



# Seizure Prediction With HIVE-CODAs: The Hierarchical Vote Collective of Domain Adaptation Methods

Peizhen Peng\*

Key Laboratory of Measurement and Control of CSE, Ministry of Education, School of Automation, Southeast University, Nanjing, China

Epileptic seizure prediction is one of the most used therapeutic adjuvant strategies for drug-resistant epilepsy. Conventional methods are usually trained and tested on the same patient due to the interindividual variability. However, the challenging problem of the domain shift between different subjects remains unsolved, resulting in low prevalence of clinical application. In this study, a generic model based on the domain adaptation (DA) technique is proposed to alleviate such problems. Ensemble learning is employed by developing a hierarchical vote collective of seven DA modules over multi-modality data, such that the predictive performance is improved by training multiple models. Moreover, to increase the feasibility of its implementation, this study mimics the data distribution of clinical sampling and tests the model under this simulated realistic condition. Based on the performance of seven subnetworks, the applicability of each DA algorithm for seizure prediction is evaluated, which is the first study that provides the assessment. Experimental results on both intracranial and scalp EEG databases demonstrate that this method can reduce the domain gap effectively compared with previous studies.

**Keywords:** seizure prediction, domain adaptation, ensemble learning, EEG, time series classification

## OPEN ACCESS

### Edited by:

Kai-Da Xu,  
Xi'an Jiaotong University, China

### Reviewed by:

Bin Chen,  
Yangzhou University, China  
Liangyu Ma,  
North China Electric Power University,  
China

### \*Correspondence:

Peizhen Peng  
zpeng0714@gmail.com

### Specialty section:

This article was submitted to  
Optics and Photonics,  
a section of the journal  
Frontiers in Physics

**Received:** 09 November 2021

**Accepted:** 01 December 2021

**Published:** 03 January 2022

### Citation:

Peng P (2022) Seizure Prediction With  
HIVE-CODAs: The Hierarchical Vote  
Collective of Domain  
Adaptation Methods.  
Front. Phys. 9:811681.  
doi: 10.3389/fphy.2021.811681

## 1 INTRODUCTION

### 1.1 Epilepsy Background

Epilepsy is a cerebral anomaly with the transient occurrence of unexpected seizures caused by excessive or hypersynchronous neuronal activities [1]. It is the second most clinically significant neurological disorder, which affects approximately 1.0% of the world's population [2]. The reliable seizure prediction device, which refers to anticipating an upcoming seizure based on continuous electroencephalogram (EEG) signals, is an emerging and important demand for drug-resistant individuals accounting for about 30% of the epileptic [3, 4]. The early warning device could significantly prevent the injury of epileptic coma, or even death.

EEG is a commonly used type of physiological signal that measures the epileptic brain activity, which records rhythmic information induced by coordinated neuronal firing with characteristic periodicity. The first-in-man forecast study was reported in 2013 [5], which offered the convincing proof of the predictability of seizure. Since then, many EEG-based algorithms adopting the data-driven technique have been presented.

## 1.2 Related Work

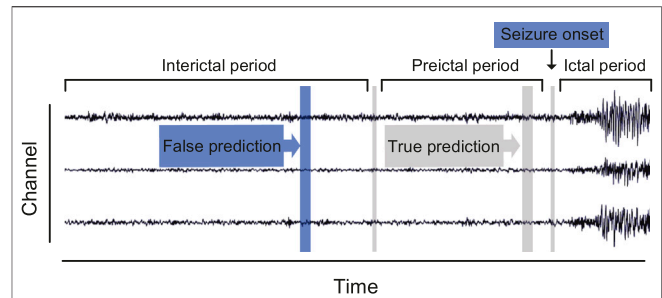
Current research studies regarding seizure prediction can be mainly categorized into two streams. The first stream typically follows a binary classification scheme, which assumes that a difference exists between the interictal and preictal stages. The ictal and postictal sequences are discarded during the data processing due to the futility of their contribution to forecast. The second stream is to detect the fluctuation of a specific index during the preictal period, such as the spike rate [6–8], zero-crossing intervals [9], and phase/amplitude locking value [10]. If the observed indicator exceeds the previously set threshold, an early warning would be declared. Owing to the multiplicity of multichannel EEG recordings, the first stream is more widely recognized than the second one. This study also adopts the strategy that distinguishes preictal states from interictal states, which is depicted in **Figure 1**.

