denotes cross entropy loss; y_i^s is the corresponding ground-truth label of \hat{x}_i^s .

We design reconstruction loss L_{rec} to enable the learned features to carry the waveform information to reconstruct the original signals, which can prevent the encoder from over-fitting class information:

$$L_{rec} = \text{MSE}(f_{de}(f_{en}(\hat{x}_i^s)), \hat{x}_i^s) \quad (6)$$

where MSE represents mean square error.

Ear side prediction loss L_{ear} tends to optimize the learned features to be aware of the ear side information among signals, which may assist in classifying subjects:

$$L_{ear} = \text{CrossEntropy}(f_p(f_{en}(\hat{x}_i^s)), y_i^d) \quad (7)$$

where y_i^d is the corresponding ground-truth ear side of \hat{x}_i^s .

Then, we can summarize the loss function for optimizing our Side-aware Meta-AutoEncoder (SMeta-AE) as follows:

$$L_{SMeta-AE} = L_{cls} + L_{rec} + L_{ear} \quad (8)$$

Based on the SMeta-AE loss, we first meta-train the model based on the support set, which virtually optimizes the model to learn a basic model. Then, we meta-test the model based on the query set, which summarizes the overall gradients of tasks to truly optimize the model and thus obtains generalized model parameters. The meta-learning can be summarized as follows:

$$\theta'_{AE}(\tau^{spt}) \leftarrow \theta_{AE} - \alpha \sum_k \nabla_{\theta_{AE}} \mathcal{L}_{SMeta-AE}^{\tau^{spt}}(\theta_{AE}) \quad (9)$$

$$\theta_{AE}^* \leftarrow \theta_{AE} - \beta \sum_k \nabla_{\theta_{AE}} \mathcal{L}_{SMeta-AE}^{\tau^{qry}}(\theta'_{AE}(\tau^{spt})) \quad (10)$$

where θ_{AE} are model parameters of SMeta-AE; $\theta'_{AE}(\tau^{spt})$ represents the virtual parameters using support sets to meta-train the model; θ_{AE}^* are the updated parameters based on the query sets; $\nabla_{\theta_{AE}} \mathcal{L}_{SMeta-AE}^{\tau^{spt}}(\theta_{AE})$ represents the gradient calculation of SMeta-AE for loss $L_{SMeta-AE}$ using θ_{AE} as original model parameters and τ_k^{spt} as the training samples; α is the meta-learning rate for meta-training; β is the learning rates for meta-testing, respectively.

D. Siamese Autoencoder Extension

In this section, we introduce a Siamese autoencoder extension, named Side-aware Meta-Siamese-AutoEncoder (SMeta-SAE), to our basic version of autoencoder following the work of Liu *et al.* [6]. Additionally, we design a subject predictor $f_{sub}(e_i, e_j) \rightarrow (0, 1)$ to predict which subject the signals belong to. Note that we simplify the contrastive learning about subject prediction to two subjects, which is sufficient for meta-learning. To conduct pairwise training for episode-wise meta-learning, we sample two subjects each time, and fuse their support and query sets in a half-to-half way: keeping half of the sets without fusion as samples that belong to the same subjects and fusing the last half of the sets of two subjects as samples that belong to the different subjects. Then, we design the loss function for SMeta-SAE based on an adversarial training loss as follows:

$$\begin{aligned} L_{SMeta-SAE} &= L_{SMeta-AE} + L_{adv} + L_{sub} \\ &= L_{cls} + L_{rec} + L_{ear} + \gamma_{i,j} * \text{MSE}(e_i, e_j) \\ &\quad + \text{CrossEntropy}(f_{sub}(e_i, e_j), u_{i,j}) \end{aligned} \quad (11)$$

where $L_{SMeta-AE}$ denotes the sum of the common loss functions L_{cls} , L_{rec} , and L_{ear} ; $\gamma_{i,j}$ is a sign function that equals 1 when $i = j$ otherwise -1; $u_{i,j}$ is the contrastive label which equals 1 when $i = j$ otherwise 0.

L_{adv} is an adversarial loss. It aligns the learned features when they are from the same subject; otherwise it optimizes the learned features to be divergent. L_{sub} is a contrastive label that enhances the model learning abilities by comparing signals from different subjects.

Then, we can easily infer the meta-learning loss for SMeta-SAE as follows:

$$\theta'_{SAE}(\tau^{spt}) \leftarrow \theta_{SAE} - \alpha \sum_k \nabla_{\theta_{SAE}} \mathcal{L}_{SMeta-SAE}^{(\tau_{k,1}^{spt}, \tau_{k,2}^{spt})}(\theta_{SAE}) \quad (12)$$

$$\theta_{SAE}^* \leftarrow \theta_{SAE} - \beta \sum_k \nabla_{\theta_{SAE}} \mathcal{L}_{SMeta-SAE}^{(\tau_{k,1}^{qry}, \tau_{k,2}^{qry})}(\theta'_{SAE}(\tau^{spt})) \quad (13)$$

where θ_{SAE} are model parameters of SMeta-SAE.

E. Training and Inference Strategy for Meta-Learning

In the training stage, we first train the AE and SAE to get a basic model. Then, we take the trained model as initial parameters to conduct conventional meta-learning and side-aware meta-learning. We do not use random parameters as an initial model for meta-learning because meta-learning will weaken the fitting ability of models. The random initialization may cause the model under-fitting.

