



# Cross-Subject Tinnitus Diagnosis Based on Multi-Band EEG Contrastive Representation Learning

Chang-Dong Wang , Member, IEEE, Xi-Ran Zhu, Xueqing Zhou, Jiahong Li, Liping Lan, Dong Huang , Member, IEEE, Yiqing Zheng, and Yuexin Cai 

**Abstract**—Electroencephalogram (EEG) is an important technology to explore the central nervous mechanism of tinnitus. However, it is hard to obtain consistent results in many previous studies for the high heterogeneity of tinnitus. In order to identify tinnitus and provide theoretical guidance for the diagnosis and treatment, we propose a robust, data-efficient multi-task learning framework called Multi-band EEG Contrastive Representation Learning (MECRL). In this study, we collect resting-state EEG data from 187 tinnitus patients and 80 healthy subjects to generate a high-quality large-scale EEG dataset on tinnitus diagnosis, and then apply the MECRL framework on the generated dataset to obtain a deep neural network model which can distinguish tinnitus patients from the healthy controls accurately. Subject-independent tinnitus diagnosis experiments are conducted and the result shows that the proposed MECRL method is significantly superior to other state-of-the-art baselines and can be well generalized to unseen topics. Meanwhile, visual experiments on key parameters of the model indicate that the high-classification weight electrodes of tinnitus' EEG signals are mainly distributed in the frontal, parietal and temporal regions. In conclusion, this study facilitates our understanding of the relationship between electrophysiology and pathophysiological changes of tinnitus and provides a new deep learning

method (MECRL) to identify the neuronal biomarkers in tinnitus.

**Index Terms**—Electroencephalograph, deep learning, tinnitus, multi-dimensional, contrastive learning.

## I. INTRODUCTION

ELECTROENCEPHALOGRAPHY (EEG) is a medical technique that reads scalp electrical activity generated by brain structures. It has been shown to represent the macroscopic activity of the surface layer of the brain underneath [1], [2], [3], [4]. Owing to its non-invasive and real-time reflection of brain state [5], EEG has been widely applied in the exploration of central mechanism, diagnosis and treatment of tinnitus. Tinnitus is the self-perception of sound with the absence of an external sound source. The previous studies regarded tinnitus as a lesion of cochlea or auditory pathways. However, many imaging studies have shown that tinnitus not only involves the auditory cortex, but also becomes a kind of central plasticity encephalopathy involving a large number of non-auditory brain regions and network connections [6]. And abnormalities in temporal, frontal, parietal, cingulate gyrus and other brain regions in different frequency bands of EEG have been widely studied [7], [8]. For example, Schmidt et al. discovered that the network connectivity with the precuneus is a invariant marker of long-term tinnitus [9]. Araneda et al. found that the executive function deficit caused by the change of prefrontal cortex would be the key factor for the generation and persistence of tinnitus [10]. And other studies have shown that these central changes associated with tinnitus reflect the dysfunction of the default mode network (DMN) [11], [12]. The default mode network may regard tinnitus as a norm, which is an important reason why tinnitus persists [13]. However, due to the complex central mechanism of tinnitus and numerous influencing factors, many previous studies have not reached comprehensive consistent results. Therefore, many attempts have been made in diagnosing tinnitus based on the EEG signals by using machine learning and deep learning methods [14], [15].

For the EEG-based machine learning methods, the spectral features of the EEG signals are firstly extracted by some methods like fast Fourier transform (FFT), and then feature selection is performed depending on clinical experience and the selected

Manuscript received 21 September 2022; revised 20 January 2023 and 15 March 2023; accepted 31 March 2023. Date of publication 5 April 2023; date of current version 3 July 2023. This work was supported in part by the National Key Research and Development Program of China under Grant 2021YFF1201200, in part by NSFC under Grants 62276277, 82071062, 82271165 and 61976097, in part by the Guangdong Basic and Applied Basic Research Foundation under Grants 2022B1515120059 and 2021A1515012038, and in part by the Zhuhai Science and Technology Innovation Bureau in 2022 under Grant 222000400085. (Corresponding author: Yuexin Cai.)

Chang-Dong Wang and Xi-Ran Zhu are with the School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou 510000, China, and with the Guangdong Province Key Laboratory of Computational Science, Guangzhou 510000, China, and also with the Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, Guangzhou 510000, China (e-mail: changdongwang@hotmail.com; zhuxr3@mail2.sysu.edu.cn).

Xueqing Zhou, Jiahong Li, Liping Lan, Yiqing Zheng, and Yuexin Cai are with the Department of Otolaryngology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou 510000, China (e-mail: zhouxq33@mail2.sysu.edu.cn; ljh1995@hotmail.com; lanlp@mail2.sysu.edu.cn; zhengyiq@mail.sysu.edu.cn; caiyx25@mail.sysu.edu.cn).

Dong Huang is with the College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510000, China (e-mail: huangdonghere@gmail.com).

Digital Object Identifier 10.1109/JBHI.2023.3264521

features are further supplied to the machine learning models (e.g.  $K$ -nearest neighbors (KNN), Naive-Bayes and Support Vector Machine (SVM)) to complete classification or regression tasks [16], [17], [18]. Some methods including Random Forest [19] and XGBoost [20] take advantage of an automatic selection of optimal feature combinations and achieve promising performance. However, these machine learning methods hard-encode the EEG signals and fail to comprehend the semantic information among them. Therefore, it is difficult to further enhance the performance. On the other hand, due to the unparalleled fitting capabilities of deep neural networks and the structural similarity between neuron and real-world nerve cells, more and more studies have been conducted in developing EEG-based deep learning methods to accomplish various downstream tasks [14], [21], [22], [23]. Unfortunately, the high individual variability of EEG [24] significantly limits the generalization capabilities of these methods, which leads to wide performance gap between seen and unseen subjects [25]. That's why many EEG-based deep learning models are restricted to the laboratory stage and have seldom been able to take the step of large-scale promotion.

In order to address the above issues and deal with the problem of high individual variability of the EEG signals, we take tinnitus diagnosis as the typical task and propose a robust, data-efficient multi-task learning framework called Multi-band EEG Contrastive Representation Learning (MECRL). The MECRL framework consists of four components, namely EEG data sampler, multi-band EEG encoder, feature fusioner and linear classifier. And three layer-by-layer progressive learning tasks are designed to capture the semantics of EEG. The whole training process of MECRL consists of the representation learning process and the classification process. During the representation learning process, the EEG data sampler generates various mini-batches by randomly sampling subjects and segments from the dataset, and different sampling strategies are applied to support different auxiliary tasks. The multi-band EEG encoder is composed of the spatial module and the temporal module which integrates the multi-dimensional information (e.g. spatiotemporal domain and frequency domain) of EEG automatically and extracts discriminative multi-band features. The feature fusioner fuses the extracted multi-band features and generates a unified representation. During the classification process, the linear classifier is attached after the encoder and the whole model is fine-tuned via the supervised learning paradigm.

A tinnitus dataset is collected by recruiting 187 tinnitus patients and 80 healthy subjects from the department of Otolaryngology, Sun Yat-sen Memorial hospital, Sun Yat-sen University. Experiments are conducted on this tinnitus dataset. The results show that the proposed MECRL framework achieves superior performance over the state-of-the-art baseline models in the cross-subject tinnitus diagnosis task because it not only takes the multi-dimensional information of EEG into account, but also utilizes multiple self-supervised learning tasks to help the encoder understand the semantics of EEG and align the individual variability. Furthermore, the results demonstrate the validity of the data from high-weight electrodes in tinnitus diagnosis, which can undoubtedly help us explore the causes of chronic tinnitus.

Overall, the key contributions of this paper are as follows:

- We have collected a resting-state EEG dataset by recruiting 187 tinnitus patients and 80 healthy subjects from the department of Otolaryngology, Sun Yat-sen Memorial hospital, Sun Yat-sen University. Compared with the previous studies [14], [26], the dataset has higher spatial precision (more electrodes) and a larger number of subjects, which eases the high individual variability of EEG data when applying deep learning methods and makes it possible for us to carry out a larger range of subject-independent experiments to uncover macroscopic EEG differences between the chronic tinnitus patients and the healthy individuals. The dataset can be accessed in [drive.google.com/drive/folders/1Su1IWGyZIED-IINUTVgCY9dIL\\_1k1ip\\_](https://drive.google.com/drive/folders/1Su1IWGyZIED-IINUTVgCY9dIL_1k1ip_).
- A novel multi-task learning framework called Multi-band EEG Contrastive Representation Learning (MECRL) is proposed. The framework helps the deep residual network-based encoder understand the semantics of EEG data and align the individual variability by constructing different self-supervised auxiliary tasks. After the training process, discriminative representations will be generated by the trained model which are used to distinguish the tinnitus patients from the healthy controls accurately.
- Subject-independent comparison experiments and visualization experiments are conducted on the collected EEG dataset to demonstrate the superiority of our framework on cross-subject tinnitus diagnosis comparing with state-of-the-art methods. Meanwhile, the visualization of the key parameters of the model helps us explore the central nervous mechanism of chronic tinnitus with our medical collaborators.