Approaches using the binary classification scheme commonly adopt machine learning techniques like support vector machines [11–13], random forests [14], and *k*-nearest neighbor [15]. For the past few years, many deep learning frameworks, including convolutional neural network (CNN) [16–19], 3D CNN [20], long short-term memory (LSTM) network [21–23], and cascades of DNN [24], have been exploited to analyze continuously acquired epileptic EEG signals. However, there are still many promising algorithms to be developed and applied further. Ensemble learning is considered the state-of-the-art solution for many challenging problems. For instance, several representative approaches, including HIVE-COTE [25], boosting, bagging, and stacking, have achieved high performance for time series classification. Such methods are appealing because it has stronger generalization ability than a single model by training multiple subnetworks and combining their predictions. For this reason, we attempt to probe into its effectiveness for seizure prediction.

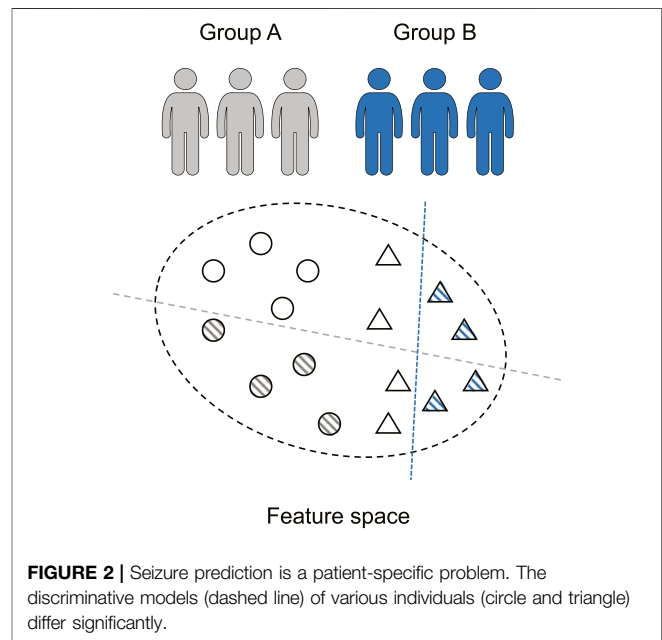
Most recently, various machine learning–based studies have achieved high performance. However, these methods are not yet in widespread use. Most of these research studies only provide patient-specific results, namely, both training and testing sets are collected from one subject. The reason for adopting this strategy is that large interindividual variability is ubiquitous among patients with epilepsy [26–28]. Therefore, an ensemble that contains a number of domain adaptation modules is developed in this study to reduce the impact of epileptic individual variability.

## 1.3 Significance

Although conventional studies achieve encouraging successes in seizure prediction task, their translation to the application remains challenging, in part due to their limited domain adaptability across different subjects. EEG patterns vary significantly from patient to patient as shown in **Figure 2**, and the issue regarding the model generalization ability remains unsolved. In previous studies, the training and testing sets are from the same patients, which can obtain a very high sensitivity (< 85% on average). Although such trials are important for personalized medicine, they are inconsistent with the clinical scenario in most cases. In other words, conventional models may



**FIGURE 1** | Definition of three brain states in continuous epileptic EEG recordings.



**FIGURE 2** | Seizure prediction is a patient-specific problem. The discriminative models (dashed line) of various individuals (circle and triangle) differ significantly.

perform well in one patient but be less effective in another, since the domain gap between different subjects is partly ignored. In practice, the training set is mainly composed of the previous patient data, and only a small amount of user samples can be used for training. The training set consisted of various subjects is the source domain. The “unseen” user is the target domain. In the existing literature, few research studies explore the domain shift issue. Therefore, a general seizure prediction model that is similar to clinical situation remains to explore and perfect further.

To alleviate the low popularization of clinical application and circumvent the impact of interindividual variability, domain adaptation (DA) is introduced for seizure prediction. However, few studies aim at using these techniques in epileptic EEG. In the fields such as image recognition and emotion-related EEG, successful applications of domain adaptation approaches have been reported [29, 30]. There are three main streams of DA algorithms. The first stream exploits adversarial learning to extract invariant information among source and target domains. The second stream extends sample size with data augmentation to access the target domain pattern in advance.

The third stream establishes general features based on specific prior knowledge.