In the inference stage, we lack observed samples for the conventional meta-learning, so we directly use the trained model to conduct inference. For side-aware meta-learning, as shown in the lower part of Fig. 2, we use side information from the target dataset to fine-tune the model parameters for each subject as follows:

$$\theta_k \xleftarrow{\text{fine-tune}} \theta - \beta \nabla_{\theta} \mathcal{L}_{ear}^{\tau_k^{test} \in \hat{T}} \quad (14)$$

where θ_k denotes the fine-tuned model parameters of AE or SAE for subject k ; β is the learning rate; $\hat{T} = \{\hat{x}_1^g, \hat{x}_2^g, \dots\}$.

IV. EXPERIMENT

A. Experiment Setting

In the experiments, we examine the efficacy of our method using two tinnitus dataset [51], [52] through Auditory brain-stem responses (ABRs) through EEG. The tinnitus patients in the datasets are human subjects with ‘hidden hearing loss’. These patients with tinnitus show a normal audiogram but an abnormal pattern in ABRs. Only 5 tinnitus and 3 control subjects were unable to hear 16 kHz in the source dataset. ABRs are evoked potentials recorded with EEG sensors, which originate from the early stages of the auditory pathway. ABRs in the source dataset [51] were bandpass filtered (100-1500 Hz) and averaged (≥ 8000 repetitions for 90, and ≥ 6000 repetitions for 100 dB peSPL) with an 11 click/s rate. Signals in the target dataset [52] were filtered between 30-1500 Hz and by a fourth-order Butterworth filter. The activity in response to 102 dB peSPL clicks presented with a rate of 7 clicks/s was averaged (a total of 7040 presentations per ear) and corrected by a pre-stimulus baseline. The source dataset was approved by the University College London ethics committee with ethics ID number 2039/002. The declaration is the Helsinki declaration. The target dataset was approved by the National Research Ethics Service Greater Manchester West Ethics Committee with REC reference 15/NW/0133 and IRAS project ID 168221. The source dataset [51] contains 408 signals of a duration of 10 ms from 38 subjects for training, where each subject only has data of one side (left or right ear). The target dataset [52] has 80 signals of a duration of 8 ms from 40 subjects for testing, where each subject in the target dataset has records for both left and right ears, respectively. The source dataset included 33 female subjects, 15 with tinnitus (mean age 36.3 ± 2.6 years) and 18 controls (mean age 33.2 ± 1.9 years). The target dataset consisted of 20 tinnitus participants (10 female participants, mean age 25.7 ± 1.3 years) and 20 control participants (10 female participants, mean age 25.5 ± 1.3 years). Control participants were individually matched to tinnitus participants based on age (up to 18 months) and gender. All tinnitus participants in the source and target datasets were required to have chronic tinnitus (> 4 months, stable, and non-pulsatile). The averaged time point numbers are 500 and 131 for the source and target datasets, respectively. The information about symptom side can be found in the work [51], [52]. Both datasets are measured by a Medelec Synergy T-EP system (Oxford Instruments Medical) by placing Disposable electrodes (Nicolet Biomedical) on the high forehead and the ipsilateral and contralateral mastoids, where the electrode impedances are two k Ω . Set $n_s=400$ and $n_g=131$ to make signal slices have the same time duration (i.e., 8 ms).

For both SMeta-AE and SMeta-SAE, we use a 1-way 2-shot learning scheme for meta-learning following the work of Finn *et al.* [19]. We use this setting because we want to have a subject-specific fine-tuned model for each subject to ease the negative influence of subject variance on classification, and the target dataset only has two samples for each subject, i.e.,

2 samples to fit in the inference stage. The window interval is 20 when sampling signals in the source dataset. We set $\alpha=1e-3$, $\beta=1e-3$, query size as 8, and batch size $b=16$ for SMeta-AE and $b=5$ for SMeta-SAE. The epoch maximum is 200/1000 for SMeta-AE/SAE during training.

We compare our method with 9 competitive machine learning models: (a) XGBoost [53], (b) nu-SVC [54], (c) nCSP [55], (d) DeepNet [56], (e) ShallowNet [56] (f) AEXGB [29], (g) EEGNet [57], (h) AE [58], and (i) SAE [6]. We view tinnitus patients as positive samples and control subjects as negative samples. We use criteria, 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).

B. Cross-Dataset Tinnitus Diagnosis

We run each model 50 times and use the best performance as the final result. The best performance and standard deviation are shown in Table I. In the compared methods, AE and SAE achieve the highest accuracy. Meanwhile, AE obtains the highest scores in TNR, N-F1, and PPV, which means that AE has a better ability in classifying negative samples than SAE. SAE has the best scores in TPR and P-F1, which shows that SAE is effective in classifying positive samples. DeepNet gets the highest score in NPV. Overall, deep-learning-based methods have the best performance in machine learning methods. We show both conventional meta-learning and side-aware meta-learning on AE and SAE. Compared to these methods, we can observe that both meta-learning and side-aware meta-learning can effectively improve the model performance. Specifically, SMeta-based method can improve up to 1.8%, 2.8%, 1.7%, 3.5%, 5.0% in NPV, N-F1, PPV, P-F1, and ACC, respectively. The meta-based method can improve up to 2.5%, 2.3%, 3.1%, 2.5%, in TNR, N-F1, PPV, and Acc, respectively. Meta-learning mainly improves the ability to classify negative samples, while SMeta can improve the ability to classify both positive and negative samples. However, SMeta methods tend to be less stable than Meta methods, which may be caused by the over-fitting towards side information during testing.