The rest of this paper is organized as follows. Section II describes the proposed MECRL framework. Section III presents the details and results of the subject-independent experiments exhaustively. Section IV visualizes the key parameters of our model and discusses the pathogenic mechanism of the chronic tinnitus. Section V draws the conclusion of this paper.

## II. METHODS

The Multi-band EEG Contrastive Representation Learning (MECRL) framework consists of four components, namely EEG data sampler, multi-band EEG encoder, feature fusioner, and linear classifier. The first three components constitute the representation learning process, which is shown in Fig. 1. To help the encoder learn about EEG and align the individual variability, the following three contrastive learning tasks are designed:

- *Task A*: Because the EEG signals of tinnitus patients differ significantly from those of the healthy controls, the features extracted from different tinnitus patients should be more similar than those extracted from the healthy controls. Task A is a subject-wise task.
- *Task B*: Due to the individual variability, the features extracted from the EEG segments of the same subject should be more similar than those extracted from different subjects. Task B is designed to allow the model to identify individual biases and would not confuse them with the

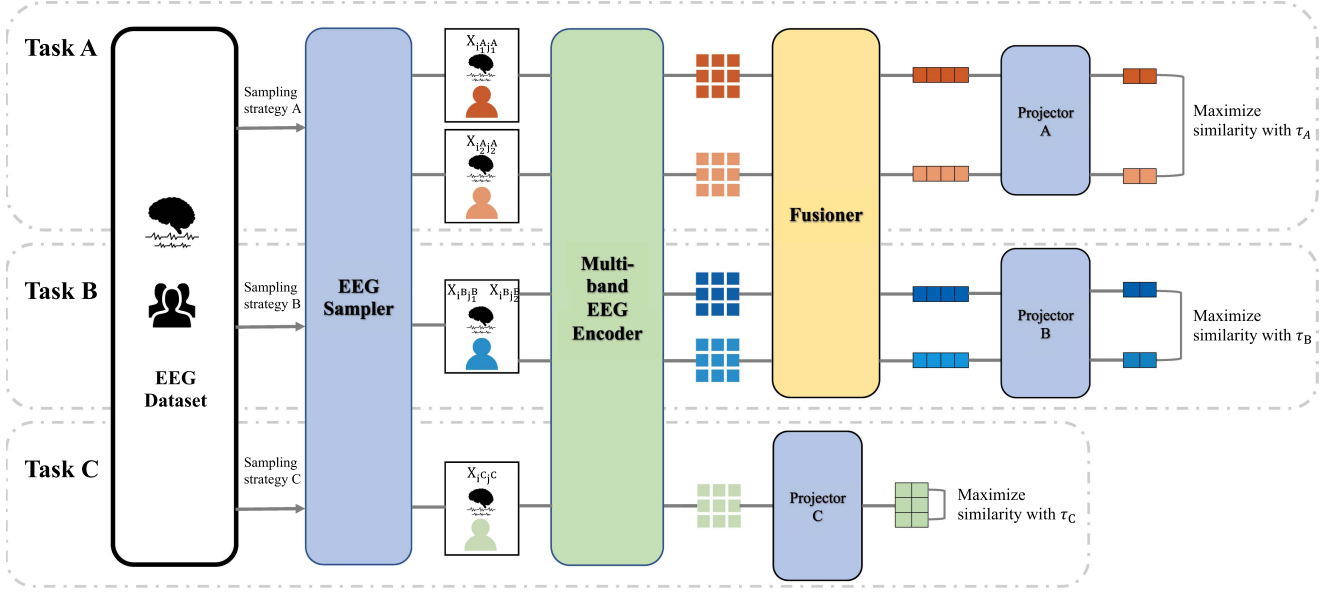


Fig. 1. Overview of the representation learning process. In order to support various auxiliary tasks, the EEG sampler uses different sampling strategies to generate mini-batches, and then the corresponding mini-batches will be fed to the model and the corresponding contrastive loss will be calculated at each step of representation learning process. By minimizing the contrastive loss, the model parameters are continuously optimized, so that the model can integrate multi-dimensional information in the EEG signals and learn discriminative representations across subjects.

general differences in tinnitus disorders. It is a segment-wise task.

- *Task C*: Because multiple clips obtained by frequency division of the same raw EEG segment are temporally consistent, the features extracted from the same segment should be more similar than those extracted from other segments. Task C is a band-wise task.

Task A is a supervised contrastive learning task [27] while task B and task C are self-supervised contrastive learning tasks [28]. The three tasks correspond to different aspects of the EEG semantics in tinnitus diagnosis, respectively. And the similarity between the positive samples of these three contrastive learning tasks is progressive, which means we can obtain a more semantically informative and uniformly distributed feature space with appropriate parameter settings, especially the temperature coefficient [29]. This enables the multi-band EEG encoder to extract higher-quality and more robust cross-subject representations. The whole training process consists of representation learning process and classification process. During the representation learning process, the multi-band EEG encoder is trained with the designed contrastive learning tasks. In order to assist each task in training, the EEG sampler takes different sampling strategies for different tasks to build differentiable mini-batches, which enriches the model in learning materials. After the representation learning process, a mapping is established from EEG segments to discriminative feature vectors. In the classification process, we add a linear classifier after the multi-band EEG encoder. The labeled data and cross-entropy loss are used to train the linear classifier and fine-tune the encoder. Finally, the well-trained model would be applied to classify the tinnitus patients and the healthy controls.

In the following subsections, we will introduce the components of the MECRL framework in detail.

### A. EEG Data Sampler

The EEG datasets tend to have a small number of individuals, which makes it difficult for the EEG-based deep learning models to capture discriminative features and learn task-related representations across subjects due to the overfitting issue. To address this issue, some studies divide EEG data into isometric segments and treat these segments as irrelevant data to enlarge the dataset. However, this practice will introduce personal variance when segments from a single subject appear multiple times on the dataset, which may confuse the model, leading them to mistake individual characteristics as universal ones. To solve this problem and support the multi-task learning framework, we design an EEG data sampler to generate mini-batches by randomly sampling subjects and segments from the dataset. Suppose that the EEG data in the tinnitus dataset is  $\{X_i^c \mid i = 1, 2, 3, \dots, N; X_i^c \in \mathbb{R}^{B \times M \times T_i}\}$ , where  $c$  means the class of the data,  $B$  is the number of bands,  $M$  is the number of electrodes, and  $T_i$  is the number of time points of the EEG signal from subject  $i$ . In order to adapt to the multi-task learning framework, the EEG data sampler adopts the following three sampling strategies to support different contrastive learning tasks.

- *Strategy A*: To explore the general difference of the EEG signals between the tinnitus patients and the healthy controls, the sampler first randomly samples equal numbers of tinnitus subjects and healthy subjects. And then a single EEG segment will be sampled from each selected subject to guarantee that data of the same individual will not

appear in the mini-batch, which helps the EEG encoder capture the general differences in the EEG signals between the tinnitus patients and the healthy controls, and avoids the introduction of individual biases.

- *Strategy B*: To help the model recognize the individual biases, the sampler first randomly samples two EEG segments from a single subject as positive samples, and then constructs negative samples by sampling segments from multiple separate subjects.
- *Strategy C*: To capture the temporal consistency amount multiple clips obtained by frequency division of the same raw EEG segment, the sampler first randomly samples clips of a raw EEG segment from a single subject as positive samples, and then constructs negative samples by sampling clips from the same subject.

By applying different sampling strategies, the EEG data sampler provides abundant learning materials to the models. Moreover, the random sampling method and the coordination of multiple sampling strategies introduce more misinformation into the training process, which makes it easier for the model to escape from the local optimum to avoid overfitting, and also makes the learned representations more robust.

## B. Multi-Band EEG Encoder

The multi-band EEG encoder takes multi-band EEG segments as inputs and transforms them into subject-independent representation related to tinnitus diagnosis. Structurally, it is a neural network model composed of multiple deep neural networks in parallel and each deep neural network is responsible for processing EEG data of a single frequency band. In order to integrate the rich spatial and temporal information in EEG data, each network consists of a regularized spatial convolution layer as spatial module to extract spatial correlation of electrode positions and a deep residual network as temporal module to explore dependence of the EEG signals in the time domain. These two modules will be described in the following subsections.

