Generalizable Sample-Efficient Siamese Autoencoder for Tinnitus Diagnosis in Listeners With Subjective Tinnitus

Zhe Liu, *Member, IEEE*, Lina Yao, *Member, IEEE*, Xianzhi Wang, *Member, IEEE*, Jessica J. M. Monaghan, *Member, IEEE*, Roland Schaette, Zihuai He, and David McAlpine

***Abstract*— Electroencephalogram (EEG)-based neurofeedback has been widely studied for tinnitus therapy in recent years. Most existing research relies on experts' cognitive prediction, and studies based on machine learning and deep learning are either data-hungry or not well generalizable to new subjects. In this paper, we propose a robust, data-efficient model for distinguishing tinnitus from the healthy state based on EEG-based tinnitus neurofeedback. We propose trend descriptor, a feature extractor with lower fineness, to reduce the effect of electrode noises on EEG signals, and a siamese encoder-decoder network boosted in a supervised manner to learn accurate alignment and to acquire high-quality transferable mappings across subjects and EEG signal channels. Our experiments show the proposed method significantly outperforms state-of-the-art algorithms when analyzing subjects' EEG neurofeedback to 90dB and 100dB sound, achieving an accuracy of 91.67%-94.44% in predicting tinnitus and control subjects in a subject-independent setting. Our ablation studies on mixed subjects and parameters show the method's stability in performance.**

***Index Terms*— EEG, subject-independent, siamese autoencoder, domain alignment, trend descriptor, tinnitus.**

I. Introduction

TINNITUS is a type of phantom perception caused by neural activities related to disorder of the auditory system. It is a common disease, covering over 17% of the general population and up to 33% of the elderly [1], and has been widely studied over the last decades. Despite, tinnitus is frequently associated with unrecoverable hearing loss, it can also appear with hidden hearing loss [2], [3], a clinical condition in which hearing thresholds are normal, but patients complaint for difficulties in listening in challenging conditions [4]. Currently, the diagnosis of tinnitus largely relies on patients' cognitive reactions to questionnaires and auditory tests [5].

Extensive experiments and studies on exploring the causes of tinnitus have led to the widely accepted opinion that tinnitus may be triggered by temporarily stressful and annoying situations, but turned into a permanent symptom by an unknown mechanism in central auditory pathways [6], [7]. This mechanism enlarges and reinforces the relationship between the unpleasant situations and the tinnitus, and turns out to be a persisting state at last. Some research proves that patients' responses to tinnitus will decrease if they turn their attention from tinnitus to other irrelevant tasks [8]. This discovery suggests that we may alleviate and terminate the aberrant neural activity in the central auditory system by breaking the association between tinnitus and negative emotions or situations by habituation. Following this idea, sound therapy aims to alleviate tinnitus via training patients using session sounds. Until now, sound controlled tinnitus therapy has proven effective [9] and become a common treatment to tinnitus [10]–[12]. Considering that the subject variances of different types of tinnitus (e.g., hidden hearing loss) and individual variability (e.g., age), which will lead to difficulty of recognizing proper sound therapy by humans, it is crucial to schedule treatment based on individual patients' feedback in such therapy [4].

Compared with previous traditional auditory tests [13], [14] that investigate patients' tinnitus based on cognitive judgement of patients, auditory brainstem response (ABR) recorded through electroencephalogram (EEG) allows obtaining real-time numerical feedback from the nervous system using non-invasive wearable devices. While the neurofeedback can be an effective data source for experts—who analyze and decide the proper sound treatment for patients manually [15]–[17]—machine learning and deep learning methods, e.g., support vector machine (SVM) [18], neural network [19], [20], and autoencoder [21], have achieved extraordinary performance in EEG-based neurofeedback analysis. Recently, generative models have shown the potential for overcoming subject variances in tinnitus neurofeedback analysis [22], [23], given its capability in domain alignment and domain transfer. As subject

Manuscript received November 23, 2020; revised April 5, 2021 and June 6, 2021; accepted June 29, 2021. Date of publication July 7, 2021; date of current version July 27, 2021. *(Corresponding author: Lina Yao.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University College London (UCL) ethics committee under Application No. 2039/002, and performed in line with the Helsinki declaration.

Zhe Liu and Lina Yao are with the School of Computer Science and Engineering, University of New South Wales, Sydney, NSW 2052, Australia (e-mail: zhe.liu@student.unsw.edu.au; lina.yao@unsw.edu.au).

Xianzhi Wang is with the School of Computer Science, University of Technology Sydney, Ultimo, NSW 2007, Australia.

Jessica J. M. Monaghan is with the National Acoustic Laboratories, Macquarie Park, NSW 2113, Australia, and also with the Department of Linguistics, Macquarie University, Sydney, NSW 2109, Australia.

Roland Schaette is with the University College London (UCL) Ear Institute, London WC1X 8EE, U.K.

Zihuai He is with the School of Medicine, Stanford University, Stanford, CA 94305 USA.

David McAlpine is with the Department of Linguistics, Macquarie University, Sydney, NSW 2109, Australia.

Digital Object Identifier 10.1109/TNSRE.2021.3095298

variance can be viewed as characteristic information of a domain, generative models can learn how to transfer the information from one domain to another. Besides, generative models can adopt domain alignment to embed the samples from different subjects into a unified space and thus achieve better classification. However, these methods rely on a large amount of labeled data while tinnitus datasets are usually small-scaled, and they fail to well-combine the deep learning methods with hand-engineering features.

This paper introduces a novel hand-engineered descriptor and an automatic representation learning model named Siamese Autoencoder for small-scale datasets and subject-independent experiments (i.e., no testing subjects in the training set). Our designed features achieve remarkable performance in distinguishing tinnitus patients from control subjects. They can be used for either determining whether a patient has recovered to a healthy state or predicting the usefulness of certain sound in neurofeedback-based sound therapy for tinnitus treatment. We make the following contributions in this paper:

- We design a novel Siamese Autoencoder with extra auxiliary accuracy loss, domain loss and alignment loss for better autoencoder optimization, domain transfer, and domain alignment, respectively. We further propose a trend descriptor that can reduce the effect of electrode noise to complement the autoencoder representations. The proposed two features, in combination, achieve good generalization in different experiments.
- We introduce a new split method, i.e., the Anchor Split training procedure, for our Siamese Autoencoder. The new split method contains the required information for domain alignment, domain transfer, and class prediction, and is a more suitable split method than conventional randomly cross-validation split methods during training.
- We experimentally show our approach outperforms state-of-the-art algorithms by 12.5% and 5.56% under two different conditions, respectively. We also study the impact of our proposed two features on the final performance and present ablation studies on the hyper-parameters and sound classification.

II. Related Work

Much research has been conducted about EEG-based neurofeedback neural activity training to alleviate brain system disorders. This treatment has been widely used in treating epilepsy [24] and attention deficit hyperactivity disorder [25], and has achieved outstanding performance. Tinnitus researchers similarly used neurofeedback to monitor EEG status and applied brain training to keep brain maintaining in a healthy state.