Inspired by the success in other areas, we hope to extend DA to the field of seizure prediction. Since many DA techniques [31–33] have been provided, an ensemble learning-based model, the hierarchical vote collective of DA subnetworks (HIVE-CODAs), is proposed in this study. HIVE-CODAs combine the advantages of various DA methods. Besides, it can evaluate the applicability of each DA algorithm. In general, the main contributions of this study are summarized as follows:

- A generic model, HIVE-CODA, is proposed to tackle the DA problem for seizure prediction. It is the first attempt to reduce the domain disparity between different patients and to test the model under simulated clinical sampling conditions.
- Ensemble learning is introduced into this model by developing a hierarchical vote collective. Such a framework can improve the predictive performance and generalization ability due to the combination of multiple DA subnetworks.
- This study is the first study to evaluate the applicability of different DA algorithms for seizure prediction, which is crucial for follow-up studies.

Based on DA techniques and ensemble learning, the proposed model provides an above par disturbance rejection property, making the model more robust and practical for clinical application. Experiments on two public databases, the Freiburg Hospital EEG database and the CHB-MIT EEG database [34, 35], are conducted for model evaluation. Results indicate that HIVE-CODA achieves better domain adaptability compared with other state-of-the-art baselines.

## 2 DATA ACQUISITION AND PREPROCESSING

### 2.1 Patients

Two public EEG datasets, the Freiburg Hospital Intracranial EEG database [34] and the CHB-MIT scalp EEG database [35], are adopted to evaluate the generalization capability of HIVE-CODAs. The Freiburg Hospital EEG database includes time series of 87 seizures from 21 people with medically intractable focal epilepsy, ranging from 10 to 50 years old (8 male and 13 female patients). EEG signals are recorded invasively with six electrodes (3 near the epileptic focus and the other three distal to the epileptogenic zone). The sampling rate for all patients is 256 Hz (data of Patient No. 12 are sampled at 512 Hz but are down-sampled to 256 Hz).

The CHB-MIT database consists of scalp EEG sequences of 22 epileptic subjects, including five male patients ranging from 3 to 22 years and 17 female patients from 1.5 to 19 years. The EEG signals are recorded at a 256 Hz sampling rate with 16-bit analog-to-digital converters. Most samples are acquired from surface electrodes of 23 channels following the 10–20 standard system for

electrodes placement. Each patient has a subfolder that contains 9 to 42 recordings.

### 2.2 Data Selection and Labeling

Power line noise removal is implemented to denoise the data. We discarded the frequency bands of 47–53 and 97–103 Hz in the intracranial EEG set and the frequency bands of 57–63 and 117–123 Hz in the scalp EEG set. It is because noise commonly appears at 50 Hz for the Freiburg database and 60 Hz for the CHB-MIT database. Moreover, a subject selection is performed. Only patients with at least two seizures but fewer than 15 seizures per day are used for prediction, since less than two seizures are not enough to support training and more than 15 seizures make the forecast purportless. The chosen subjects are listed in **Tables 1, 2**.

A prerequisite for seizure prediction is the reliable distinction between preictal and interictal samples. We set 30 min before seizure onset as the seizure prediction horizon (SPH), which follows empirical evidence of comparison tests as applied multiple preictal lengths, and the seizure occurrence period is set to 0. A seizure should occur within 30 min after the predictor returns a positive. The raw EEG recordings are then divided into continuous, non-overlapping fragments by a 5-s time window. The sample number for each subject is sufficient (>7,200) to support training. Besides, we remarked that the amount of interictal samples is much larger than preictal samples. To remedy the sample imbalance, a random subsample on the interictal signals is performed to make an equal quantity of preictal and interictal training samples.

## 3 METHODS

To learn the domain-invariant representation, we proposed a generic seizure prediction model: the hierarchical vote collective of DA subnetworks (HIVE-CODAs). HIVE-CODA is an ensemble that combines 7 DA modules over multi-modality data. Each subnetwork is assigned a weight *via* the probabilistic voting scheme to equilibrate its value. By analyzing the most contributive DA component and its feature space, we provided a preliminary conclusion about the generalized information during the preictal period among various individuals.