We further show the detailed prediction of left and right ear in Table II and Table III, respectively. We can find that SMeta-based methods still achieve the best performance on both left and right ears in Acc. SMeta-SAE improves the accuracy up to 2.5% and 5.0% compared to the best comparison in left and right ear, respectively. This may be related to the different degrees of tinnitus in tinnitus subjects. For example, subject 57 is a subject with severe left-sided tinnitus, which is difficult for Meta-AE to classify. However, SMeta can better utilize side information to distinguish tinnitus than traditional Meta, and thus correctly classify subject 57. In the aspect of balanced classification ability, SMeta-based methods are suitable for the negative subject in both left and right ears. It can also well handle tinnitus patients in the right ear. While meta-learning may impair the model performance of TPR in the left ear, both Meta- and SMeta-methods can make the model classify tinnitus patients with higher precision, i.e., PPV.

TABLE I
BEST PERFORMANCE OF CROSS-DATASET TINNITUS DIAGNOSIS ON BOTH EARS (STANDARD DEVIATION)

Model	Both Sides						
	NPV	TNR	N-F1	PPV	TPR	P-F1	Acc
XGBoost	0.467(0.000)	0.525(0.000)	0.494(0.000)	0.457(0.000)	0.400(0.000)	0.427(0.000)	0.463(0.000)
Nu-SVC	0.467(0.000)	0.350(0.000)	0.400(0.000)	0.480(0.000)	0.600(0.000)	0.533(0.000)	0.475(0.000)
nCSP	0.568(0.055)	0.625(0.078)	0.595(0.058)	0.583(0.054)	0.525(0.085)	0.553(0.063)	0.575(0.053)
DeepNet	0.714(0.243)	0.125(0.042)	0.213(0.070)	0.521(0.010)	0.950(0.019)	0.673(0.009)	0.538(0.018)
ShallowNet	0.581(0.044)	0.625(0.147)	0.602(0.112)	0.595(0.033)	0.550(0.110)	0.571(0.040)	0.588(0.035)
AE-XGB	0.696(0.071)	0.400(0.170)	0.508(0.158)	0.579(0.033)	0.825(0.126)	0.680(0.041)	0.613(0.041)
EEGNET	0.605(0.081)	0.650(0.230)	0.627(0.181)	0.622(0.048)	0.575(0.159)	0.597(0.055)	0.613(0.049)
AE	0.660(0.047)	0.775(0.077)	0.713(0.032)	0.727(0.104)	0.600(0.154)	0.658(0.143)	0.688(0.056)
SAE	0.703(0.029)	0.650(0.063)	0.675(0.018)	0.674(0.027)	0.725(0.076)	0.699(0.034)	0.688(0.013)
Meta-AE	0.667(0.013)	0.800(0.036)	0.727(0.012)	0.750(0.020)	0.600(0.036)	0.667(0.014)	0.700(0.007)
Meta-SAE	0.681(0.018)	0.800(0.045)	0.736(0.014)	0.758(0.023)	0.625(0.048)	0.685(0.020)	0.713(0.010)
SMeta-AE	0.714(0.020)	0.750(0.025)	0.732(0.010)	0.737(0.012)	0.700(0.040)	0.718(0.022)	0.725(0.013)
SMeta-SAE	0.732(0.019)	0.750(0.061)	0.741(0.022)	0.744(0.040)	0.725(0.053)	0.734(0.020)	0.738(0.013)

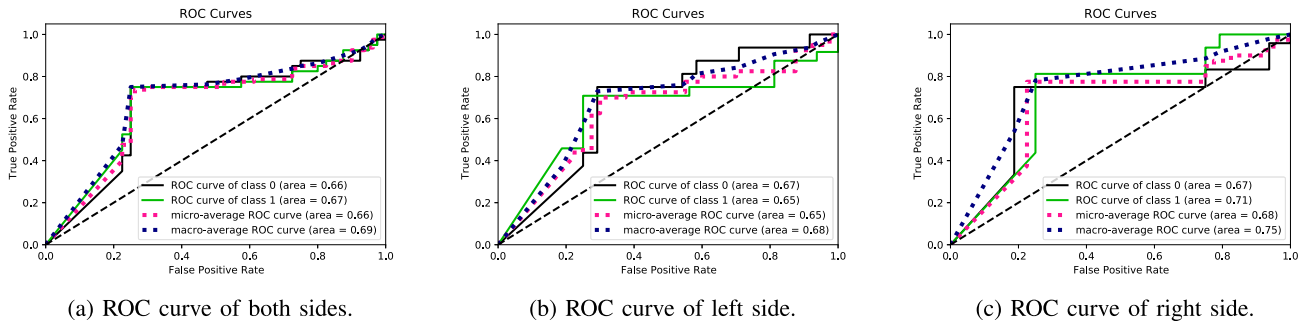


Fig. 4. ROC curves of SMeta-SAE. Class 0 and class 1 denote control subject and tinnitus patient, respectively.