1) *Spatial Module*: Tinnitus is considered to be a long-term neurological disorder by mainstream scholars, which may be caused by abnormal neural activities involving multiple brain regions and frequency bands. To explore the functional connections among different brain regions and remove the redundant electrodes, we utilize a spatial module containing a regularized spatial convolution layer to assign weights to the different electrodes. The module can be formulated as:

$$\begin{aligned} \bar{X}_{ij}^b &= f_{spatial}(X_{ij}^b, W^b), \\ L_{region} &= \sum_b \left\| W^b \right\|_1 \end{aligned} \quad (1)$$

where  $X_{ij}^b \in \mathbb{R}^{M \times T}$  denotes the  $j$ -th EEG segment of subject  $i$  in the  $b$ -th frequency band,  $W^b \in \mathbb{R}^{K_1 \times M \times 1}$  denotes the convolution kernel of the spatial convolution in the  $b$ -th neural network and  $\bar{X}_{ij}^b \in \mathbb{R}^{K_1 \times T}$  denotes the output of the spatial module.  $T$  means the length of EEG segments and  $K_1$  represents the number of spatial convolution filters.  $f_{spatial}$  means the process of the spatial convolution.

In recent years, the EEG data tends to contain more and more electrodes with the upgrading of acquisition equipment. This presents both opportunities and challenges for EEG-based deep learning methods since most of these electrodes are redundant, which may introduce extra noise signals and even worsen the performance of the downstream tasks. In order to reduce the spatial redundancy and explore which brain regions are more indicative of tinnitus diagnosis, we utilize a  $L_1$  regularization loss (i.e.  $L_{region}$ ) on the kernel of the spatial convolution layer to force them to be sparse, which allows only valid information to pass smoothly and enter the next layer.

2) *Temporal Module*: The EEG signals dynamically change over time, which contains abundant temporal information related to various neural activities. To detect the dependence of the EEG signals hiding in time domain, we apply a deep residual network as the temporal module which is not only with strong fitting ability, but also tractable in the optimization process. The network architecture is exactly analogous to [30], and the process can be formulated as:

$$h_{ij}^b = f_{temporal}(\bar{X}_{ij}^b) \quad (2)$$

where  $h_{ij}^b \in \mathbb{R}^{K_2}$  is the feature extracted from the  $j$ -th EEG segment of subject  $i$  in the  $b$ -th frequency band, and  $K_2$  represents the dimension of the extracted feature.  $f_{temporal}^b$  represents the function map of the  $b$ -th deep residual network.

## C. Feature Fusioner

The feature fusioner is utilized to fuse the extracted multi-band features and generate a unified representation. To comprehensively consider the information obtained from each single band, attention mechanism is introduced and the process can be formulated as:

$$\bar{h}_{ij} = \sum_{b=1}^B f_{fusion}(h_{ij}^b) \cdot h_{ij}^b \quad (3)$$

where  $h_{ij}^b$  denotes the feature from the  $b$ -th frequency band, and  $f_{fusion}$  denotes the attention layer used to fuse the multi-band features. In order to enhance the expressiveness of the model, a projector composed of three layers of MLPs is introduced [31]:

$$z_{ij} = g(\bar{h}_{ij}, \delta) \quad (4)$$

where  $\delta(\ast)$  means the activation function. And it should be noted that different contrastive learning tasks have their own projectors.

## D. Representation Learning: Loss Function and Training

The three contrastive learning tasks correspond to the three different aspects of the EEG semantics. The loss function of each task could be formulated as follows:

$$\begin{aligned} L_A &= \frac{2}{N_A} \sum_{n=1}^{\frac{1}{2}N_A} L_A^n, \\ L_A^n &= -\frac{1}{\frac{1}{2}N_A - 1} \end{aligned} \quad (5)$$



$$\begin{aligned} & \times \log \frac{\sum_{k=1}^{N_A} l_{n \neq k} \cdot l_{\tilde{y}_n = \tilde{y}_k} \cdot \exp\left(\frac{\text{sim}(z_{i_n^A j_n^A}, z_{i_k^A j_k^A})}{\tau_A}\right)}{\sum_{k=1}^{N_A} l_{n \neq k} \cdot l_{\tilde{y}_n \neq \tilde{y}_k} \cdot \exp\left(\frac{\text{sim}(z_{i_n^A j_n^A}, z_{i_k^A j_k^A})}{\tau_A}\right)} \\ L_B &= -\log \frac{\exp(\text{sim}(z_{i_1^B j_1^B}, z_{i_1^B j_2^B})/\tau_B)}{\sum_{k=2}^{N_B} \exp(\text{sim}(z_{i_1^B j_1^B}, z_{i_k^B j_k^B})/\tau_B)} \quad (6) \\ L_C &= \sum_{i=1}^B -\frac{1}{B-1} \cdot \log \frac{\sum_{b=1}^B \exp(\text{sim}(h_{j_1}^i, h_{j_1}^b)/\tau_C)}{\sum_{k=2}^{N_C} \exp(\text{sim}(h_{j_1}^i, h_{j_k}^i)/\tau_C)} \quad (7) \end{aligned}$$

where  $\tau_A, \tau_B, \tau_C \in \mathbb{R}^+$  are scalar temperature parameters.  $l_{n \neq k}$  is an indicator function whose value is 1 when  $n = k$  and 0 when  $n \neq k$ .  $l_{\tilde{y}_n = \tilde{y}_k}$  returns 1 when the label of the  $n$ -th sampled subject is the same as that of the  $k$ -th sampled subject and returns 0 when the labels are different, while  $l_{\tilde{y}_n \neq \tilde{y}_k}$  is just the opposite.  $N_A, N_B, N_C$  denote the batch size of each task and  $\text{sim}(\ast)$  represents a function that measures the cosine similarity of the input vectors.

Finally, we can get the loss function of the representation learning process (denoted as  $L_{rpt}$ ) as follows:

$$L_{rpt} = L_{region} + L_A + L_B + L_C. \quad (8)$$

That is,  $L_{rpt}$  is the sum of  $L_{region}, L_A, L_B$  and  $L_C$ . During the representation learning process, the optimizer adjusts model parameters by reducing  $L_{rpt}$ , and the model can better understand the EEG data after continuous optimization.

### E. Linear Classifier

After the representation learning process, our multi-band EEG encoder gains the capacity of transforming the EEG segments into cross-subject representations. In the classification process, we attach a linear classifier after the feature fusioner to train the classifier and fine-tune the previous components. The classifier is trained via the supervised learning paradigm, using cross-entropy (i.e.  $L_{cls}$ ) as the loss function which can be formulated as:

$$\hat{y} = f_{classifier}(\bar{h}) \quad (9)$$

$$L_{cls} = -y \cdot \log \hat{y} + (1 - y) \cdot \log(1 - \hat{y}) \quad (10)$$

where  $f_{classifier}$  denotes the function map of the linear classifier,  $\hat{y}$  denotes the probability of tinnitus and  $y$  denotes the true label.

## III. EXPERIMENTS

### A. Data and Preprocessing

1) *Participants*: The study groups consist of a tinnitus group with 187 tinnitus patients (77 females and 110 males, mean age = 43.24 years, ranging from 14 to 73 years, std = 14.205) and a control group with 80 healthy subjects (37 females and 43 males, mean age = 39.84 years, ranging from 20 to 62 years, std = 14.362). All tinnitus patients have suffered tinnitus for at least 3 months. Furthermore, the tinnitus patients with Meniere's

disease, pulsatile tinnitus, central nervous system disorders, otosclerosis and the history of previous middle ear surgery are excluded. All tinnitus patients are recruited from the department of Otolaryngology, Sun Yat-sen Memorial hospital, Sun Yat-sen University. Before collecting data, routine audiological examinations including otoscopy and pure-tone audiometry are performed for all participants. Moreover, the Tinnitus Handicap Inventory (THI) questionnaire and tinnitus specific assessments including tinnitus pitch and loudness matching measurements are performed for all tinnitus patients. Tinnitus severity is assessed by the THI questionnaire which has 25 items to evaluate the self-perceived level of handicap caused by tinnitus, based on a scale of 0–100. And the hearing state is calculated from the average of 500, 1000, 2000, 4000 and 8000 Hz. Low frequency tinnitus pitch means less than or equal to 1000 Hz, medium frequency tinnitus pitch means 1000 Hz to 4000 Hz, and high frequency tinnitus pitch indicates more than 4000 Hz. The details about the study groups are listed in Table I.