### 3.1 Clinical Situation Simulation

Conventional approaches only provide the patient-specific results. Such frameworks may obtain high precision but are not consistent with the signal recording situation in real life. It is difficult to collect a large number of long-term EEG samples from one specific patient during clinical treatment, such that the sample size is unable to support the training process. Therefore, we considered using DA technology to apply data from other subjects to predictor training for the particular subject.

The training and testing strategy is depicted in **Figure 3**. The training and validation sets consist of existing patient data and one seizure of the target subject, while the remaining target seizures served as the testing set. The selection of seizure for

**TABLE 1** | Details of the Freiburg Hospital test set.

Patient	Gender	Age (years)	Seizure type	No. of seizures
Pt 1	F	15	SP	4
Pt 2	M	38	SP, CP, GTC	3
Pt 3	M	14	SP, CP	5
Pt 4-	F	26	SP, CP, GTC	5
Pt 5	F	16	SP, CP, GTC	5
Pt 6	F	31	CP, GTC	3
Pt 8	F	32	SP, CP	2
Pt 9	M	44	CP, GTC	4
Pt 10	M	47	SP, CP, GTC	5
Pt 11	F	10	SP, CP, GTC	4
Pt 12	F	42	SP, CP, GTC	3
Pt 13	F	22	SP, CP, GTC	2
Pt 14	F	41	CP, GTC	4
Pt 15	M	31	SP, CP, GTC	4
Pt 16	F	50	SP, CP, GTC	5
Pt 17	M	28	SP, CP, GTC	5
Pt 18	F	25	SP, CP	5
Pt 19	F	28	SP, CP, GTC	4
Pt 20	M	33	SP, CP, GTC	5
Pt 21	M	13	SP, CP	5

F, female; M, male; SP, simple partial; CP, complex partial; and GTC, generalized tonic-clonic.

training refers to the idea of the leave-one-out cross-validation (LOOCV) approach [36]. Moreover, the combined data are partitioned into five folds, and 80% of the samples are assigned to the training set, while the remaining 20% is named for the validation set to prevent overfitting.

### 3.2 Modular Hierarchical Structure

HIVE-CODAs include seven constituent modules: subject-invariant domain adaption (SIDA) [37], conditional deep convolutional generative adversarial networks (C-DCGANs) [38], plug-and-play domain adaptation (PPDA) [39], maximum independence domain adaptation (MIDA) [40], maximum mean discrepancy-adversarial autoencoders (MMD-AAEs) [41], model-agnostic learning of semantic features (MASF) [42], and cone manifold domain adaptation (CMDA) [43]. The modular hierarchical structure is depicted in **Figure 4**.

Since few domain adaptation techniques of epileptic EEG have been reported, we applied seven state-of-the-art approaches from the related fields to constitute the subnetworks of HIVE-CODAs. Several modules require images as inputs, instead of time series, such that we generate the spectrograms from EEG segments using the short-time Fourier transform (STFT) [44]. The raw EEG recordings are translated into two-dimensional matrices composed of frequency and time axes. Then the EEG fragments and their spectrograms will be sent forward to the corresponding modules depending on their modalities.

### 3.3 Modules Based on Adversarial Learning

1) *MMD-AAE*: We developed the MMD-AAE module referring to the study mentioned in reference [41], which aims at assessing the effectiveness of maximum mean discrepancy (MMD) measure and adversarial autoencoders (AAEs). An MMD-based regularization term is proposed to align the distributions

**TABLE 2** | Details of the CHB-MIT test set.

Patient	Gender	Age (years)	Seizure type	No. of seizures
Pt 1	F	11	SP, CP	7
Pt 2	M	11	SP, CP, GTC	3
Pt 3	F	14	SP, CP	6
Pt 5	F	7	CP, GTC	5
Pt 6	F	2	CP, GTC	4
Pt 7	F	15	SP, CP, GTC	3
Pt 8	M	4	SP, CP, GTC	5
Pt 9	F	10	CP, GTC	4
Pt 10	M	3	SP, CP, GTC	6
Pt 13	F	3	SP, CP, GTC	5
Pt 14	F	9	CP, GTC	5
Pt 17	F	12	SP, CP, GTC	3
Pt 18	F	18	SP, CP	6
Pt 19	F	19	SP, CP, GTC	3
Pt 20	F	6	SP, CP, GTC	5
Pt 21	F	13	SP, CP	4

F, female; M, male; SP, simple partial; CP, complex partial; and GTC, generalized tonic-clonic.

among various subjects. The AAE architecture is applied to learn the latent codes that are universal to all domains. The sharable information is captured by matching the aligned distribution to an arbitrary prior distribution. Thus the MMD-AAE may circumvent the overfitting to source data.