TABLE II
BEST PERFORMANCE OF CROSS-DATASET TINNITUS DIAGNOSIS ON THE LEFT EAR (STANDARD DEVIATION)

Model	Left Side						
	NPV	TNR	N-F1	PPV	TPR	P-F1	Acc
XGBoost	0.455(0.000)	0.500(0.000)	0.476(0.000)	0.444(0.000)	0.400(0.000)	0.421(0.000)	0.450(0.000)
Nu-SVC	0.467(0.000)	0.350(0.000)	0.400(0.000)	0.480(0.000)	0.600(0.000)	0.533(0.000)	0.475(0.000)
nCSP	0.591(0.088)	0.650(0.109)	0.619(0.090)	0.611(0.078)	0.550(0.114)	0.579(0.091)	0.600(0.081)
DeepNet	0.500(0.245)	0.100(0.053)	0.167(0.085)	0.500(0.016)	0.900(0.026)	0.643(0.017)	0.575(0.029)
ShallowNet	0.524(0.063)	0.550(0.136)	0.537(0.110)	0.526(0.036)	0.500(0.122)	0.513(0.055)	0.525(0.044)
AE-XGB	0.615(0.143)	0.400(0.179)	0.485(0.169)	0.556(0.049)	0.750(0.146)	0.638(0.065)	0.575(0.060)
EEGNET	0.600(0.115)	0.600(0.240)	0.600(0.196)	0.600(0.054)	0.600(0.169)	0.600(0.068)	0.600(0.062)
AE	0.652(0.053)	0.750(0.078)	0.698(0.043)	0.706(0.157)	0.600(0.159)	0.649(0.161)	0.675(0.069)
SAE	0.750(0.038)	0.600(0.077)	0.667(0.035)	0.667(0.036)	0.800(0.073)	0.727(0.030)	0.700(0.023)
Meta-AE	0.682(0.018)	0.750(0.034)	0.714(0.016)	0.722(0.019)	0.650(0.037)	0.684(0.020)	0.700(0.015)
Meta-SAE	0.682(0.021)	0.750(0.038)	0.714(0.018)	0.722(0.022)	0.650(0.044)	0.684(0.024)	0.700(0.017)
SMeta-AE	0.714(0.030)	0.750(0.029)	0.732(0.023)	0.737(0.023)	0.700(0.047)	0.718(0.032)	0.725(0.026)
SMeta-SAE	0.714(0.024)	0.750(0.056)	0.732(0.024)	0.737(0.036)	0.700(0.051)	0.718(0.023)	0.725(0.019)

Referring to the work [6], AE could achieve an original accuracy of 72.2% for 100 dB SPL in the source dataset. We transfer this model to the target dataset (102 dB SPL) by meta-learning and achieve an accuracy of 73.8%. This indicates that side-aware meta-learning can utilize the low-frequency signals (30-100 Hz) and related experiments (90 dB SPL) to enhance the classification ability of the original deep learning model.

In conclusion, side-aware meta-learning can effectively improve the ability of the model to classify tinnitus patients and control subjects of different gender and ages. The side information of ears can boost the model training and fitting process to further enhance the meta-learning. It is feasible to use side-aware meta-learning to develop a reliable classification tool for ubiquitous patients with diverse physical conditions, e.g., age and gender, based on limited datasets.

TABLE III
BEST PERFORMANCE OF CROSS-DATASET TINNITUS DIAGNOSIS ON THE RIGHT EAR (STANDARD DEVIATION)

Model	Right Side						
	NPV	TNR	N-F1	PPV	TPR	P-F1	Acc
XGBoost	0.478(0.000)	0.550(0.000)	0.512(0.000)	0.471(0.000)	0.400(0.000)	0.432(0.000)	0.475(0.000)
Nu-SVC	0.467(0.000)	0.350(0.000)	0.400(0.000)	0.480(0.000)	0.600(0.000)	0.533(0.000)	0.475(0.000)
nCSP	0.545(0.068)	0.600(0.099)	0.571(0.077)	0.556(0.062)	0.500(0.109)	0.526(0.078)	0.550(0.063)
DeepNet	1.000(0.416)	0.150(0.047)	0.261(0.082)	0.541(0.011)	1.000(0.028)	0.702(0.012)	0.500(0.021)
ShallowNet	0.636(0.049)	0.700(0.166)	0.667(0.125)	0.667(0.045)	0.600(0.104)	0.632(0.033)	0.650(0.043)
AE-XGB	0.800(0.115)	0.400(0.174)	0.533(0.164)	0.600(0.039)	0.900(0.127)	0.720(0.044)	0.650(0.051)
EEGNET	0.609(0.060)	0.700(0.226)	0.651(0.170)	0.647(0.054)	0.550(0.155)	0.595(0.050)	0.625(0.046)
AE	0.667(0.049)	0.800(0.098)	0.727(0.037)	0.750(0.110)	0.600(0.160)	0.667(0.140)	0.700(0.055)
SAE	0.667(0.038)	0.700(0.063)	0.683(0.027)	0.684(0.037)	0.650(0.090)	0.667(0.054)	0.675(0.031)
Meta-AE	0.654(0.019)	0.850(0.048)	0.739(0.019)	0.786(0.034)	0.550(0.046)	0.647(0.022)	0.700(0.016)
Meta-SAE	0.680(0.021)	0.850(0.060)	0.756(0.020)	0.800(0.037)	0.600(0.060)	0.686(0.024)	0.725(0.012)
SMeta-AE	0.714(0.025)	0.750(0.037)	0.732(0.017)	0.737(0.026)	0.700(0.050)	0.718(0.028)	0.725(0.019)
SMeta-SAE	0.750(0.030)	0.750(0.073)	0.750(0.030)	0.750(0.053)	0.750(0.070)	0.750(0.035)	0.750(0.025)