2) *EEG Data Collection*: A high-density EEG system with 128 channels (EGI, Eugene) and a NetAmps 200 amplifier are used to collect the resting-state EEG recordings from all participants. The sampling rate is 1000 Hz and impedances are kept below 50 k $\Omega$ . The CZ electrode is used as the reference electrode. During the EEG data collection process, participants are asked to sit on a chair, open their eyes and focus on a cross mark on the computer screen to keep awake. The whole process lasts for about 7 minutes.

3) *Data Preprocessing*: We apply the EEGLAB v13.0.0 toolbox in MATLAB R2013a to preprocess the raw EEG data. The raw data is first resampled at 250 Hz, band-pass filtered between 0.5 Hz and 80 Hz, and notch filtered at 50 Hz, and re-referenced to the 56th and 107th electrodes which locate in the bilateral mastoid. Next, evident artifacts are removed from the raw EEG data manually after visual inspection, and the independent component analysis (ICA) is utilized to remove muscle artifacts, eye movement, and heartbeats in the raw data. Then, the EEG data is segmented into 2 s slices and a basic finite impulse response (FIR) filter of EEGLAB toolbox is applied to transform the slices to 8 frequency bands which are significant for tinnitus diagnosis, including delta (2–3.5 Hz), theta (4–7.5 Hz), alpha1 (8–10 Hz), alpha2 (10–12 Hz), beta1 (13–18 Hz), beta2 (18.5–21 Hz), beta3 (21.5–30 Hz) and gamma (30.5–44 Hz) [32], [33], [34], [35], [36]. The process of extracting each frequency band of the EEG signals is shown in Fig. 2. Finally, for each frequency band, we calculate the mean and standard deviation of the voltage value of each EEG segment on each electrode, and then electrode-level zero-mean normalization is performed on each EEG segment, which normalizes the data while maintaining the dynamic pattern of the EEG signals.

### B. Model Training

The entire training process of MECRL consists of the representation learning process and the classification process. In the representation learning process, the EEG sampler first applies different sampling strategies (strategy A, B and C, respectively) to generate mini-batches (mini-batch A, B and C, respectively),

TABLE I  
THE INFORMATION SHEET ABOUT THE STUDY GROUPS

	Tinnitus patients ( $n=187$ )	Normal subjects ( $n=80$ )	$t/X^2, P$
<b>Gender</b> (male: female)	77 : 110	43 : 37	$X^2 = 3.58, P > 0.05$
<b>Age</b> (years)	$43.24 \pm 14.205$	$39.84 \pm 14.362$	$t = 1.78, P > 0.05$
<b>Hearing state:</b> Right ear (dB HL)	$27.32 \pm 16.23$	$< 20$	—
<b>Hearing state:</b> Left ear (dB HL)	$29.84 \pm 17.40$	$< 20$	—
<b>Tinnitus laterality</b> (left: right: bilateral)	54 : 39 : 94	—	—
<b>Tinnitus pitch</b> (pure tone: narrow noise)	177 : 10	—	—
<b>Tinnitus frequency</b> (low: medium: high frequencies)	25 : 16 : 146	—	—
<b>Tinnitus loudness</b> (VAS scores)	$4.50 \pm 3.14$	—	—
<b>Tinnitus severity</b> (THI scores)	$42.34 \pm 24.99$	—	—

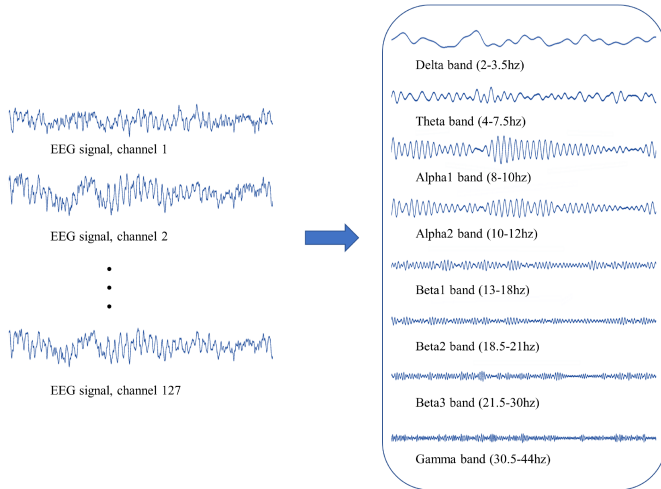


Fig. 2. The process of extracting 8 frequency bands from raw EEG signals.

and then the generated mini-batches will be fed to the multi-band EEG encoder, which is made up of 8 parallel deep neural networks composed of the spatial module and the temporal module. The EEG segments of different frequency bands first pass through the spatial module to integrate spatial information and streamline channels (electrodes), so that only the data of electrodes with better tinnitus diagnosis effect in each frequency band will flow into the next layer. And then the EEG data is further streamed into the temporal module to learn its dynamic patterns in time domain. After that, the outputs of the multi-band EEG encoder will be fed to the feature fusioner to generate a unified representation. Finally, different contrastive loss ( $L_A$ ,  $L_B$ ,  $L_C$ , respectively) will be calculated and the model parameters are continuously optimized via minimizing the loss by the optimizer. After the representation learning process, we attach an MLP classifier behind the model and train it using the supervised learning paradigm, which constitutes the classification process.

### C. Experiment Setting

In order to ensure scientific nature of the subject-independent tinnitus diagnosis experiments, the training set and test set are obtained by stratified sampling (randomly sample 90% of the subjects in the tinnitus group and control group as the training set, and the remaining subjects as the test set), which makes

the subjects in the test set unseen by the model. Under this experimental setting, we can better evaluate the effectiveness of different models in clinical tinnitus diagnosis.

For the architecture of the multi-band EEG encoder, the number of spatial convolution filters is set to 16 and the number of temporal convolution filters increases from 16 to 128 from the first layer to the deeper layer. The temporal convolution filter length is set as 32 and the kernel length of average pooling is set as 16 to extract relatively stable averaged features. The output dimension of the projector is set to 32. The temperature coefficients of the three auxiliary tasks are set to 0.1, 0.07, and 0.04 respectively. In the representation learning process, we train the model for 300 epochs with early stopping (maximal tolerance of 30 epochs without descending validation loss). An Adam optimizer with a cosine annealing learning rate scheduler and a three-time warm restart are applied to optimize the model. The initial learning rate of the optimizer is set to 0.0007, and the weight decay is set to 0.015 empirically.

For the linear classifier in the classification process, there are two hidden layers with 64 and 32 units respectively. Rectified linear units (ReLUs) are used between the two layers. We use cross-entropy loss and an Adam optimizer to optimize the parameters. The learning rate is set as 0.0005 empirically, and the weight decay is set to 0.025. The batch size is set as 64 empirically. The classifier is trained for 100 epochs.

### D. Evaluation Measures

In our experiments, five evaluation measures are adopted to evaluate the performance of different models, namely ACC, AUC, Precision, Recall and F1-score. ACC is the segment-wise accuracy of the model for tinnitus diagnosis. AUC stands for “Area under the ROC Curve” which measures the entire two-dimensional area underneath the entire ROC curve (receiver operating characteristic curve). ROC is a graph showing the performance of a classification model at all classification thresholds while AUC provides an aggregate measure of performance across all possible classification thresholds. Precision refers to the proportion of true tinnitus EEG segments among the tinnitus segments predicted by the model, while Recall reflects the proportion of EEG segments which are correctly predicted in all tinnitus EEG segments. F1-score combines the precision and recall of a model into a single metric by taking their harmonic mean. The larger ACC, AUC, Precision, Recall and F1-score are, the better the classification effect is.