The early work on judging the health of neurofeedback mainly relied on the cognitive judgment of experts with little statistics information. The first attempt of neurofeedback training on tinnitus treatment was carried out by Gosepath *et al.* [15]. They researched on 40 patients and 15 control subjects. The treatment protocol was to increase the Alpha-band activity and decrease the Beta-band activity of EEG, which was considered to be related to tinnitus. The experiment lasted for seven years, and all the patients expressed a significant decrement of their tinnitus by the Tinnitus Questionnaire (TQ). Schenk *et al.* [17] further replicated the experiment in 40 patients. They conducted a stress-test while monitoring the EEG of patients and control subjects, and the experiment results showed that all the patients scored tinnitus with less annoying and stressful in TQ by tinnitus after training. Among the patients, 23 of 40 patients succeed in increasing Alpha-band activities, but 13 of them failed to decrease Beta-band activities. The experiment confirmed the effectiveness in treating tinnitus with neurofeedback-based EEG training but also showed that simple lifting of tinnitus patients' EEG did not equal to the healthy EEG state.

Researchers further began to use some statistical tools to help judge whether patients' EEG recover to a healthy state. Weiler *et al.* [16] extracted the power value of Alpha-band, Beta-band, Theta-band and Delta-band activities and compared the z-score of tinnitus patients with 20 control subjects. The protocol was to ameliorate the waves of patients to have a similar z-score with the other observed 20 control subjects. The experiment showed that tinnitus might be correlated to the above all four bands. Milner *et al.* [26] applied slow cortical potential (SCP) in neurofeedback tinnitus treatment, which was a descriptor of overall cortical distribution shifts in neural activities. The case report of this SCP-based neurofeedback training showed that the reduction of frequency in Delta-band and Theta-band could effectively decrease the tinnitus loudness and pitch. These studies showed that tinnitus might have a complicated relationship with EEG patterns, which was hard for simple models to catch. The subject variance might lead to the different patterns of the EEG. All the above-mentioned research stayed on the stage of relying on the cognitive judgments of experts and case-specific analysis. Therefore, the research did not have a clear criterion, and the experiment results differed from each other.

Recently, researchers sought the help of machine learning and deep learning methods for a solid criterion in the analysis of tinnitus. Some efforts aim to distinguish tinnitus patients from control subjects by machine learning. Sun *et al.* [20] extracted Principal Components Analysis (PCA), Fast Fourier Transformation (FFT), and frequency-domain statistical features for analysis. Similarly, Li *et al.* [18] preprocessed data in the frequency domain, and further extracted the features by cosine mapping and main-phase computing. Both of the works received good performance in the experiments. However, these studies [18], [20] were subject-dependent, which meant that some of the test samples come from the same subjects in training. Then, short-time sampling from the same subjects would produce some similar samples, so subject-dependent experiments may contain similar samples in both train and test samples, which would overestimate the performance of models. Wang *et al.* [27] studied the subject-independent experiments in classifying tinnitus patients from control subjects. It adopted FFT and concatenated the multi-view information from multiple channels and bands, which achieved good performance with the least squares SVM in a dataset of 29 volunteers.

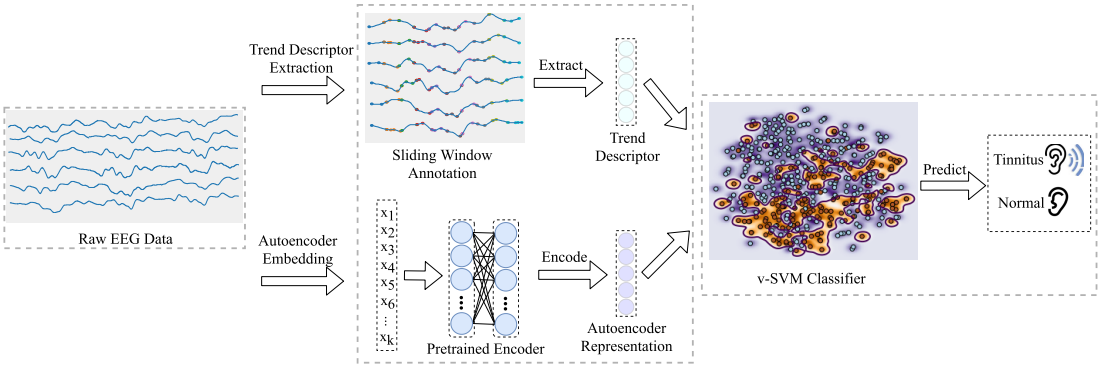


Fig. 1. Model overview. We create two low-dimensional representations of raw data by a) using a trend descriptor and b) encoding the raw data with an encoder from the Siamese Autoencoder. The sliding window annotation highlights the peaks and troughs in consecutive sliding windows in different colors. For the encoder input, x_k denotes the kth time point in an EEG signal X_n. We depict the trend descriptor and the encoded representation with blue and purple circles, respectively. v-SVM takes the combination of two features to predict whether the EEG data belong to a tinnitus patient or a healthy subject.

Unfortunately, these tinnitus-related work only considered how to distinguish patients from control subjects in a quiet environment: control subjects would not hear anything, but tinnitus patients would hear the phantom sound, which might lead to a clear difference in EEG. Moreover, their model failed to consider to help sound therapy for tinnitus, which needed models to predict tinnitus patients from control subjects while hearing a sound. Therefore, their models could not handle the job of neurofeedback-based sound therapy for tinnitus.

Other EEG study fields have attracted many advanced models and analysis tools. Some studies [28] utilized deep learning to predict the sleep stages and activities in subject-independent experiments; others [22], [29] applied generative models to generate low-dimension features; finally, recent studies [30] adopted domain transfer by generative models to enhance the EEG analysis. However, these powerful generative models would be limited in the data scale, which needed large-scale datasets to optimize the massive network parameters.

In summary, most studies [31], [32] only took statistics features (e.g. power value) and relied on the large scale dataset for training; in contrast, most tinnitus EEG datasets had a limited scale. Little work has been done to provide a reliable subject-independent tool to assist sound therapy for tinnitus. Therefore, we propose a novel trend descriptor that contains better overall shape information, and we design a Siamese Autoencoder intended for small-scale datasets and subject variance learning in sound therapy.

III. Methodology

A. Model Overview

Fig. 1 shows an overview of our proposed model with feature construction and class prediction as two main steps. The model takes the raw EEG signals as inputs and generates two sources of features: a trend descriptor (which separates a single signal into sliding windows and summarizes the trend in each) and an encoder of the Siamese Autoencoder (which learns subject differences and aligns samples from different subjects into a unified domain). The classifier operates on these two feature sources to predict the class: tinnitus or control. Throughout this section, we define the EEG signal set as $X = \{X_1, X_2, \ldots, X_n\}$, where X_i denotes a signal for i.

B. Trend Descriptor Extraction

We propose a lower fineness trend descriptor to solve the electrode noise problem in EEG signals. The trend descriptor first finds the peak and trough points in sliding windows and then collects the magnitude and number of peaks and troughs in each window.