TABLE IV
PARAMETER STUDY OF SMETA-AE ON BATCH SIZE b

Batch Size	Both Sides					Left Side					Right Side				
	NPV	TNR	PPV	TPR	Acc	NPV	TNR	PPV	TPR	Acc	NPV	TNR	PPV	TPR	Acc
1	0.718	0.700	0.707	0.725	0.713	0.737	0.700	0.714	0.750	0.725	0.700	0.700	0.700	0.700	0.700
2	0.690	0.725	0.711	0.675	0.700	0.700	0.700	0.700	0.700	0.700	0.682	0.750	0.722	0.650	0.700
4	0.682	0.750	0.722	0.650	0.700	0.682	0.750	0.722	0.650	0.700	0.682	0.750	0.722	0.650	0.700
8	0.660	0.775	0.727	0.600	0.688	0.652	0.750	0.706	0.600	0.675	0.667	0.800	0.750	0.600	0.700
16	0.714	0.750	0.737	0.700	0.725	0.714	0.750	0.737	0.700	0.725	0.714	0.750	0.737	0.700	0.725
32	0.714	0.750	0.737	0.700	0.725	0.714	0.750	0.737	0.700	0.725	0.714	0.750	0.737	0.700	0.725
64	0.705	0.775	0.750	0.675	0.725	0.714	0.750	0.737	0.700	0.725	0.696	0.800	0.765	0.650	0.725
128	0.682	0.750	0.722	0.650	0.700	0.714	0.750	0.737	0.700	0.725	0.652	0.750	0.706	0.600	0.675

TABLE V
PARAMETER STUDY OF SMETA-AE ON DIFFERENT LEARNING STEPS

Learning Step	Both Sides					Left Side					Right Side				
	NPV	TNR	PPV	TPR	Acc	NPV	TNR	PPV	TPR	Acc	NPV	TNR	PPV	TPR	Acc
1	0.714	0.750	0.737	0.700	0.725	0.750	0.750	0.750	0.750	0.750	0.682	0.750	0.722	0.650	0.700
2	0.711	0.675	0.690	0.725	0.700	0.722	0.650	0.682	0.750	0.700	0.700	0.700	0.700	0.700	0.700
3	0.725	0.725	0.725	0.725	0.725	0.750	0.750	0.750	0.750	0.750	0.700	0.700	0.700	0.700	0.700
4	0.698	0.750	0.730	0.675	0.713	0.714	0.750	0.737	0.700	0.725	0.682	0.750	0.722	0.650	0.700
5	0.714	0.750	0.737	0.700	0.725	0.714	0.750	0.737	0.700	0.725	0.714	0.750	0.737	0.700	0.725
6	0.667	0.800	0.750	0.600	0.700	0.682	0.750	0.722	0.650	0.700	0.654	0.850	0.786	0.550	0.700
7	0.674	0.775	0.735	0.625	0.700	0.682	0.750	0.722	0.650	0.700	0.667	0.800	0.750	0.600	0.700
8	0.682	0.750	0.722	0.650	0.700	0.714	0.750	0.737	0.700	0.725	0.652	0.750	0.706	0.600	0.675
9	0.690	0.725	0.711	0.675	0.700	0.667	0.700	0.684	0.650	0.675	0.714	0.750	0.737	0.700	0.725
10	0.653	0.800	0.742	0.575	0.688	0.682	0.750	0.722	0.650	0.700	0.630	0.850	0.769	0.500	0.675

C. Ablation Study

1) *ROC Curves*: As shown in Figs. 4 (a-c), we show the Receiver Operating Characteristic (ROC) curves of SMeta-SAE, which can achieve a high Area Under Curve (AUC) score for ROC in three conditions. When classifying both left and right sides at the same time, the areas are similar. SMeta-SAE obtains the highest AUC score in precisely classifying the right ear of tinnitus patients. Overall, SMeta-SAE shows a better AUC in the right ear than the left ear.

2) *Hyper-Parameter Study*: We use the parameters in Section IV-A as the default setting and fix the random seed as 200 to test the influence of different hyper-parameters on SMeta-AE. We show the ablation study of batch size b ,

learning steps for meta-training, meta-learning rate α , learning rate β in Table IV, Table V, Table VI, and Table VII, respectively. We can observe that our model is stable over most parameters. Only a few parameters (e.g., $b=128$ and $\alpha=9e-4$) will slightly impair the model performance, which may be caused by the bad initialization of the random seed, but the model performance is still competitive.

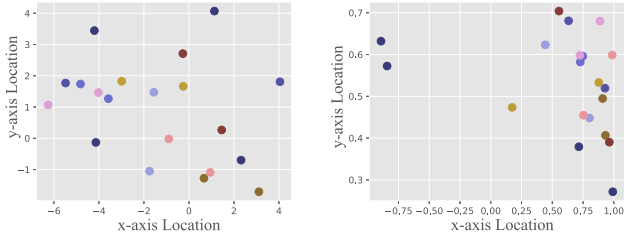
3) *Visualization of Learned Features*: We take 10 subjects from the target dataset and visualize the learned features of SMeta-SAE based on t-SNE [59] in Figs. 5 (b). The samples of the same subject are plotted with the same color. Compared to the raw signals in Figs. 5 (a), the learned features can effectively decrease the distance between samples of the same

TABLE VI
PARAMETER STUDY OF SMETA-AE ON DIFFERENT META-LEARNING RATE (α)