TABLE II

THE MEAN AND STANDARD DEVIATION (OVER 10 RUNS) OF CLASSIFICATION ACCURACY OBTAINED BY DIFFERENT METHODS ON SUBJECT-INDEPENDENT TINNITUS DIAGNOSIS UNDER THE 9:1 SETTING. IN EACH COLUMN, THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Method	ACC	AUC	Precision	Recall	F1-score	ACC improvement
v-SVM [37]	0.5817( <b>0.025</b> )	0.6123( <b>0.018</b> )	0.5933( <b>0.020</b> )	0.5968( <b>0.027</b> )	0.6007( <b>0.026</b> )	57.023%
MLP [38]	0.6546(0.031)	0.6763(0.036)	0.6849(0.025)	0.6618(0.036)	0.6753(0.034)	39.536%
EEGNet [39]	0.7366(0.067)	0.7427(0.071)	0.7582(0.073)	0.7285(0.064)	0.7423(0.083)	24.002%
SiameseAE [14]	0.7972(0.042)	0.8018(0.048)	0.8241(0.052)	0.8068(0.051)	0.8193(0.048)	14.576%
SMeta-SAE [40]	0.8131(0.038)	0.8257(0.046)	0.8314(0.049)	0.8169(0.047)	0.8255(0.046)	12.336%
4D-CNN [41]	0.8429(0.084)	0.8474(0.088)	0.8563(0.079)	0.8622(0.116)	0.8672(0.098)	8.364%
<b>MECRL</b>	<b>0.9134</b> (0.031)	<b>0.9256</b> (0.034)	<b>0.9444</b> (0.053)	<b>0.9500</b> (0.044)	<b>0.9472</b> (0.049)	—

### E. Subject-Independent Tinnitus Diagnosis Experiment

To demonstrate the superiority of our MECRL framework, we select the traditional machine learning methods based on feature engineering and several competitive EEG-based deep learning methods as baselines and reproduce them on our EEG dataset.

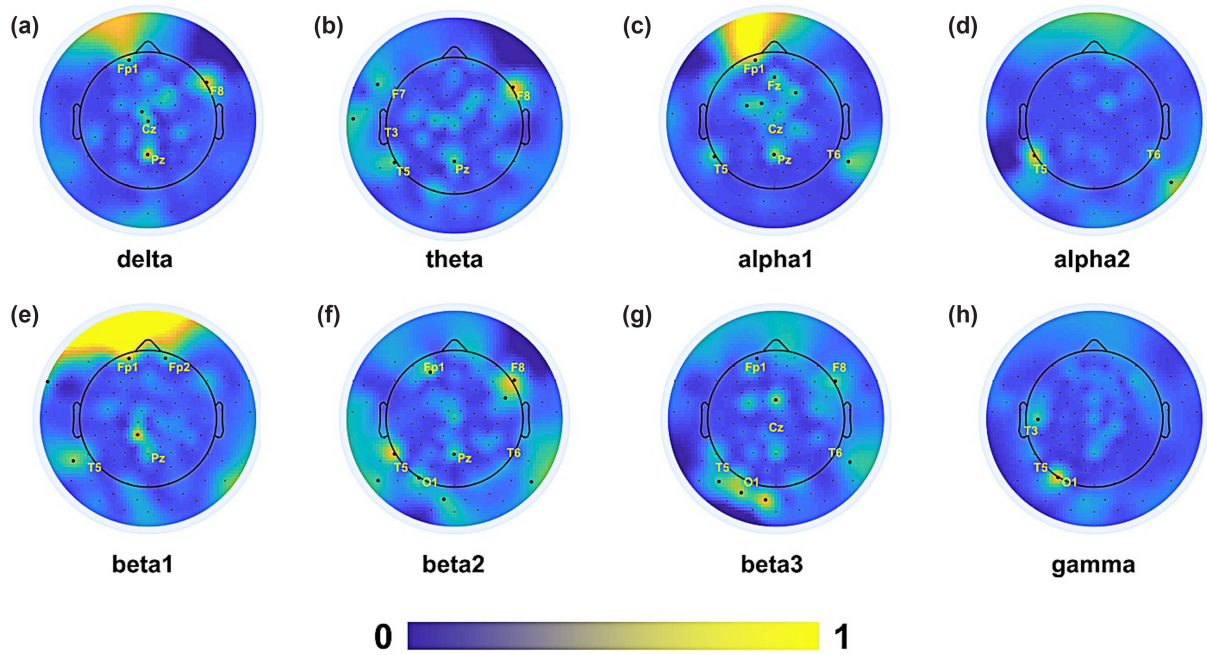
- 1) v-SVM [37] is a variant of the Support Vector Machine (SVM) algorithm, which can be interpreted as a maximal separation between subsets of the convex hulls of the data. The convex hulls are controlled by choice of the parameter  $v$ . Because the traditional machine learning methods can hardly handle complex EEG signals, Power Spectrum Density (PSD) of each electrode is calculated as the input, which is a common way of hard-encoding the EEG data in disease diagnosis.
- 2) MLP [38] stands for Multilayer Perceptron, which is a basic deep learning model and has strong nonlinear fitting ability. It takes the PSD of electrodes of EEG data as input and outputs the prevalence of tinnitus.
- 3) EEGNet [39] does not hard-encode the EEG data, but uses Convolutional Neural Networks (CNN) to automatically extract features and complete the classification tasks. However, since there is no special mechanism for individual variability, this method suffers from the overfitting in practical applications.
- 4) SiameseAE [14] is an auto-encoder based Siamese network which takes ABRs (auditory brainstem responses) as input for tinnitus diagnosis. ABRs are single-channel evoked potentials recorded with EEG sensor. To adapt the SiameseAE method to the EEG dataset, we use the potentials of the corresponding electrode (electrode no.13) located on the midfrontal line as its input. The method designs a variety of different loss functions to align individual differences and achieves good results. However, the auto-encoder structure is too simple and does not take the spatial and frequency information of the EEG data into account, which limits the further improvement of the diagnosis performance.
- 5) SMeta-SAE [40] is an ABR-based cross-dataset tinnitus diagnosis method using SiameseAE as the backbone model. The introduction of meta-learning alleviates the problem of missing large-scale datasets. However, similar to SiameseAE, it does not consider the spatial and frequency information.
- 6) 4D-CNN [41] is a deep learning method that fully considers the multi-dimensional information of the EEG data.

Through frequency division and topographic map extraction, the two-dimensional EEG signals are converted into four-dimensional signals as the input of 4D convolutional neural network to complete the classification tasks. Similar to EEGNet, the overfitting problem limits the performance of the method in the cross-subject tinnitus diagnosis task.

The results of the subject-independent tinnitus diagnosis experiment are listed in Table II, in which the mean and the standard deviation are reported over 10 runs with different stratified sampling results. From Table II, we can find that the proposed MECRL method achieves outstanding accuracy on tinnitus diagnosis as high as 91.34%. It outperforms the second and third best methods (i.e. 4D-CNN and SMeta-SAE) by achieving improvements of 8.364% and 12.336% respectively, which reflects the excellent performance of the MECRL method. Moreover, we find that the methods taking handcrafted features as input (i.e. v-SVM and MLP) generally perform worse than the methods extracting deep features from the raw EEG data (i.e. EEGNet, SiameseAE, SMeta-SAE, 4D-CNN and MECRL). This is because the transformation from raw EEG data to handcrafted features will lose a lot of useful information for tinnitus diagnosis. On the contrary, the deep neural network can dynamically extract effective information from the EEG signals according to the needs of downstream tasks. However, since EEGNet and 4D-CNN are trained under a simple supervised learning paradigm, it is easy for the deep neural network model to take random noise in complex signals as discriminative features in the case of insufficient training samples, making the model perform poorly when faced with data from unseen subjects. Unlike them, MECRL deconstructs the complex EEG data through a series of self-supervised learning tasks. These tasks are progressive and semantically complementary, so that the MECRL model can effectively map EEG segments into semantically rich and discriminative representation. That's why MECRL outperforms other methods.

In terms of stability, because the high complexity of input data will bring greater uncertainty, the standard deviation of the accuracy of v-SVM and MLP is smaller, which means that they are more consistent in multiple experiments. Different from other EEG-based deep learning methods, the representation learning process of MECRL is equivalent to self-supervised pre-training of the model parameters. Therefore, compared with other methods that start training from completely random initialization parameters, the MECRL method has stronger stability, i.e. smaller standard deviation.