First, we introduce our lower fineness extreme point searching method. Given an arbitrary input X_i, we define x_k as the kth time point in the original input X_i. Let α be window size and β be sliding step, we first use sliding windows to obtain sub-windows q_w to construct the sub-window set $Q_i = \{q_1, q_2, \ldots, q_w\}$ to represent X_i as inputs for trend descriptor analysis. More specifically, we define δ as a searching scope to judge the trend of a time point, and μ as a hyperparameter which lowers the fineness of searching extreme points and ignores the abnormal fluctuation caused by electrode noise.

Then, we have the sign function for time point x_k:

$$Flow(x_k) = \begin{cases} Up & (p > d + \mu) \\ Stable & (d + \mu \geq p \geq d - \mu) \\ Down & (p < d - \mu) \end{cases} \quad (1)$$

where p, d denote the largest point and the smallest point in the scope $\{x_{k+1}, x_{k+2}, \ldots, x_{k+\delta}\}$, respectively.

Then, the peaks and troughs in the sliding window are readily identified by observing when the sign function changes from Up to Down, or vice versa. We let the set of peaks and troughs be E, and we can express the trend character of an arbitrary sub-window q_w by a 4-dimension vector composed of Up/Down degree, and Up/Down times $Trend(q_w)$:

$$\begin{aligned} R &= \{E_j : E_j < E_{j+1}, E_j \in q_w\} \\ D &= \{E_j : E_j > E_{j+1}, E_j \in q_w\} \\ Trend(q_w) &= \{|R|, \sum_{E_i \in R} abs(E_i - E_{i+1}), \\ &\quad |D|, \sum_{E_i \in D} abs(E_i - E_{i+1})\} \end{aligned} \quad (2)$$

Algorithm 1 Trend Descriptor Extraction With Default Values: Window Size $\alpha = 50$, Step $\beta = 50$, Search Scope $\delta = 10$, $\mu = 2$

Require: EEG signal set $X = \{X_1, X_2, \ldots, X_n\}$

for $i \in [1, n]$ **do**
  Split X_i into slices $Q_i = \{q_1, q_2, \ldots, q_w\}$
  for $j \in [1, w]$ **do**
    for $m \in [1, \alpha]$ **do**
      Sign $x_m \in q_j$ by Eq. (1)
    end for
    Calculate $Trend(q_j)$ by Eq. (2)
  end for
  Concatenate and obtain $Trend(X_i)$ by Eq. (3)
end for
for $i \in [1, n]$ **do**
  Normalize $Trend(X_i)$ by Eq. (4)
end for

Since E is time-ordered, a peak can only be followed by a trough and vice-versa. Therefore, R and D denote the prior point sets of peak and trough points to help locate the peaks/troughs in sub-window q_w. We denote by $|R|$, $|D|$ the number of elements in R, D, and by abs the absolute value. Then, we can obtain the trend descriptor of a signal $Trend(X_i)$ by concatenating the trend descriptor of sub-windows of Q_i:

$$Trend(X_i) = \{Trend(q_w) : q_w \in Q_i\} \tag{3}$$

The dimension of the trend descriptor will be 4 times the sub-window number, depending on sliding-window size and step. We show the procedure in Algorithm 1. Finally, we apply z-score normalization:

$$\begin{aligned} Trend(X_i) &= Normalize(\{Trend(q_i) : q_i \in Q_i\}) \\ &= \{\frac{Trend(q_i) - \overline{Trend(q_i^X)}}{\sigma(Trend(q_i^X))} : q_i \in Q_i\} \end{aligned} \tag{4}$$

where $q_i^X = \{q_i^{X_i} : X_i \in X\}$ denotes the set of sliding windows from the same time of all signals; $\overline{Trend(q_i^X)}$ denotes the mean value; and $\sigma(Trend(q_i^X))$ denotes the standard deviation.

C. Siamese Autoencoder Representation Learning

The proposed Siamese Autoencoder (Fig. 2) utilizes two autoencoders as two extractors to provide low-dimensional embeddings that carry out necessary information from original input sensor data supervised by reconstruction loss L_{AE}. Different from the conventional siamese methods that only use one criterion to supervise the siamese training, we use both auxiliary classifier loss and domain prediction loss as siamese loss $L_{Siamese}$ to supervise Siamese Autoencoder to correctly predict the class information and subject domain information, respectively. Moreover, to let autoencoders have the ability to project different subjects' samples into a unified domain, we adopt an align loss L_{align} to supervise the encoded embeddings of both autoencoders.

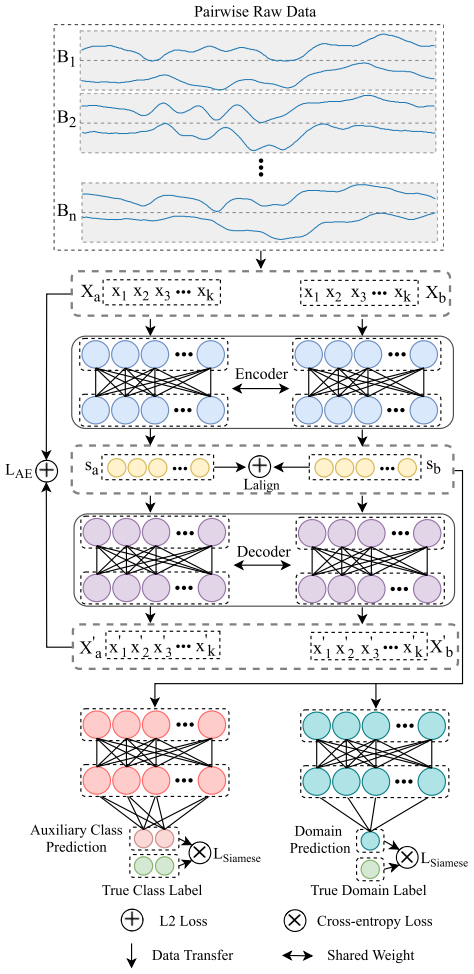


Fig. 2. Siamese Autoencoder Architecture. X_a, X_b represent the raw data from a train pair and X'_a, X'_b denote the corresponding generated data. s_a, s_b are the encoded representation. $L_{Siamese}$ consists of auxiliary class loss and domain loss.

The proposed Siamese Autoencoder takes pairwise raw data, meaning that for N input samples, we have $\binom{N}{2}$ training pairs. Therefore, we can generate enough training samples to optimize the deep-learning model even with the small-scale EEG dataset. Given a raw data set X, we compose the pairwise training batch $B = \{B_1, B_2, \ldots, B_n\}$ by $B_i = \{X_a, X_b : X_a, X_b \in X\}$ (details of construction of B in Section III-E). We define s as the autoencoder representation, and the representation pair of B_i as $S_i = \{s_a^i, s_b^i\}$, where s_a^i, s_b^i denote two corresponding representations of B_i, respectively.

In the following, we introduce the autoencoder structure, the siamese structure, and domain alignment representation learning for the proposed Siamese Autoencoder, respectively.