2) *SIDA*: We also estimated the performance of SIDA on epileptic EEG, which combines power spectral density (PSD) features and adversarial learning [37]. SIDA focuses on the extraction of the invariant representations among different domains. The sharable information is jointly learned with the task loss  $\mathcal{L}_{task}$  and subject confusion loss  $\mathcal{L}_{subj}$ . The training procedure adopts the adversarial strategy, which is implemented with a gradient reversal layer. Suppose that there are  $N$  source samples  $\{x_i\}_{i=1}^N$ , the process can be explicitly written as follows:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{task}(h(f(x_i; \theta); \gamma), c_i) - \lambda \frac{1}{N} \sum_{i=1}^N \mathcal{L}_{subj}(g(f(x_i; \theta); \phi), s_i), \quad (1)$$

$$\hat{\theta}, \hat{\gamma} = \arg \min_{\theta, \gamma} (\mathcal{L}(\theta, \gamma, \hat{\phi})), \quad (2)$$

$$\hat{\phi} = \arg \max_{\phi} (\mathcal{L}(\hat{\theta}, \hat{\gamma}, \phi)), \quad (3)$$

where  $\theta, \gamma, \phi$  represent the network parameters and  $\hat{\theta}, \hat{\gamma}, \hat{\phi}$  are their updated forms.  $\lambda$  is the trade-off positive parameter.  $h(\cdot)$  and  $g(\cdot)$  are the classification task and subject discrimination outputs.  $c_i, s_i$  denote the corresponding labels for  $\mathcal{L}_{task}$  and  $\mathcal{L}_{subj}$ , respectively. Note that, a specific feature extraction component is assembled in HIVE-CODAs since the inputs of SIDA are PSD features in accordance with the study mentioned in reference [37].

### 3.4 Modules Based on Data Augmentation

1) *C-DCGANs*: By introducing C-DCGANs [38], we tested the feasibility of using data augmentation and convolutional neural networks (CNN) to remedy the domain discrepancy. The main

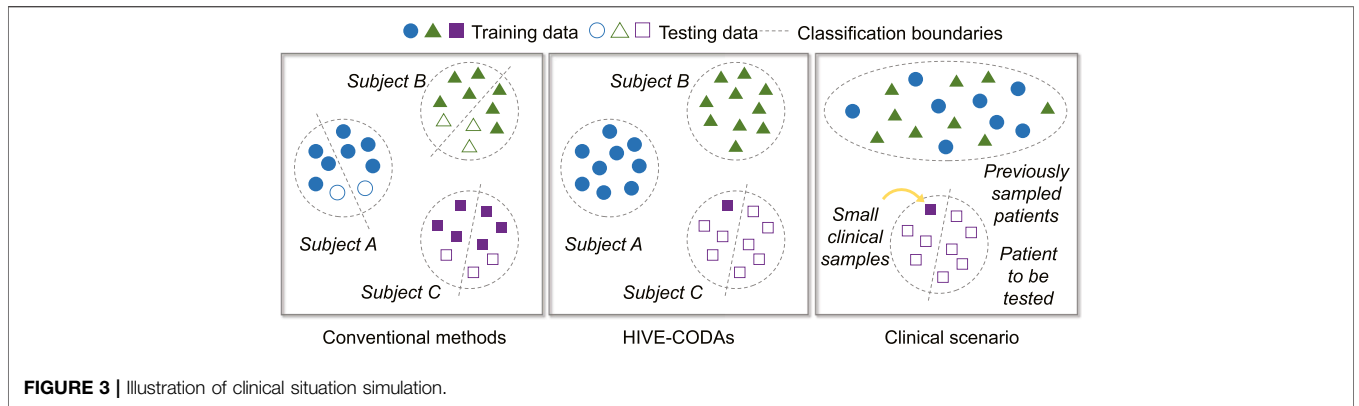


FIGURE 3 | Illustration of clinical situation simulation.