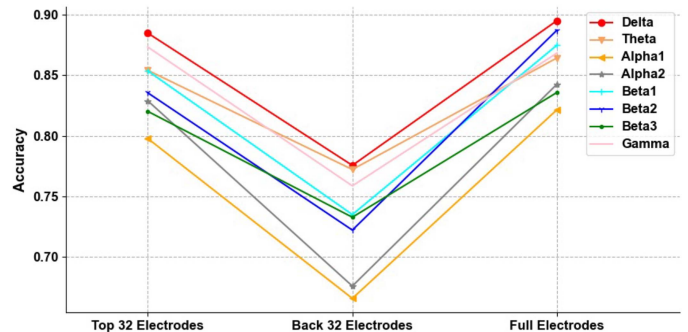


**Fig. 3.** The topographic maps generated by visualizing the weight distribution of scalp electrodes in 8 frequency bands. The black dots in bold indicate electrodes with a weight greater than 0.55 and the top 10% weights in the corresponding frequency band, and the core electrodes among them are marked with the international standard name (10%-10% system). (a) In the delta band, the electrodes of high weight are mainly distributed in the left prefrontal pole (Fp1), right frontal (F8), central parietal (Pz), central cranial (Cz) and nearby electrode groups; (b) In the theta band, the electrodes of high weight are mainly distributed in the central parietal (Pz), bilateral frontal (F7 and F8), left temporal (T3 and T5) and nearby electrode groups; (c) In the alpha1 band, the electrodes of high weight are mainly distributed in the left prefrontal pole (Fp1), central frontal (Fz), central cranial (Cz), bilateral temporal (T5 and T6) and nearby electrode groups; (d) In the alpha2 band, the electrodes of high weight are mainly distributed in the bilateral temporal (T5 and T6); (e) In the beta1 band, the electrodes of high weight are mainly distributed in the bilateral prefrontal pole (Fp1 and Fp2), central parietal (Pz) and left temporal (T5); (f) In the beta2 band, the electrodes of high weight are mainly distributed in the left prefrontal pole (Fp1), right frontal (F8), central parietal (Pz), bilateral temporal (T5 and T6) and left occipital (O1); (g) In the beta3 band, the electrodes of high weight are mainly distributed in the left prefrontal pole (Fp1), right frontal (F8), central cranial (Cz), bilateral temporal (T5 and T6) and left occipital (O1); (h) In the gamma band, the electrodes of high weight are mainly distributed in the left temporal (T3 and T5) and left occipital (O1).

### F. Parameter Analysis

By applying the regularized spatial module, the EEG data from brain electrodes in different spatial locations is fed into the backbone encoder with different weights. With the advantage of the training process, the weights of electrodes in brain regions that are more important for tinnitus diagnosis would be continuously amplified, while the weights of electrodes in incoherent brain regions would tend to be zero. To explore the role of the EEG data in different frequency bands for tinnitus diagnosis and trace abnormal brain regions, we extract the weight coefficients of the electrodes in each brain region, which are visualized in Fig. 3. In the figure, we can find that the electrodes with high weights in each frequency band are sparse and concentrated, which implies that many electrodes in EEG are likely to be unhelpful for specific downstream tasks. The results also illustrate that high-weight electrodes in each frequency band are spatially clustered and the brain areas in which these electrodes are located may be related to the onset of chronic tinnitus.

To demonstrate the validity of the traceability of abnormal brain regions associated with tinnitus, we retrain the single-band version of the model with the spatial module removed using the EEG data of the 32 electrodes with the highest weight and the 32 electrodes with the lowest weight in each frequency band



**Fig. 4.** Parameter analysis: Accuracy of model retained with data from different electrodes.

respectively. The results are shown in Fig. 4. From the figure, we can see that the performance of the model trained with high-weight electrodes in each frequency band is much better than that of the model trained with low-weight electrodes. What's more, the performance of the model using the 32 electrodes with the highest weight is very close to the model using all electrodes, and even slightly better than the all-electrode model in the gamma frequency band. This suggests that the brain regions where



**TABLE III**  
ABLATION STUDY: COMPARISON BETWEEN MECRL AND ITS FOUR VARIANTS

Method	ACC	AUC	Precision	Recall	F1-score
MECRL-Task A	0.7778	0.7924	0.8333	0.8004	0.8165
MECRL-Task B	0.8641	0.8875	0.8713	0.8366	0.8536
MECRL-Task C	0.8826	0.8719	0.8315	0.8853	0.8576
MECRL-RPT	0.7807	0.7667	0.7921	0.7702	0.7810
<b>MECRL</b>	<b>0.9134</b>	<b>0.9256</b>	<b>0.9444</b>	<b>0.9500</b>	<b>0.9472</b>

these high-weight electrodes are located are associated with the onset and diagnosis of tinnitus. In addition, we can find that the model performance of the single-band version is much worse than that of the multi-band version. It shows the essentiality of frequency domain information for tinnitus diagnosis, which is one of the reasons for the poor performance of EEGNet, SiameseAE and SMeta-SAE. More medical discussions can be found in Section IV.

### G. Ablation Study

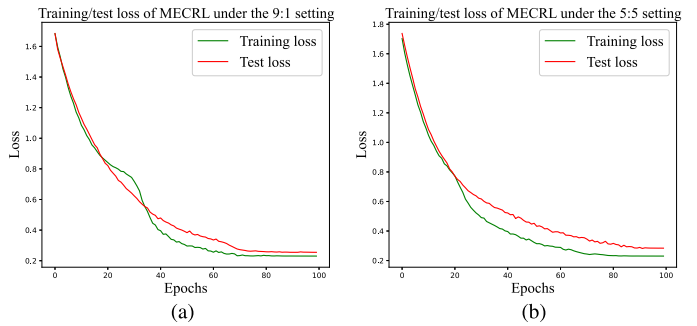
In this section, ablation study is conducted to demonstrate the role of different auxiliary tasks as well as the representation learning process in our MECRL framework. Specifically, we compare the proposed MECRL framework with four variant models:

- *MECRL-Task A*: The multi-band EEG encoder is trained only with task B and C.
- *MECRL-Task B*: The multi-band EEG encoder is trained only with task A and C.
- *MECRL-Task C*: The multi-band EEG encoder is trained only with task A and B.
- *MECRL-RPT*: The representation learning process is removed and the model is trained under the traditional supervised learning paradigm.

The results in Table III show that the designed auxiliary tasks as well as the representation learning process can effectively help the model to better learn cross-individual feature representations from the EEG data, which is helpful for the tinnitus diagnosis.

### H. Overfitting Analysis

In order to further demonstrate the robustness of the proposed MECRL framework in the case of scarcity of training data, we modify the original experimental setting on generating the training set and the test set. Specifically, we randomly sample 50% of the subjects in the tinnitus group and control group as the training set, and the remaining subjects as the test set. That is, different from the 9:1 training-test ratio, we have 5:5 training-test ratio. Under the new experimental setting (i.e. the 5:5 training-test setting), the training data is almost cut in half, and the larger test set allows us to evaluate the performance of the models more accurately. Moreover, we display the loss curve of MECRL in the classification process under the two experimental settings in Fig. 5 to verify whether it is overfitting. And the results of the subject-independent tinnitus diagnosis experiment under 5:5 setting is shown in Table IV. From the table, we can find that under the experimental setting of 5:5, MECRL still



**Fig. 5.** The loss curve of MECRL under (a) The 9:1 setting and (b) The 5:5 setting.

performs better than the baselines. Meanwhile, we can observe that compared with the experimental setting of 9:1, the accuracy of EEGNet, SiameseAE, SMeta-SAE, 4D-CNN and MECRL decreased by 3.41%, 3.06%, 2.84%, 4.82% and 1.34% respectively under the experimental setting of 5:5. This is because the reduction of training data makes it difficult to optimize the parameters for data-driven deep neural networks. However, the drop in MECRL is the lowest among all deep learning methods, reflecting its robustness in the face of the dilemma of scarcity of training data. This is due to the self-supervised pre-training process (i.e. the representation learning process) that makes the generated feature space have a good internal structure and rich semantic information.

In addition, from Fig. 5, we can find that the training loss and test loss can achieve convergence regardless of the experimental setting, which shows that the overfitting phenomenon in MECRL is not as serious as other baselines. This can also explain its excellent performance in the EEG-based tinnitus diagnosis task.

## IV. MEDICAL DISCUSSION

Resting state EEG markers are promising for large-scale, low-cost, noninvasive screening of tinnitus subjects. Robust automatic classification of such neural signatures that are predictive at the individual patient level would both advance the neurobiological understanding of tinnitus and provide important clinical implications. Therefore, our work aims to differentiate tinnitus patients from healthy people, and further identify characteristic tinnitus features extracted from the EEG data using a robust, data-efficient multi-task learning framework called Multi-band EEG Contrastive Representation Learning (MECRL). The deep model finally achieves higher accuracy (ACC 0.9134 and AUC 0.9256) for tinnitus diagnosis than other models.