1) Autoencoder Structure: An autoencoder consists of two components: an *Encoder* and a *Decoder*. The *Encoder* extracts the low-dimension representation, and *Decoder* reconstructs the raw data from the representation. The representation should contain all the essential information from the raw data to allow the *Decoder* to recover the raw data.

Given an arbitrary data X_i, we construct the *Encoder* with two fully connected (FC) layers activated by the hyperbolic tangent (Tanh) function. Here, we use the Tanh function

because it can convert the input data into $\mathbb{E} \in [-1, 1]$ and help avoid the vanishing and exploding gradient problems. Then we can obtain the representation s_i of X_i by:

$$\begin{aligned} Layer(X_i) &= Tanh(W * X_i + b) \\ s_i &= Encoder(X_i) \\ &= Layer(Layer(X_i)) \end{aligned} \tag{5}$$

where W, b denote the weights and bias of the $Encoder$, respectively. Similarly, $Decoder$ consists of two FC layers, but only the first layer has the Tanh activation function. Our raw data scope is out of the range of the Tanh function, so we drop the Tanh function in the decoder output layer. Then, we have the generated raw data X_i' by:

$$\begin{aligned} X_i' &= Decoder(s_i) \\ &= W * (Layer(s_i)) + b \end{aligned} \tag{6}$$

where W, b denote weights and bias of $Decoder$, respectively. We want the generated X_i' to be as similar as possible to the raw data X_i. Therefore, we apply the L2 loss function to ensure the $Encoder$ and the $Decoder$ preserve the necessary information from the raw data. This leads to the pairwise autoencoder loss function, L_{AE}, for an arbitrary pairwise data B_i:

$$L_{AE} = \sum_{X_a \in B_i} ||X_a - X_a'||_2^2 + \sum_{X_b \in B_i} ||X_b - X_b'||_2^2 \tag{7}$$

where $||X_a - X_a'||_2^2$ denotes the L2 loss between the raw data and the generated data. L_{AE} will ensure the representation S_i carries the essential information of raw data B_i during training.

2) Siamese Structure: This section introduces the auxiliary class enhanced siamese structure for domain transfer learning. Traditional siamese structure compares two samples and determines whether they are from the same class. In our model, we compare the pairwise representations S_i and predict whether they belong to the same subject domain. This domain prediction lets $Encoder$ transfer the domain information from different subjects and learn how to generate the representation with discriminative subject identity information. We also propose an auxiliary class loss, which prevents $Encoder$ from generating white noise representations.

Given an arbitrary pairwise representations $S_i = \{s_a^i, s_b^i\}$ from $Encoder$, we define the subject domain label D as 1 when s_a^i, s_b^i are from the same subject otherwise 0. Siamese structure consists of two classifiers $Class$ and $Domain$ for predicting class labels Y and domain labels D of a pairwise representation S_i, respectively. Both classifiers contain two FC layers and take Sigmoid Function as the activation function of outputs. We apply the Cross-entropy Loss as our classification loss function. Then, siamese structure loss $L_{Siamese}$:

$$\begin{aligned} L_{cls}(y, \hat{y}) &= -y \log \hat{y} - (1 - y) log(1 - \hat{y}) \\ L_{Siamese} &= \sum_{S_i \in S} L_{cls}(Y_a, Class(s_a^i)) + L_{cls}(Y_b, Class(s_b^i)) \\ &\quad + L_{cls}(D, Domain(S_i)) \end{aligned} \tag{8}$$

where $L_{cls}(y, \hat{y})$ denotes the Cross-entropy Loss function between true label y and predicted probability $\hat{y}$; Y_a, Y_b denote the class labels of s_a^i, s_b^i, respectively. The first two terms represent the accuracy loss, which offers the label information during the training of autoencoder. Label information could assist the optimization of autoencoder towards better prediction results, and keep the key information for class prediction. The last term is the domain loss, which hopes the representations can contain the subject identity information. Subject identity information will make the representations more diverse and enhance the robustness of algorithms on predicting class labels from different subjects. Therefore, $L_{Siamese}$ will make the representations produced by autoencoder more easily to predict the class labels and the subject source-domain.

3) Domain Alignment Representation Learning: The siamese structure improves the subject information and class information carried by the representations, but it fails to align the sample domain into a unified classification space. Therefore, we propose L_{align} to adopt domain alignment of the generated representations from $Encoder$.

Given an arbitrary training batch set $S = \{S_1, S_2, \ldots, S_n\}$, where $S_i = \{s_a^i, s_b^i\}$, then we have:

$$\begin{aligned} L_{align} &= \sum_{S^i \in U} ||s_a^i - s_b^i||_2^2 - \sum_{s^i \in T} ||s_a^i - s_b^i||_2^2 \\ s.t.\ U &= \{S^i : Y_{s_a^i} = Y_{s_b^i}\}, \quad T = \{S^i : Y_{s_a^i} \neq Y_{s_b^i}\} \end{aligned} \tag{9}$$

where U, T denote the set of pairwise data from same and different classes, respectively; s_a^i, s_b^i denote the corresponding two representations of S_i. L_{align} is designed to minimize the pairwise representations difference from the same class, and maximize the representation difference from the different classes, in which way the class domain of different subjects will be adapted into a unified domain after optimization. Based on Eq. (7), Eq. (8), and Eq. (9), we could get the loss function of our Siamese Autoencoder:

$$L = L_{AE} + L_{Siamese} + L_{align} \tag{10}$$

where L_{AE} is the basic autoencoder loss function; $L_{Siamese}$ is intended for subject variance learning by domain transfer; L_{align} is designed for domain alignment in classification.

D. Classifier

We use v-Support Vector Machine (v-SVM) to predict class labels. v-SVM takes both trend descriptors and autoencoder representations as the inputs of the classifier. It aims to find a hyper-plane that maximizes the sample distance to this hyper-plane. In our tinnitus prediction, which is a binary classification problem, we define the binary labels as {-1,1}. Then, we denote the classification hyper-plane by

$$0 = w^T r + b \tag{11}$$

where w denotes the normal vector of the hyper-plane; b denotes a real number; r denotes the mapped point by kernel function (e.g. radial basis function kernel). Note, w and r are vectors with multi-dimensions. Then, we have the sign function for class label y:

$$y_{r_i} = \begin{cases} 1 & w^T r_i + b \geq 1 \\ -1 & w^T r_i + b \leq -1 \end{cases} \tag{12}$$

Now, our goal transforms into maximizing the gap between classification margin 1 and -1, and we figure out the distance of an arbitrary sample r_i to the hyper-plane by

$$\gamma_i = \frac{y_i(w^T r_i + b)}{||w||} \tag{13}$$

where $||w||$ denotes the norm of hyper-plane normal vector, and y_i denotes the class label of sample r_i.