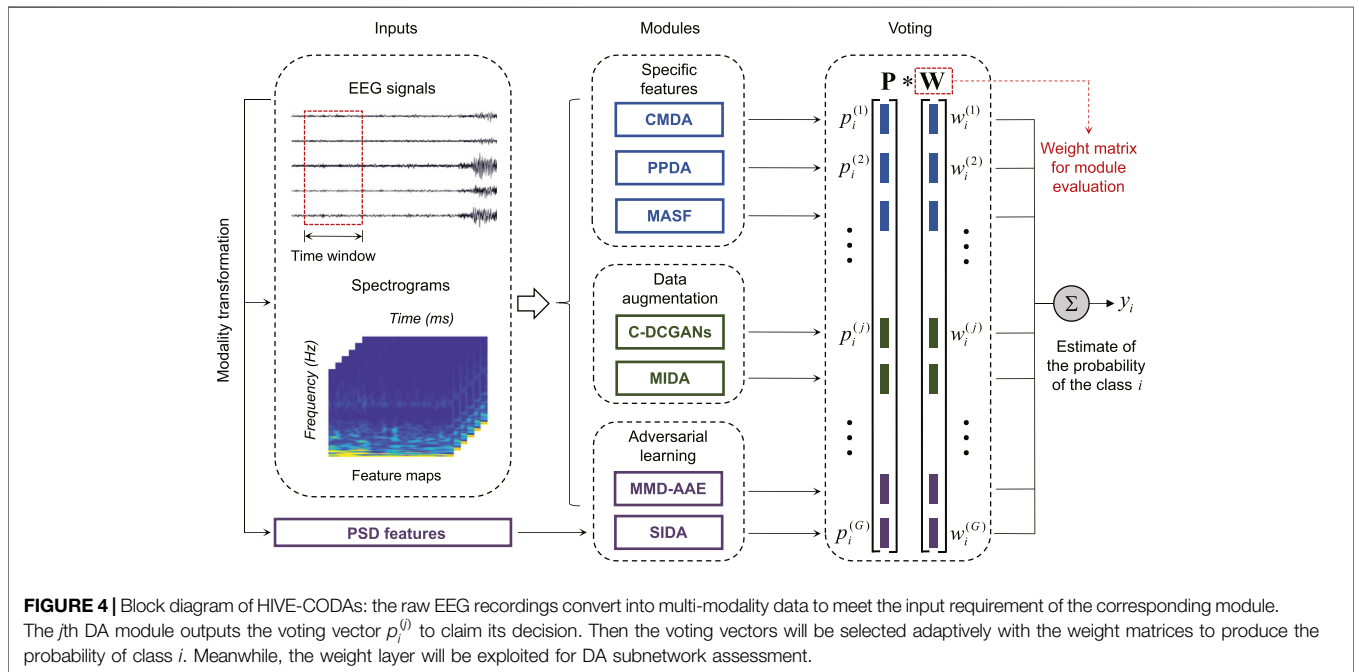


FIGURE 4 | Block diagram of HIVE-CODAs: the raw EEG recordings convert into multi-modality data to meet the input requirement of the corresponding module. The  $j$ th DA module outputs the voting vector  $p_i^{(j)}$  to claim its decision. Then the voting vectors will be selected adaptively with the weight matrices to produce the probability of class  $i$ . Meanwhile, the weight layer will be exploited for DA subnetwork assessment.

idea of C-DCGANs is increasing generalization capability via artificial EEG data generation. A generative adversarial network (GAN) is exploited to expand the training set, and an end-to-end CNN is employed as the classifier. We remarked that C-DCGANs also involve the adversarial learning due to the application of GAN. However, the generation function of GAN is highlighted in HIVE-CODAs, instead of the minmax optimization, such that we placed emphasis on the assessment of data augmentation.

2) *MIDA*: *MIDA* subnetwork is developed to measure the importance of background information and feature augmentation. In *MIDA* framework, an inner product space is established, where feature vectors are maximally independent in the sense of a Hilbert–Schmidt independence criterion (HSIC) [40]. The feature augmentation is performed via generating latent representations based on the background knowledge like acquisition time. The original feature vectors are expanded by concatenating with the produced features. According to the study mentioned in reference [40], we exploited the domain label

(which domain a sample belongs) as the background information since no device label and the acquisition time is provided in the epileptic EEG database.