To further explore potential brain regions that contribute to the successful tinnitus classification, we extract and visualize the spatial weights in the spatial module which are shown in Fig. 3. From Fig. 3 we can find that electrodes helpful for the classification and diagnosis of tinnitus are mainly distributed in the frontal (Fp1 and Fz), parietal (Pz), central cranial (Cz), temporal (T5 and T3) and occipital region (O1) in each frequency band, as well as nearby auxiliary electrode. In summary, our

TABLE IV

THE MEAN AND STANDARD DEVIATION (OVER 10 RUNS) OF CLASSIFICATION ACCURACY OBTAINED BY DIFFERENT METHODS ON SUBJECT-INDEPENDENT TINNITUS DIAGNOSIS UNDER THE 5:5 SETTING. IN EACH COLUMN, THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Method	ACC	AUC	Precision	Recall	F1-score	ACC improvement
v-SVM [37]	0.5648( <b>0.029</b> )	0.5974( <b>0.026</b> )	0.5826( <b>0.030</b> )	0.5817( <b>0.029</b> )	0.5844( <b>0.032</b> )	59.561%
MLP [38]	0.6333(0.036)	0.6287(0.040)	0.6412(0.032)	0.6477(0.036)	0.6405(0.033)	42.302%
EEGNet [39]	0.7123(0.097)	0.7165(0.080)	0.7199(0.075)	0.7018(0.086)	0.7144(0.089)	28.321%
SiameseAE [14]	0.7728(0.062)	0.7894(0.074)	0.7811(0.067)	0.7934(0.068)	0.7901(0.071)	16.615%
SMeta-SAE [40]	0.7900(0.052)	0.7831(0.063)	0.7866(0.067)	0.8015(0.070)	0.7994(0.072)	14.076%
4D-CNN [41]	0.8038(0.121)	0.8126(0.105)	0.8273(0.094)	0.8068(0.082)	0.8143(0.091)	10.740%
<b>MECRL</b>	<b>0.9012</b> (0.058)	<b>0.9116</b> (0.060)	<b>0.9145</b> (0.041)	<b>0.9088</b> (0.064)	<b>0.9143</b> (0.057)	—

model mostly indicates the importance of these electrodes in the diagnosis and classification of tinnitus.

Essentially, the EEG biomarkers highlighted here assist the interpretation of the mechanism of central neurophysiological changes in tinnitus. These high-classification weighted electrodes are distributed in the frontal, parietal, temporal and occipital regions, which are the core regions of the default mode network (DMN). The DMN is generally considered to consist of core regions including the auditory cortex, dorsolateral prefrontal cortex, parahippocampal gyrus and precuneus/posterior cingulate gyrus [42], [43], which are normally active in the absence of external stimuli. The DMN's function includes monitoring the external environment, maintaining self-awareness, generating spontaneous thoughts, and processing episodic memory, cognition, and emotion [44]. Consistent with the previous studies [11], [12], [45], our results suggest that the DMN abnormality is a significant central neural characteristic of tinnitus. Dysregulation of the default mode network, which controls self-representation processing, is related to the pathogenesis of tinnitus pathology [9]. In other words, the sound (tinnitus) becomes an integral part of patients' self [46]. This may be why tinnitus has become persistent and difficult to treat.

Furthermore, the pathological DMN is associated with changes in brain oscillatory activity. The special activity of DMN is largely organized synchronously mediated by alpha rhythms [47]. Early EEG studies have suggested that the attenuation of alpha rhythm and the increase of theta and high frequency band (beta and gamma) are the electrophysiological manifestations of tinnitus [48], [49], [50]. However, the significance of DMN dysfunction in tinnitus is still a mystery [51]. It is not clear that the neural oscillation of DMN measured with EEG is related to spatial distribution of the electrode nodes over the cortex [52], [53]. Interestingly, by using machine learning to analyze multiple EEG bands, our results show that the number of electrodes with classification weight significance is more in the alpha1, beta1, beta2, and beta3 bands, but less in the alpha2 band. This suggests that the slowing of the alpha rhythm of tinnitus EEG may be a phenomenon, but the more underlying pathological significance is the alteration of DMN under the alpha rhythm. This is a significant neuro-biomarker for EEG oscillations of tinnitus, which facilitates our understanding of the relationship between electrophysiology and pathophysiology changes of tinnitus.

Limitations must be acknowledged in this study. This is a cross-sectional study with a limited sample size. We mainly investigate the differences in cortical activity between chronic tinnitus and healthy subjects. The high heterogeneity factors of

tinnitus such as laterality, hearing condition, tinnitus severity on brain network are not further explored in the present study. In the future study, it would be very interesting to explore the influence of various factors on the central brain networks in tinnitus patients by categorizing them on the basis of the tinnitus characteristics.

## V. CONCLUSION

In this study, we propose a robust, data-efficient multi-task framework named MECRL for cross-subject tinnitus diagnosis. It can effectively solve the overfitting problem suffered by the deep neural networks caused by the lack of large-scale dataset and high individual variability. In the framework, we design several layer-by-layer progressive learning tasks to deconstruct the semantics of EEG data, which help the deep residual network-based encoder integrate multi-dimensional information and align the individual variability. Moreover, the self-supervised learning tasks do not depend on the annotated data and complement each other. This allows us to efficiently train deep models without large-scale EEG datasets and manually extracting discriminative subject-independent representations that are used to distinguish the tinnitus subjects from the healthy controls precisely. Experiments are conducted on a tinnitus dataset collected from the department of Otolaryngology, Sun Yat-sen Memorial hospital, Sun Yat-sen University. And the results have confirmed the effectiveness of the proposed MECRL framework over the state-of-the-art baselines.

## REFERENCES

- [1] M. Teplan et al., "Fundamentals of EEG measurement," *Meas. Sci. Rev.*, vol. 2, no. 2, pp. 1–11, 2002.
- [2] J. C. Henry, "Electroencephalography: Basic principles, clinical applications, and related fields," *Neurology*, vol. 67, no. 11, pp. 2092–2092, 2006.
- [3] B.-K. Min, M. S. Hämäläinen, and D. Pantazis, "New cognitive neurotechnology facilitates studies of cortical-subcortical interactions," *Trends Biotechnol.*, vol. 38, no. 9, pp. 952–962, 2020.
- [4] B. Lee et al., "Synergy through integration of wearable EEG and virtual reality for mild cognitive impairment and mild dementia screening," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 7, pp. 2909–2919, Jul. 2022.
- [5] S. Gaubert et al., "EEG evidence of compensatory mechanisms in preclinical Alzheimer's disease," *Brain*, vol. 142, no. 7, pp. 2096–2112, 2019.
- [6] A. B. Elgoyhen, B. Langguth, D. D. Ridder, and S. Vanneste, "Tinnitus: Perspectives from human neuroimaging," *Nature Rev. Neurosci.*, vol. 16, no. 10, pp. 632–642, 2015.
- [7] J.-J. Song, S. Vanneste, W. Schlee, P. V. D. Heyning, and D. D. Ridder, "Onset-related differences in neural substrates of tinnitus-related distress: The anterior cingulate cortex in late-onset tinnitus, and the frontal cortex in early-onset tinnitus," *Brain Struct. Function*, vol. 220, no. 1, pp. 571–584, 2015.
- [8] S. Rosemann and J. P. Rauschecker, "Neuroanatomical alterations in middle frontal gyrus and the precuneus related to tinnitus and tinnitus distress," *Hear. Res.*, vol. 424, 2022, Art. no. 108595.