Finally, we have the loss function for v-SVM [33]:

$$\begin{aligned} &\arg\min \frac{||w||^2}{2} - \rho\upsilon + \frac{1}{l}\sum_{i}^{l}\zeta_i \\ &s.t.\ y_i[w^T r_i + b] \geq \rho - \zeta_i, \quad i \in [1, l] \\ &\quad \rho = y_i - \varepsilon, \rho \geq 0 \\ &\quad \zeta_i \geq 0, \quad i \in [1, l] \end{aligned} \tag{14}$$

where υ denotes a hyper-parameter between 0 and 1; ε denotes the error tolerance of prediction; ρ denotes the tolerated sample sign; l denotes the sample number; ζ_i denotes the relaxation variable. The optimization details are in [33].

E. Anchor Split Training and Procedure

Training batch split is a critical factor of generative models' performance in domain adaptation [34]–[36]. Therefore, we propose Anchor split for our Siamese Autoencoder that includes subject difference and class difference in each training batch. The proposed Anchor-split method allows the autoencoder to learn the characters of subjects and classes evenly. First, we select a random sample from each subject as Anchor to ensure each training batch will contain the information of all training subjects. Second, we randomly pick a batch size number of positive and negative unique samples belonging to each Anchor without duplicated samples, respectively. The selected samples will be combined as pairs with the corresponding Anchors that carry the class information. Then, we further choose a random batch size number of positive and negative unique samples belonging to the different subjects from Anchors and merge them as pairs to transfer the subject variance information to the autoencoder. Third, we keep the selected pairs to build a data warehouse for each Anchor and shuffle them randomly. For each iteration of a training round (epoch), we will select a certain number of pairwise data from each Anchor's data warehouses to obtain a training batch. In each epoch, we repeat the three steps to relocate Anchors and rebuild the corresponding data warehouses to improve the training diversity. We illustrate the training process in Algorithm 2.

Algorithm 2 Training Procedure With Default Values: Batch Size = 128, Max Epoch = 300, Representation Dimension = 128, Learning Rate (Lr) = 0.001

Require: EEG signal set $X = \{X_1, X_2, \ldots, X_n\}$
1: Split train/test subjects
2: Extract trend descriptor by Algorithm 1
3: **while** epoch $\leq$ max epoch **do**
4: Set Anchors in train subjects
5: Build data warehouse within train subjects
6: **while** Data warehouse not empty **do**
7: Pick up a pairwise train batch
8: Calculate L by Eq. (10)
9: $W, b \leftarrow Adam(lr, L)$
10: **end while**
11: **end while**
12: Encode samples of test subjects by Eq. (5)
13: Use v-SVM to predict classes with trend descriptors and representations of test subjects by Eq. (14)

IV. Experiment

A. Experiment Setting

In the experiments, we aim to distinguish tinnitus patients without obvious hearing loss and control subjects under different sound environments, which may be caused by the hidden hearing loss. While limited ABR datasets can be obtained, our experiments is carried out on a small-scale dataset [4] under 90 dB SPL and 100 dB SPL environments. The dataset measured ABR by a Medelec Synergy T-EP system (Oxford Instruments Medical). Disposable electrodes (Nicolet Biomedical) are placed on the high forehead and the ipsilateral and contralateral mastoids. Channel number is 1 and sampling frequency is 50.0 kHz. Electrode impedances were 2 kΩ. Stimuli were 50 μs clicks presented via Telephonics TDH 49 headphones at a rate of 11 clicks/s. The bilateral signals were bandpass filtered (100-1500 Hz) and averaged (8000 repetitions for 90 dB SPL, 6000 repetitions for 100 dB SPL). All the subjects are female. Mean age of patients and control subjects are 36.3 years and 33.2 years, respectively. All participants did not have obvious hearing loss and had normal hearing (hearing thresholds $\leq$ 20 dB hearing level (HL) from 125 Hz to 8 kHz). There was no significant difference in the average hearing thresholds up to 12 kHz [4]. Therefore, in our experiments, distinguishing tinnitus patients is consistent with detecting hidden hearing loss. We conduct subject-independent experiments in distinguishing tinnitus subjects from control subjects for 90dB SPL and 100dB SPL click-evoked EEG data, respectively. We selected averaged waveforms from two tinnitus patients and two control subjects as signals for testing and used the remaining waveforms for training. This provided 216 waveforms for training and 49 waveforms for testing in the 90dB SPL data, generating 23,220 pairs of training data; we also have 148 waveforms for training and 43 waveforms for testing in 100dB SPL experiment, which contains 10,878 pairs of train samples. By this means, we generated sufficient data for training our model.

For both experiments, we define tinnitus patients as 1 (positive) and control subjects as 0 (negative), and set our Siamese Autoencoder with the same parameters: lr = 0.005, batch size = 128, max epoch = 300, representation dimension = 64; we also set window size α = 40, step $\beta = 40$, search scope $\delta = 10$, $\mu = 2$ for our trend descriptor. Due to the significant subject variance, we use the fixed test sets for the experiments to carry out the further quantitative

TABLE I
TINNITUS PATIENTS PREDICTION FROM CONTROL OBJECTS UNDER 90DB SPL

Model	0-Precision	0-F1	1-Precision	1-F1	Macro-Precision	Macro-Recall	Macro-F1 (Weighted-F1)	Accuracy
v-SVM	0.5333	0.5926	0.5556	0.4762	0.5444	0.5417	0.5344	0.5417
nCSP	0.6364	0.6087	0.6154	0.6400	0.6259	0.6250	0.6243	0.6250
CRAM	0.5202	0.6844	1.0000	0.5030	0.7601	0.6680	0.5937	0.6140
AE-XGB	0.6923	0.7200	0.7273	0.6957	0.7098	0.7083	0.7078	0.7083
EEGNet	0.8571	0.6316	0.6471	0.7586	0.7521	0.7083	0.6951	0.7083
ShallowNet	0.7059	0.8276	1.0000	0.7368	0.8529	0.7917	0.7822	0.7917
DeepNet	0.8000	0.4706	0.5789	0.7097	0.6895	0.6250	0.5901	0.6250
SiameseAE	0.8571	0.9231	1.0000	0.9091	0.9286	0.9167	0.9161	0.9167

TABLE II
TINNITUS PATIENTS PREDICTION FROM CONTROL OBJECTS UNDER 100DB SPL

Model	0-Precision	0-F1	1-Precision	1-F1	Macro-Precision	Macro-Recall	Macro-F1	Weighted-F1	Accuracy
v-SVM	0.4000	0.3077	0.5385	0.6087	0.4692	0.4750	0.4582	0.4749	0.5000
nCSP	0.5833	0.6364	0.6250	0.5556	0.6042	0.6000	0.5960	0.5960	0.6000
CRAM	0.3807	0.5514	1.0000	0.4913	0.6903	0.6628	0.5214	0.5089	0.5233
AE-XGB	0.5556	0.5882	0.6667	0.6316	0.6111	0.6125	0.6099	0.6123	0.6111
EEGNet	0.6667	0.8000	1.0000	0.7500	0.8333	0.8000	0.7750	0.7722	0.7778
ShallowNet	0.6154	0.7619	1.0000	0.6667	0.8077	0.7500	0.7143	0.7090	0.7222
DeepNet	0.8750	0.8750	0.9000	0.9000	0.8875	0.8875	0.8875	0.8889	0.8889
SiameseAE	0.8889	0.9412	1.0000	0.9474	0.9444	0.9500	0.9443	0.9446	0.9444

analysis. We compared our method with several competitive methods: (**a**) one of the v-Support Vector Machine (v-SVM), i.e., nu-SVM [37], using raw data as the baseline, (**b**) Improved Covariance Matrix Estimators (nCSP) [38], (**c**) Convolutional Recurrent Attention Model (CRAM) [19], (**d**) Autoencoder enhanced Extreme Gradient Boosting (AEXGB) [22], (**e**) Compact Convolutional Neural Network for EEG (EEGNet) [39], (**f**) Shallow Convolutional Network (ShallowNet) [40], and (**g**) Deep Convolutional Network (DeepNet) [40].