α	Both Sides					Left Side					Right Side				
	NPV	TNR	PPV	TPR	Acc	NPV	TNR	PPV	TPR	Acc	NPV	TNR	PPV	TPR	Acc
6.0E-04	0.698	0.750	0.730	0.675	0.713	0.714	0.750	0.737	0.700	0.725	0.682	0.750	0.722	0.650	0.700
7.0E-04	0.698	0.750	0.730	0.675	0.713	0.714	0.750	0.737	0.700	0.725	0.682	0.750	0.722	0.650	0.700
8.0E-04	0.698	0.750	0.730	0.675	0.713	0.714	0.750	0.737	0.700	0.725	0.682	0.750	0.722	0.650	0.700
9.0E-04	0.682	0.750	0.722	0.650	0.700	0.714	0.750	0.737	0.700	0.725	0.652	0.750	0.706	0.600	0.675
1.0E-03	0.714	0.750	0.737	0.700	0.725	0.714	0.750	0.737	0.700	0.725	0.714	0.750	0.737	0.700	0.725
1.1E-03	0.674	0.775	0.735	0.625	0.700	0.682	0.750	0.722	0.650	0.700	0.667	0.800	0.750	0.600	0.700
1.2E-03	0.698	0.750	0.730	0.675	0.713	0.682	0.750	0.722	0.650	0.700	0.714	0.750	0.737	0.700	0.725
1.3E-03	0.698	0.750	0.730	0.675	0.713	0.714	0.750	0.737	0.700	0.725	0.682	0.750	0.722	0.650	0.700
1.4E-03	0.707	0.725	0.718	0.700	0.713	0.700	0.700	0.700	0.700	0.700	0.714	0.750	0.737	0.700	0.725
1.5E-03	0.698	0.750	0.730	0.675	0.713	0.714	0.750	0.737	0.700	0.725	0.682	0.750	0.722	0.650	0.700

TABLE VII
PARAMETER STUDY OF SMETA-AE ON DIFFERENT LEARNING RATE (β)

β	Both Sides					Left Side					Right Side				
	NPV	TNR	PPV	TPR	Acc	NPV	TNR	PPV	TPR	Acc	NPV	TNR	PPV	TPR	Acc
6.0E-04	0.698	0.750	0.730	0.675	0.713	0.714	0.750	0.737	0.700	0.725	0.682	0.750	0.722	0.650	0.700
7.0E-04	0.689	0.775	0.743	0.650	0.713	0.714	0.750	0.737	0.700	0.725	0.667	0.800	0.750	0.600	0.700
8.0E-04	0.690	0.725	0.711	0.675	0.700	0.737	0.700	0.714	0.750	0.725	0.652	0.750	0.706	0.600	0.675
9.0E-04	0.682	0.750	0.722	0.650	0.700	0.682	0.750	0.722	0.650	0.700	0.682	0.750	0.722	0.650	0.700
1.0E-03	0.714	0.750	0.737	0.700	0.725	0.714	0.750	0.737	0.700	0.725	0.714	0.750	0.737	0.700	0.725
1.1E-03	0.690	0.725	0.711	0.675	0.700	0.700	0.700	0.700	0.700	0.700	0.682	0.750	0.722	0.650	0.700
1.2E-03	0.682	0.750	0.722	0.650	0.700	0.682	0.750	0.722	0.650	0.700	0.682	0.750	0.722	0.650	0.700
1.3E-03	0.690	0.725	0.711	0.675	0.700	0.700	0.700	0.700	0.700	0.700	0.682	0.750	0.722	0.650	0.700
1.4E-03	0.698	0.750	0.730	0.675	0.713	0.714	0.750	0.737	0.700	0.725	0.682	0.750	0.722	0.650	0.700
1.5E-03	0.682	0.750	0.722	0.650	0.700	0.667	0.700	0.684	0.650	0.675	0.696	0.800	0.765	0.650	0.725



(a) Raw signals.

(b) Learned features.

Fig. 5. Visualization of (a) raw signals and (b) the learned features of SMeta-SAE, respectively. x-axis and y-axis represent the two-dimensional location of features in the embedding space calculated by t-SNE. We take 10 subjects from the target dataset for exhibition. Each subject (distinguished by different colors) has two signals (represented as spots).

subject, i.e., the x-axis and y-axis ranges are reduced about 5 times, which shows the effectiveness of SMeta-learning in clustering new signals if they belong to the same subjects without any prior collected data from these subjects.

D. Conclusion

We propose a novel side-aware meta-autoencoder to conduct tinnitus diagnosis across datasets for different ages and genders. We design a sliding window with down-sampling to align data, a subject-specific task sampling way for our side-aware meta-learning, and an inference stage to fit side information. We conduct extensive experiments to analyze the model performance on both ears and single ears, which is superior to conventional meta-learning. The experimental results indicate that the side information of the tested ear

is beneficial for model training. Moreover, we visualize the embedding space to show the effectiveness of our method in improving the quality of the learned features. These results indicate that it is feasible to transfer model classification ability across datasets with different data collection processes. In the future, we plan to validate our method on more challenging datasets that collect data from more divergent environments.

REFERENCES

- [1] G. D. Searchfield *et al.*, "A state-of-art review of digital technologies for the next generation of tinnitus therapeutics," *Frontiers Digit. Health*, vol. 3, Aug. 2021, Art. no. 724370.
- [2] N. Hunn, "The market for hearable devices 2016–2020," WiFore Wireless Consulting, Marketing Rep., 2016. [Online]. Available: <http://bit.ly/hearables2020>
- [3] P. Crum, "Hearables: Here come the: Technology tucked inside your ears will augment your daily life," *IEEE Spectr.*, vol. 56, no. 5, pp. 38–43, May 2019.
- [4] P. Koprinkova-Hristova, V. Mladenov, and N. K. Kasabov, "Artificial neural networks," *Eur. Urol.*, vol. 40, no. 1, p. 245, 2015.
- [5] J. Allgaier, P. Neff, W. Schlee, S. Schoisswohl, and R. Pryss, "Deep learning end-to-end approach for the prediction of tinnitus based on EEG data," in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Nov. 2021, pp. 816–819.
- [6] Z. Liu *et al.*, "Generalizable sample-efficient Siamese autoencoder for tinnitus diagnosis in listeners with subjective tinnitus," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1452–1461, 2021.
- [7] S. Shahsavarani, I. T. Abraham, B. J. Zimmerman, Y. M. Baryshnikov, and F. T. Husain, "Comparing cyclicity analysis with pre-established functional connectivity methods to identify individuals and subject groups using resting state fMRI," *Frontiers Comput. Neurosci.*, vol. 13, p. 94, Jan. 2020.
- [8] Z.-R. Sun *et al.*, "Multi-view intact space learning for tinnitus classification in resting state EEG," *Neural Process. Lett.*, vol. 49, no. 2, pp. 611–624, Apr. 2019.