- [9] S. A. Schmidt, J. Carpenter-Thompson, and F. T. Husain, "Connectivity of precuneus to the default mode and dorsal attention networks: A possible invariant marker of long-term tinnitus," *NeuroImage: Clin.*, vol. 16, pp. 196–204, 2017.
- [10] R. Araneda et al., "A key role of the prefrontal cortex in the maintenance of chronic tinnitus: An fMRI study using a stroop task," *NeuroImage: Clin.*, vol. 17, pp. 325–334, 2018.
- [11] J.-J. Song et al., "The balance between Bayesian inference and default mode determines the generation of tinnitus from decreased auditory input: A volume entropy-based study," *Hum. Brain Mapping*, vol. 42, no. 12, pp. 4059–4073, 2021.
- [12] Y.-C. Chen et al., "Tinnitus distress is associated with enhanced resting-state functional connectivity within the default mode network," *Neuropsychiatric Dis. Treat.*, vol. 14, pp. 1919–1927, 2018.
- [13] S.-Y. Lee et al., "Is the posterior cingulate cortex an on-off switch for tinnitus?: A comparison between hearing loss subjects with and without tinnitus," *Hear. Res.*, vol. 411, 2021, Art. no. 108356.
- [14] 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.
- [15] Y. Li, Z. Liu, L. Yao, J. J. M. Monaghan, and D. McAlpine, "Disentangled and side-aware unsupervised domain adaptation for cross-dataset subjective tinnitus diagnosis," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 1, pp. 538–549, 2023.
- [16] M.-P. Hosseini, A. Hosseini, and K. Ahi, "A review on machine learning for EEG signal processing in bioengineering," *IEEE Rev. Biomed. Eng.*, vol. 14, pp. 204–218, 2020.
- [17] S. Vanneste, J.-J. Song, and D. D. Ridder, "Thalamocortical dysrhythmia detected by machine learning," *Nature Commun.*, vol. 9, no. 1, pp. 1–13, 2018.
- [18] D. A. Engemann et al., "Robust EEG-based cross-site and cross-protocol classification of states of consciousness," *Brain*, vol. 141, no. 11, pp. 3179–3192, 2018.
- [19] D. R. Edla, K. Mangalorekar, G. Dhavalikar, and S. Dodia, "Classification of EEG data for human mental state analysis using random forest classifier," *Procedia Comput. Sci.*, vol. 132, pp. 1523–1532, 2018.
- [20] S. Parui, A. K. R. Bajija, D. Samanta, and N. Chakravorty, "Emotion recognition from EEG signal using XGBoost algorithm," in *Proc. IEEE 16th India Council Int. Conf.*, 2019, pp. 1–4.
- [21] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, 2019, Art. no. 031001.
- [22] M. Lee et al., "Quantifying arousal and awareness in altered states of consciousness using interpretable deep learning," *Nature Commun.*, vol. 13, no. 1, pp. 1–14, 2022.
- [23] X. Shen, X. Liu, X. Hu, D. Zhang, and S. Song, "Contrastive learning of subject-invariant EEG representations for cross-subject emotion recognition," *IEEE Trans. Affect. Comput.*, early access, Apr. 04, 2022, doi: [10.1109/TAFFC.2022.3164516](https://doi.org/10.1109/TAFFC.2022.3164516).
- [24] A. Kondacs and M. Szabó, "Long-term intra-individual variability of the background EEG in normals," *Clin. Neurophysiol.*, vol. 110, no. 10, pp. 1708–1716, 1999.
- [25] L. Xu, M. Xu, Y. Ke, X. An, S. Liu, and D. Ming, "Cross-dataset variability problem in EEG decoding with deep learning," *Front. Hum. Neurosci.*, vol. 14, 2020, Art. no. 103.
- [26] P. M. Rossini, F. Miraglia, and F. Vecchio, "Early dementia diagnosis, MCI-to-dementia risk prediction, and the role of machine learning methods for feature extraction from integrated biomarkers, in particular for EEG signal analysis," *Alzheimer's Dement.*, vol. 18, pp. 2699–2706, 2022.
- [27] P. Khosla et al., "Supervised contrastive learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 18661–18673.
- [28] R. Krishnan, P. Rajpurkar, and E. J. Topol, "Self-supervised learning in medicine and healthcare," *Nature Biomed. Eng.*, vol. 6, pp. 1346–1352, 2022.
- [29] F. Wang and H. Liu, "Understanding the behaviour of contrastive loss," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 2495–2504.
- [30] Y. Awni et al., "Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network," *Nature Med.*, vol. 25, no. 1, pp. 65–69, 2019.
- [31] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 1597–1607.
- [32] A. Mohan, N. Moreno, J.-J. Song, D. D. Ridder, and S. Vanneste, "Evidence for behaviorally segregated, spatiotemporally overlapping subnetworks in phantom sound perception," *Brain Connectivity*, vol. 7, no. 3, pp. 197–210, 2017.
- [33] S. Vanneste, K. Joos, J. Ost, and D. D. Ridder, "Influencing connectivity and cross-frequency coupling by real-time source localized neurofeedback of the posterior cingulate cortex reduces tinnitus related distress," *Neurobiol. Stress*, vol. 8, pp. 211–224, 2018.
- [34] A. Mohan, D. D. Ridder, and S. Vanneste, "Robustness and dynamicity of functional networks in phantom sound," *Neuroimage*, vol. 146, pp. 171–187, 2017.
- [35] S. Vanneste and D. D. Ridder, "Deafferentation-based pathophysiological differences in phantom sound: Tinnitus with and without hearing loss," *Neuroimage*, vol. 129, pp. 80–94, 2016.
- [36] J.-J. Song, A. K. Punte, D. D. Ridder, S. Vanneste, and P. V. de Heyning, "Neural substrates predicting improvement of tinnitus after cochlear implantation in patients with single-sided deafness," *Hear. Res.*, vol. 299, pp. 1–9, 2013.
- [37] D. J. Crisp and C. J. C. Burges, "A geometric interpretation of v-SVM classifiers," in *Proc. Adv. Neural Inf. Process. Syst.*, S. A. Solla, T. K. Leen, and K. Müller, Eds., 1999, pp. 244–250.
- [38] R. Atangana, D. Tchiotso, G. Kenne, and L. Chanel, "EEG signal classification using LDA and MLP classifier," *Health Informat. Int. J.*, vol. 9, no. 1, pp. 14–32, 2020.
- [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, 2018, Art. no. 056013.
- [40] Z. Liu, Y. Li, L. Yao, M. Lucas, J. J. Monaghan, and Y. Zhang, "Side-aware meta-learning for cross-dataset listener diagnosis with subjective tinnitus," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2352–2361, 2022.
- [41] C.-T. Lin, C.-H. Chuang, Y.-C. Hung, C.-N. Fang, D. Wu, and Y.-K. Wang, "A driving performance forecasting system based on brain dynamic state analysis using 4-D convolutional neural networks," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 4959–4967, Oct. 2021.
- [42] M. D. Greicius, B. Krasnow, A. L. Reiss, and V. Menon, "Functional connectivity in the resting brain: A network analysis of the default mode hypothesis," *Proc. Nat. Acad. Sci.*, vol. 100, no. 1, pp. 253–258, 2003.
- [43] G. L. Shulman et al., "Common blood flow changes across visual tasks: I increases in subcortical structures and cerebellum but not in nonvisual cortex," *J. Cogn. Neurosci.*, vol. 9, no. 5, pp. 624–647, 1997.
- [44] R. L. Buckner, J. R. Andrews-Hanna, and D. L. Schacter, "The brain's default network: Anatomy, function, and relevance to disease," *Ann. New York Acad. Sci.*, vol. 1124, no. 1, pp. 1–38, 2008.
- [45] S. A. Schmidt, K. Akrofi, J. R. Carpenter-Thompson, and F. T. Husain, "Default mode, dorsal attention and auditory resting state networks exhibit differential functional connectivity in tinnitus and hearing loss," *PLoS One*, vol. 8, no. 10, 2013, Art. no. e76488.
- [46] D. D. Ridder, S. Vanneste, J.-J. Song, and D. Adhia, "Tinnitus and the triple network model: A perspective," *Clin. Exp. Otorhinolaryngol.*, vol. 15, no. 3, pp. 205–212, 2022.
- [47] K. Jann, T. Dierks, C. Boesch, M. Kottlow, W. Strik, and T. Koenig, "Bold correlates of EEG alpha phase-locking and the fMRI default mode network," *Neuroimage*, vol. 45, no. 3, pp. 903–916, 2009.
- [48] I. Adamchic, T. Toth, C. Hauptmann, and P. A. Tass, "Reversing pathologically increased EEG power by acoustic coordinated reset neuromodulation," *Hum. Brain Mapping*, vol. 35, no. 5, pp. 2099–2118, 2014.
- [49] M. Meyer, P. Neff, A. Grest, C. Hemsley, S. Weidt, and T. Kleinjung, "EEG oscillatory power dissociates between distress-and depression-related psychopathology in subjective tinnitus," *Brain Res.*, vol. 1663, pp. 194–204, 2017.
- [50] M. Moazami-Goudarzi, L. Michels, N. Weisz, and D. Jeanmonod, "Temporo-insular enhancement of EEG low and high frequencies in patients with chronic tinnitus. QEEG study of chronic tinnitus patients," *BMC Neurosci.*, vol. 11, no. 1, pp. 1–12, 2010.
- [51] F. T. Husain and S. A. Schmidt, "Using resting state functional connectivity to unravel networks of tinnitus," *Hear. Res.*, vol. 307, pp. 153–162, 2014.
- [52] J. Samogin et al., "Frequency-dependent functional connectivity in resting state networks," *Hum. Brain Mapping*, vol. 41, no. 18, pp. 5187–5198, 2020.
- [53] T. Ossandón et al., "Transient suppression of broadband gamma power in the default-mode network is correlated with task complexity and subject performance," *J. Neurosci.*, vol. 31, no. 41, pp. 14521–14530, 2011.