We define Macro and Weighted criteria for evaluation:

$$Macro: \frac{1}{|\mathcal{L}|}\sum_{l\in\mathcal{L}}\phi(y_l,\hat{y}_l)$$
$$Weighted: \frac{1}{\sum_{l\in\mathcal{L}}|\hat{y}_l|}\sum_{l\in\mathcal{L}}|\hat{y}_l|\phi(y_l,\hat{y}_l) \quad (15)$$

where $\mathcal{L}$ denotes the set of labels; $y_l, \hat{y}_l$ denote the predicted label and true labels; $|y_l|$ denotes the number of predicted labels which have the label l; $\phi(y_l, \hat{y}_l)$ denotes the computation of Precision, Recall, or F1-Score for true/predicted labels.

B. Results

Table I and Table II show our best performance in distinguishing tinnitus subjects from control subjects on 90dB SPL and 100dB SPL EEG data, respectively. Note, 90dB SPL experiment has the same number of positive and negative samples, so Weighted-F1 has the same value as Macro-F1. v-SVM cannot distinguish the tinnitus patients with raw data; AE-XGB shows the autoencoder can effectively improve the quality of the raw data; while traditional machine learning methods like nCSP fail to learn the difference between tinnitus and healthy EEG in subject-independent experiments. Deep learning models have better performance in both experiments—ShallowNet achieves 79.17% accuracy under 90dB SPL, and DeepNet obtains 88.89% accuracy under 100dB SPL, but neither of them has good performance in both experiments.

Our proposed method outperforms other algorithms in both experiments, which shows the siamese generative model's stronger ability to catch the subject variance and the robustness of our algorithm. We achieve 100% precision and over 0.90 F1-score in the prediction of tinnitus patients, which means our model can find the most patients' EEG accurately. The overall accuracy of our model is over 90%, which shows the eligibility in the EEG-based sound therapy.

TABLE III
TREND DESCRIPTOR (TD) AND SIAMESE AUTOENCODER (SA) INDEPENDENT PERFORMANCE UNDER 90DB AND 100DB SPL

SPL	Feature	0-F1	1-F1	Weighted-F1	Accuracy
90dB	TD	0.500	0.500	0.497	0.500
90dB	SA	0.811	0.792	0.788	0.792
100dB	TD	0.775	0.775	0.778	0.778
100dB	SA	0.762	0.667	0.709	0.722

C. Trend Descriptor and Siamese Autoencoder Analysis

We use the same data split as in former experiments to explore the impact of the trend descriptor and the Siamese Autoencoder, respectively. Specifically, we use v-SVM to show the performance of hand-crafted features and the classifier in siamese networks to show the effectiveness of autoencoder-generated features. Table III shows either of the

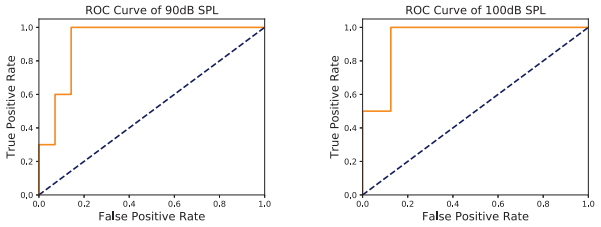
TABLE IV
PRESENTATION LEVEL PREDICTION ON MIXED SUBJECTS

Model	0-Precision	0-F1	1-Precision	1-F1	Macro-Precision	Macro-Recall	Macro-F1	Weighted-F1	Accuracy
v-SVM	0.7955	0.8642	0.8667	0.7027	0.8311	0.7684	0.7835	0.8040	0.8136
nCSP	0.6250	0.6494	0.6471	0.6197	0.6360	0.6351	0.6345	0.6345	0.6351
CRAM	0.9105	0.9379	0.9398	0.8897	0.9252	0.9058	0.9138	0.9196	0.9205
AE-XGB	0.8537	0.8974	0.8889	0.8000	0.8713	0.8366	0.8487	0.8611	0.8644
EEGNet	0.8182	0.8889	0.9333	0.7568	0.8758	0.8047	0.8228	0.8396	0.8475
ShallowNet	0.9722	0.9589	0.9130	0.9333	0.9426	0.9502	0.9461	0.9494	0.9492
DeepNet	0.8780	0.9231	0.9444	0.8500	0.9112	0.8729	0.8865	0.8958	0.8983
SiameseAE	1.0000	0.9722	0.9167	0.9565	0.9583	0.9730	0.9644	0.9664	0.9661

TABLE V
LEARNING RATE PERFORMANCE ON MEAN SCORE (STANDARD DEVIATION)

	90dB SPL				100dB SPL			
lr	0-F1	1-F1	Weighted-F1	Accuracy	0-F1	1-F1	Weighted-F1	Accuracy
0.001	0.846(0.038)	0.836(0.025)	0.841(0.031)	0.842(0.031)	0.917(0.046)	0.926(0.044)	0.922(0.044)	0.922(0.044)
0.002	0.846(0.035)	0.818(0.041)	0.832(0.038)	0.833(0.037)	0.938(0.056)	0.949(0.045)	0.944(0.050)	0.944(0.050)
0.003	0.831(0.010)	0.797(0.040)	0.814(0.023)	0.817(0.020)	0.899(0.021)	0.901(0.023)	0.900(0.022)	0.900(0.022)
0.004	0.888(0.027)	0.878(0.037)	0.883(0.031)	0.883(0.031)	0.886(0.006)	0.891(0.004)	0.889(0.000)	0.889(0.000)
0.005	0.891(0.013)	0.874(0.022)	0.882(0.017)	0.883(0.017)	0.919(0.046)	0.925(0.043)	0.922(0.044)	0.922(0.044)
0.006	0.879(0.048)	0.870(0.058)	0.875(0.053)	0.875(0.053)	0.886(0.037)	0.891(0.033)	0.889(0.035)	0.889(0.035)
0.007	0.871(0.035)	0.858(0.035)	0.864(0.035)	0.865(0.035)	0.863(0.032)	0.870(0.023)	0.867(0.027)	0.867(0.027)
0.008	0.913(0.032)	0.898(0.039)	0.906(0.035)	0.906(0.035)	0.873(0.045)	0.882(0.039)	0.878(0.041)	0.878(0.042)
0.009	0.879(0.030)	0.846(0.042)	0.862(0.036)	0.865(0.035)	0.926(0.049)	0.934(0.044)	0.931(0.046)	0.931(0.046)
0.010	0.906(0.021)	0.906(0.015)	0.906(0.018)	0.906(0.018)	0.869(0.027)	0.880(0.022)	0.875(0.024)	0.875(0.024)

two features cannot compete with their combination in performance in either experiment. Trend descriptor has better performance in 100dB SPL while merely provides useful independent information in 90dB SPL; Siamese Autoencoder can extract meaningful information from low signal-to-noise ratio data but has worse performance than trend descriptor in 100dB SPL. Compared with the previous experiment, combining the two features delivers much better performance in both experiments, meaning the features can complement and improve each other.