- [9] P.-Z. Li, J.-H. Li, and C.-D. Wang, "A SVM-based EEG signal analysis: An auxiliary therapy for tinnitus," in *Proc. Int. Conf. Brain Inspired Cogn. Syst.* Springer, 2016, pp. 207–219.
- [10] M. Saeidi *et al.*, "Neural decoding of EEG signals with machine learning: A systematic review," *Brain Sci.*, vol. 11, no. 11, p. 1525, Nov. 2021.
- [11] Y. Emami and C. Bayrak, "EEG analysis of evoked potentials of the brain to develop a mathematical model for classifying tinnitus datasets," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, May 2017, pp. 379–384.
- [12] E. W. Weiler, K. Brill, K. H. Tachiki, and D. Schneider, "Neurofeedback and quantitative electroencephalography," *Int. Tinnitus J.*, vol. 8, no. 2, pp. 87–93, 2002.
- [13] S.-J. Wang, Y.-X. Cai, Z.-R. Sun, C.-D. Wang, and Y.-Q. Zheng, "Tinnitus EEG classification based on multi-frequency bands," in *Proc. Int. Conf. Neural Inf. Process.* Guangzhou, China: Springer, 2017, pp. 788–797.
- [14] Y. Cimtay and E. Ekmekcioglu, "Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset EEG emotion recognition," *Sensors*, vol. 20, no. 7, p. 2034, Apr. 2020.
- [15] L. Xu, M. Xu, Y. Ke, X. An, S. Liu, and D. Ming, "Cross-dataset variability problem in EEG decoding with deep learning," *Frontiers Hum. Neurosci.*, vol. 14, p. 103, Apr. 2020.
- [16] Z. Lan, O. Sourina, L. Wang, R. Scherer, and G. R. Müller-Putz, "Domain adaptation techniques for EEG-based emotion recognition: A comparative study on two public datasets," *IEEE Trans. Cogn. Dev. Syst.*, vol. 11, no. 1, pp. 85–94, Mar. 2019.
- [17] Y.-P. Lin, "Constructing a personalized cross-day EEG-based emotion-classification model using transfer learning," *IEEE J. Biomed. Health Informat.*, vol. 24, no. 5, pp. 1255–1264, May 2020.
- [18] S. Rayatdoost and M. Soleymani, "Cross-corpus EEG-based emotion recognition," in *Proc. IEEE 28th Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2018, pp. 1–6.
- [19] C. Finn, P. Abbeel, and S. Levine, "Model-agnostic meta-learning for fast adaptation of deep networks," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 1126–1135.
- [20] A. Obamuyide and A. Vlachos, "Model-agnostic meta-learning for relation classification with limited supervision," in *Proc. 57th Annu. Meeting Assoc. Comput. Linguistics*, 2019, pp. 5873–5879.
- [21] H. Bharadhwaj, "Meta-learning for user cold-start recommendation," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.
- [22] T. Hospedales, A. Antoniou, P. Micaelli, and A. Storkey, "Meta-learning in neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 9, pp. 5149–5169, Sep. 2022.
- [23] B. Langguth, P. M. Kreuzer, T. Kleinjung, and D. De Ridder, "Tinnitus: Causes and clinical management," *Lancet Neurol.*, vol. 12, no. 9, pp. 920–930, Sep. 2013.
- [24] D. A. Hall *et al.*, "Systematic review of outcome domains and instruments used in clinical trials of tinnitus treatments in adults," *Trials*, vol. 17, no. 1, pp. 1–19, Dec. 2016.
- [25] K. Gosepath, B. Nafe, E. Ziegler, and W. J. Mann, "Neurofeedback in der therapie des tinnitus," *HNO*, vol. 49, no. 1, pp. 29–35, Jan. 2001.
- [26] S. Schenk, K. Lamm, H. Gündel, and K.-H. Ladwig, "Neurofeedback-gestütztes EEG- α - und EEG- β -training," *HNO*, vol. 53, no. 1, pp. 29–38, 2005.
- [27] R. Milner, M. Lewandowska, M. Ganc, K. Cieřla, I. Niedziałek, and H. Skarzyński, "Slow cortical potential neurofeedback in chronic tinnitus therapy: A case report," *Appl. Psychophysiol. Biofeedback*, vol. 41, no. 2, pp. 225–249, Jun. 2016.
- [28] A. Supratak, H. Dong, C. Wu, and Y. Guo, "DeepSleepNet: A model for automatic sleep stage scoring based on raw single-channel EEG," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 11, pp. 1998–2008, Nov. 2017.
- [29] X. Zhang, L. Yao, D. Zhang, X. Wang, Q. Z. Sheng, and T. Gu, "Multi-person brain activity recognition via comprehensive EEG signal analysis," in *Proc. 14th EAI Int. Conf. Mobile Ubiquitous Syst., Comput., New. Services*, Nov. 2017, pp. 28–37.
- [30] Z. Liu, Y. Li, L. Yao, X. Wang, and G. Long, "Task aligned generative meta-learning for zero-shot learning," in *Proc. 35th AAAI Conf. Artif. Intell.*, 2021.
- [31] K. Gregor Hartmann, R. Tibor Schirmermeister, and T. Ball, "EEG-GAN: Generative adversarial networks for electroencephalographic (EEG) brain signals," 2018, *arXiv:1806.01875*.