(a) ROC curve of 90 dB SPL. (b) ROC curve of 100 dB SPL.

Fig. 3. ROC curves.

D. Sound Classification

We further test our model in predicting the presentation level in mixed subjects. We define 90dB as label 0 and 100dB as label 1. We select three tinnitus patients and two control subjects for the test and use the remaining samples for the train. Then we have 400 waveforms for training and 56 waveforms for testing, which means 79,800 pairwise samples for training. We compare our method with the same algorithms in the prediction of tinnitus in Table IV. Prediction of sound level is easier for algorithms, while we get a baseline of 81.36%. The proposed Siamese Autoencoder only has lower 100dB precision than some of the state-of-the-art algorithms but outperforms them in other criterion scores, which reveals our model's powerful ability to analyze EEG-related tasks.

E. Ablation Study

We investigate the robustness of our Siamese Autoencoder, by showing receiver operating characteristic (ROC) curve curves and testing the impact of learning rate (lr) of Adam optimizer, batch size, and representation dimension on our model's performance, separately, with the other parameters default to $lr = 0.001$, batch size $= 128$, and representation dimension $= 64$. We run the model for 10 times and calculate the mean performance of the best five times to reduce the adverse impact caused by randomization in our Anchor split.

1) ROC Curves: As shown in Fig. 3, our model can achieve high true positive rates when the false positive rates are smaller than 0.2 in both 90 dB SPL and 100 dB SPL experiments. The model can obtain 0.3 and 0.5 of true positive rate without false positive predictions, which indicates that our model can perform better in 100 dB SPL environment. The ROC curves illustrate the effectiveness of our model in tinnitus diagnosis under different sound environments.

2) Hyper-Parameter Study: Table V, Table VI and Table VII show varying results of different learning rates, batch sizes, representation dimensions. We can observe that larger lr leads to better performance in 90dB experiments but worse performance in 100dB experiments. Batch size has only a small influence on the 100dB experiments but a

TABLE VI
BATCH SIZE PERFORMANCE ON MEAN SCORE (STANDARD DEVIATION)

	90dB SPL				100dB SPL			
Batch	0-F1	1-F1	Weighted-F1	Accuracy	0-F1	1-F1	Weighted-F1	Accuracy
64	0.825(0.027)	0.797(0.049)	0.811(0.037)	0.813(0.036)	0.924(0.021)	0.935(0.027)	0.930(0.024)	0.931(0.024)
96	0.832(0.047)	0.832(0.047)	0.832(0.043)	0.833(0.042)	0.926(0.049)	0.934(0.044)	0.931(0.046)	0.931(0.046)
128	0.846(0.038)	0.836(0.025)	0.841(0.031)	0.842(0.031)	0.917(0.046)	0.926(0.044)	0.922(0.044)	0.922(0.044)
160	0.780(0.039)	0.737(0.028)	0.758(0.033)	0.760(0.035)	0.913(0.024)	0.919(0.031)	0.917(0.028)	0.917(0.028)
192	0.797(0.028)	0.762(0.043)	0.780(0.035)	0.781(0.035)	0.911(0.022)	0.921(0.032)	0.916(0.028)	0.917(0.028)
224	0.848(0.040)	0.812(0.070)	0.830(0.054)	0.833(0.051)	0.913(0.050)	0.919(0.047)	0.917(0.048)	0.917(0.048)
256	0.774(0.045)	0.745(0.067)	0.759(0.055)	0.760(0.054)	0.900(0.019)	0.905(0.027)	0.903(0.024)	0.903(0.024)

TABLE VII
REPRESENTATION DIMENSION PERFORMANCE ON MEAN SCORE (STANDARD DEVIATION)

	90dB SPL				100dB SPL			
Dimension	0-F1	1-F1	Weighted-F1	Accuracy	0-F1	1-F1	Weighted-F1	Accuracy
16	0.770(0.027)	0.761(0.027)	0.765(0.022)	0.767(0.021)	0.897(0.022)	0.908(0.026)	0.903(0.024)	0.903(0.024)
32	0.806(0.033)	0.774(0.038)	0.790(0.034)	0.792(0.035)	0.941(0.040)	0.947(0.039)	0.944(0.039)	0.944(0.039)
48	0.796(0.019)	0.786(0.026)	0.791(0.019)	0.792(0.018)	0.913(0.050)	0.919(0.047)	0.917(0.048)	0.917(0.048)
64	0.846(0.038)	0.836(0.025)	0.841(0.031)	0.842(0.031)	0.917(0.046)	0.926(0.044)	0.922(0.044)	0.922(0.044)
80	0.839(0.025)	0.800(0.053)	0.820(0.037)	0.823(0.035)	0.912(0.030)	0.921(0.027)	0.917(0.028)	0.917(0.028)
96	0.812(0.022)	0.812(0.022)	0.812(0.021)	0.813(0.021)	0.909(0.054)	0.922(0.046)	0.916(0.049)	0.917(0.048)
112	0.828(0.017)	0.817(0.021)	0.823(0.018)	0.823(0.018)	0.910(0.028)	0.922(0.028)	0.917(0.028)	0.917(0.028)
128	0.787(0.068)	0.746(0.086)	0.767(0.072)	0.771(0.069)	0.898(0.025)	0.906(0.024)	0.903(0.024)	0.903(0.024)

larger influence on the 90dB experiments. 90dB experiments need a relatively small or a large enough batch size to allow the model to learn the representation better, while 100dB experiments need a small batch size. Our model achieves the best performance when using 32/64-dimension vectors. The too-small or too-large, e.g., 16 and 128, may be unable to express the subject information or introduce extra noise. Overall, our model is robust over most parameters.

V. CONCLUSION

We propose a novel model that integrates a lower fineness trend descriptor and a powerful Siamese Autoencoder to distinguish tinnitus patients from control subjects based on small-scale subject-independent EEG signals. We further exhibit the model effectiveness in the presentation level prediction. The experiments indicate that our work is promising to be used as a powerful criterion in the neurofeedback sound therapy, which can give a reliable suggestion to the doctors whether the EEG patterns of patients have recovered to a healthy state.

REFERENCES

[1] P. J. Jastreboff *et al.*, "Neurophysiological approach to tinnitus patients," *Amer. J. Otology*, vol. 17, no. 2, pp. 236–240, 1996.

[2] M. H. Bennett *et al.*, "Hyperbaric oxygen for idiopathic sudden sensorineural hearing loss and tinnitus," *Cochrane Database Syst. Rev.*, vol. 10, no. 1, Oct. 2012.

[3] J. J. Nelson and K. Chen, "The relationship of tinnitus, hyperacusis, and hearing loss," *Ear, Nose Throat J.*, vol. 83, no. 7, pp. 472–476, Jul. 2004.

[4] R. Schaette 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.

[5] P. H. Wilson, J. Henry, M. Bowen, and G. Haralambous, "Tinnitus reaction questionnaire: Psychometric properties of a measure of distress associated with tinnitus," *J. Speech, Lang., Hearing Res.*, vol. 34, no. 1, pp. 197–201, Feb. 1991.

[6] F. Gonzalez-Lima and H. Scheich, "Neural substrates for tone-conditioned bradycardia demonstrated with 2-deoxyglucose. II. Auditory cortex plasticity," *Behavioural Brain Res.*, vol. 20, no. 3, pp. 281–293, Jun. 1986.

[7] D. J. Strauss, W. Delb, R. D'Amelio, Y. F. Low, and P. Falkai, "Objective quantification of the tinnitus decompensation by synchronization measures of auditory evoked single sweeps," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 16, no. 1, pp. 74–81, Feb. 2008.

[8] L. C. Oatman and B. W. Anderson, "Suppression of the auditory frequency following response during visual attention," *Electroencephalogr. Clin. Neurophysiol.*, vol. 49, nos. 3–4, pp. 314–322, Aug. 1980.

[9] P. J. Jastreboff *et al.*, "Tinnitus retraining therapy (TRT) as a method for treatment of tinnitus and hyperacusis patients," *J. Amer. Acad. Audiol.*, vol. 11, no. 3, pp. 162–177, 2000.

[10] C. W. Newman and S. A. Sandridge, "A comparison of benefit and economic value between two sound therapy tinnitus management options," *J. Amer. Acad. Audiol.*, vol. 23, no. 2, pp. 126–138, Feb. 2012.

[11] D. J. Hoare, G. D. Searchfield, A. El Refaie, and J. A. Henry, "Sound therapy for tinnitus management: Practicable options," *J. Amer. Acad. Audiol.*, vol. 25, no. 1, pp. 62–75, 2014.

[12] E. Viirre, J. A. Pineda, J. D. Hestenes, and A. Vankov, "Eeg feedback controlled sound therapy for tinnitus," U.S. Patent 7 572 234, Aug. 11, 2009.

[13] T. Pan, R. S. Tyler, H. Ji, C. Coelho, A. K. Gehringer, and S. A. Gogel, "The relationship between tinnitus pitch and the audiogram," *Int. J. Audiol.*, vol. 48, no. 5, pp. 277–294, Jan. 2009.

[14] Y. Shiomi, J. Tsuji, Y. Naito, N. Fujiki, and N. Yamamoto, "Characteristics of DPOAE audiogram in tinnitus patients," *Hearing Res.*, vol. 108, nos. 1–2, pp. 83–88, Jun. 1997.

[15] 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.

[16] 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.

[17] 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.

[18] 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.* Cham, Switzerland: Springer, 2016.

[19] D. Zhang, L. Yao, K. Chen, and J. Monaghan, "A convolutional recurrent attention model for subject-independent EEG signal analysis," *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 715–719, May 2019.

[20] 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.

[21] T. Nguyen, S. Nahavandi, A. Khosravi, D. Creighton, and I. Hettiarachchi, "EEG signal analysis for BCI application using fuzzy system," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2015, pp. 1–8.

[22] 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., Netw. Services*, Nov. 2018, pp. 28–37.

[23] Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised cross-domain image generation," 2016, *arXiv:1611.02200*. [Online]. Available: http://arxiv.org/abs/1611.02200

[24] M. B. Sterman and T. Egner, "Foundation and practice of neurofeedback for the treatment of epilepsy," *Appl. Psychophysiol. Biofeedback*, vol. 31, no. 1, p. 21, Mar. 2006.

[25] N. Lofthouse, L. E. Arnold, S. Hersch, E. Hurt, and R. DeBeus, "A review of neurofeedback treatment for pediatric ADHD," *J. Attention Disorders*, vol. 16, no. 5, pp. 351–372, Jul. 2012.

[26] R. Milner, M. Lewandowska, M. Ganc, K. Cieśla, I. Niedziałek, and H. Skarński, "Slow cortical potential neurofeedback in chronic tinnitus therapy: A case report," *Appl. Psychophysiol. Biofeedback*, vol. 41, no. 2, pp. 225–249, Jun. 2016.

[27] S.-J. Wang *et al.*, "Tinnitus EEG classification based on multi-frequency bands," in *Proc. Int. Conf. Neural Inf. Process.* Cham, Switzerland: Springer, 2017.

[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] Z. Liu, Y. Li, L. Yao, X. Wang, and G. Long, "Task aligned generative meta-learning for zero-shot learning," in *Proc. 31th AAAI Conf. Artif. Intell.*, 2021, pp. 8723–8731.

[30] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, "EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals," 2018, *arXiv:1806.01875*. [Online]. Available: http://arxiv.org/abs/1806.01875

[31] A. Phinyomark, F. Quaine, S. Charbonnier, C. Serviere, F. Tarpin-Bernard, and Y. Laurillau, "EMG feature evaluation for improving myoelectric pattern recognition robustness," *Expert Syst. Appl.*, vol. 40, no. 12, pp. 4832–4840, 2013.

[32] D. Coyle, G. Prasad, and T. M. McGinnity, "A time-series prediction approach for feature extraction in a brain-computer interface," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 4, pp. 461–467, Dec. 2005.

[33] C.-C. Chang and C.-J. Lin, "Training v-support vector regression: Theory and algorithms," *Neural Comput.*, vol. 14, no. 8, pp. 1959–1977, Aug. 2002.

[34] V. K. Verma, D. Brahma, and P. Rai, "Meta-learning for generalized zero-shot learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 34, no. 4, 2020, pp. 6062–6069.

[35] S. Sankaranarayanan, Y. Balaji, C. D. Castillo, and R. Chellappa, "Generate to adapt: Aligning domains using generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8503–8512.

[36] N. Bigdely-Shamlo, G. Ibagon, C. Kothe, and T. Mullen, "Finding the optimal cross-subject EEG data alignment method for analysis and BCI," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 1110–1115.

[37] 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.

[38] 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.

[39] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *J. Neural Eng.*, vol. 15, no. 5, Oct. 2018, Art. no. 056013. [Online]. Available: http://stacks.iop.org/1741-2552/15/i=5/a=056013

[40] R. T. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, 2017. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.23730