- [32] Z. Liu, Y. Li, L. Yao, J. J. M. Monaghan, and D. McAlpine, "Disentangled and side-aware unsupervised domain adaptation for cross-dataset subjective tinnitus diagnosis," 2022, *arXiv:2205.03230*.
- [33] Y. Li, Z. Liu, L. Yao, and X. Chang, "Attribute-modulated generative meta learning for zero-shot learning," *IEEE Trans. Multimedia*, early access, Dec. 29, 2021, doi: [10.1109/TMM.2021.3139211](https://doi.org/10.1109/TMM.2021.3139211).
- [34] Y. Gu *et al.*, "Domain-specific language model pretraining for biomedical natural language processing," *ACM Trans. Comput. Healthcare*, vol. 3, no. 1, pp. 1–23, Jan. 2022.
- [35] Y. Li, Z. Liu, L. Yao, and Z. He, "Non-local self-attentive autoencoder for genetic functionality prediction," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 2117–2120.
- [36] D. K. Naik and R. J. Mammone, "Meta-neural networks that learn by learning," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, vol. 1, Jun. 1992, pp. 437–442.
- [37] G. Koch *et al.*, "Siamese neural networks for one-shot image recognition," in *Proc. ICML Deep Learn. Workshop*, Lille, France, vol. 2, 2015, pp. 1–30.
- [38] J. Snell, K. Swersky, and R. S. Zemel, "Prototypical networks for few-shot learning," 2017, *arXiv:1703.05175*.
- [39] F. Sung, Y. Yang, L. Zhang, T. Xiang, P. H. S. Torr, and T. M. Hospedales, "Learning to compare: Relation network for few-shot learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1199–1208.
- [40] O. Vinyals *et al.*, "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 3630–3638.
- [41] V. G. Satorras and J. B. Estrach, "Few-shot learning with graph neural networks," in *Proc. Int. Conf. Learn. Represent.*, 2018, pp. 1–13.
- [42] A. Santoro, S. Bartunov, M. Botvinick, D. Wierstra, and T. Lillicrap, "Meta-learning with memory-augmented neural networks," in *Proc. 33rd Int. Conf. Mach. Learn.*, 2016, pp. 1842–1850.
- [43] T. Munkhdalai and H. Yu, "Meta networks," in *Proc. 34th Int. Conf. Mach. Learn.*, 2017, pp. 2554–2563.
- [44] N. Mishra, M. Rohaninejad, X. Chen, and P. Abbeel, "A simple neural attentive meta-learner," 2017, *arXiv:1707.03141*.
- [45] Y. Duan, J. Schulman, X. Chen, P. L. Bartlett, I. Sutskever, and P. Abbeel, "RL²: Fast reinforcement learning via slow reinforcement learning," 2016, *arXiv:1611.02779*.
- [46] J. X. Wang *et al.*, "Learning to reinforcement learn," 2016, *arXiv:1611.05763*.
- [47] M. Andrychowicz *et al.*, "Learning to learn by gradient descent by gradient descent," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3981–3989.
- [48] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, Toulon, France, 2016.
- [49] E. Grant, C. Finn, S. Levine, T. Darrell, and T. Griffiths, "Recasting gradient-based meta-learning as hierarchical Bayes," 2018, *arXiv:1801.08930*.
- [50] C. Finn, A. Rajeswaran, S. Kakade, and S. Levine, "Online meta-learning," in *Proc. 36th Int. Conf. Mach. Learn.*, 2019, pp. 1920–1930.
- [51] R. Schaeffe and D. McAlpine, "Tinnitus with a normal audiogram: Physiological evidence for hidden hearing loss and computational model," *J. Neurosci.*, vol. 31, no. 38, pp. 13452–13457, 2011.
- [52] H. Guest, K. J. Munro, G. Prendergast, S. Howe, and C. J. Plack, "Tinnitus with a normal audiogram: Relation to noise exposure but no evidence for cochlear synaptopathy," *Hearing Res.*, vol. 344, pp. 265–274, Feb. 2017.
- [53] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2016, pp. 785–794.
- [54] D. J. Crisp and C. J. Burges, "A geometric interpretation of v-SVM classifiers," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, pp. 244–250.
- [55] J. Olias, R. Martín-Clemente, M. A. Sarmiento-Vega, and S. Cruces, "EEG signal processing in MI-BCI applications with improved covariance matrix estimators," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 5, pp. 895–904, May 2019.
- [56] R. T. Schirmermeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Aug. 2017.
- [57] 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, 2018, Art. no. 056013, [Online]. Available: <http://stacks.iop.org/1741-2552/15/i=5/a=056013>
- [58] K. N. S. Kumar, D. Shivashankar, and K. Keshavamurthy, "Bio-signals compression using auto-encoder," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 11, no. 1, p. 424, Feb. 2021.
- [59] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.