### 3.5 Modules Based on Specific Features

1) *CMDA*: The *CMDA* module is adopted to evaluate the applicability of manifold on epileptic EEG. Referring to the study mentioned in reference [43], the latent feature space among various domains is regularized by modeling sharable information on the Riemannian cone manifold. Specifically, covariance matrices  $P$  of EEG segments are computed to constitute the manifold  $\mathcal{M}$ . The *CMDS* module leverages the global Riemannian mean  $\hat{P}$  and the local Riemannian mean  $\bar{P}^{(k)}$  to describe the cross-domain center and the centroid of the set  $\{P^{(k)} \mid P^{(k)} \in \mathcal{M}\}$  for the  $k$ th-domain (cite). By using the parallel transport approach, the projections of  $\{P^{(k)}\}$  on the tangent space  $\mathcal{T}_{\hat{P}}\mathcal{M}$  can describe the invariant features among source domains as follows:

**TABLE 3** | Results compared with conventional methods on the Freiburg Hospital database.

Source	Target	CNN		LSTM		HIVE-CODAs	
		$S_n$	FPR (/h)	$S_n$	FPR (/h)	$S_n$	FPR (/h)
S.C.	Pt 1	0.73	0.24	0.72	0.23	0.83	0.11
S.C.	Pt 2	0.58	0.31	0.66	0.30	0.85	0.09
S.C.	Pt 3	0.67	0.26	0.67	0.25	0.85	0.13
S.C.	Pt 4	0.67	0.27	0.71	0.16	0.84	0.12
S.C.	Pt 5	0.48	0.40	0.45	0.32	0.73	0.18
S.C.	Pt 6	0.75	0.27	0.58	0.38	0.86	0.11
S.C.*	Pt 8	0.55	0.34	0.53	0.29	0.66	0.32
S.C.	Pt 9	0.65	0.19	0.75	0.16	0.83	0.15
S.C.	Pt 10	0.51	0.37	0.58	0.26	0.79	0.24
S.C.	Pt 11	0.69	0.36	0.54	<b>0.21</b>	0.87	<b>0.22</b>
S.C.	Pt 12	0.66	0.18	0.65	0.24	0.84	0.17
S.C.*	Pt 13	0.56	0.31	0.52	0.23	0.68	0.27
S.C.	Pt 14	0.47	0.48	0.57	0.29	0.76	0.23
S.C.	Pt 15	0.66	0.19	0.70	0.17	0.88	0.12
S.C.	Pt 16	0.53	0.37	0.44	0.46	0.79	0.19
S.C.	Pt 17	<b>0.63</b>	0.36	0.42	0.41	<b>0.63</b>	0.28
S.C.	Pt 18	0.72	<b>0.18</b>	0.73	<b>0.20</b>	0.83	<b>0.20</b>
S.C.	Pt 19	0.44	<b>0.29</b>	0.47	<b>0.33</b>	0.75	0.14
S.C.	Pt 20	0.43	0.37	0.46	0.33	0.74	0.26
S.C.	Pt 21	0.64	0.32	0.50	0.28	0.84	0.12
Avg.		0.60	0.31	0.56	0.27	0.80	0.18

S.C., simulated clinical samples;  $S_n$ , sensitivity; FPR, false prediction rate; and Avg., average result. Note that S.C.\* uses NO samples of the predictor user. The bold values denote outliers.

**TABLE 4** | Results compared with conventional methods on the CHB-MIT database.

Source	Target	CNN		LSTM		HIVE-CODAs	
		$S_n$	FPR (/h)	$S_n$	FPR (/h)	$S_n$	FPR (/h)
S.C.	Pt 1	0.57	0.36	0.55	0.28	0.72	0.16
S.C.	Pt 2	0.48	0.39	0.42	0.27	0.65	0.22
S.C.	Pt 3	0.64	0.30	0.58	0.32	0.75	0.27
S.C.	Pt 5	0.54	0.42	0.45	0.41	0.76	0.23
S.C.	Pt 6	0.66	0.31	0.62	0.32	0.72	0.15
S.C.	Pt 7	0.65	<b>0.25</b>	0.55	<b>0.25</b>	0.83	<b>0.28</b>
S.C.	Pt 8	0.64	0.28	0.57	0.20	0.74	0.20
S.C.	Pt 9	0.48	0.35	0.45	<b>0.26</b>	0.63	<b>0.31</b>
S.C.	Pt 10	0.47	0.32	0.51	<b>0.24</b>	0.62	<b>0.34</b>
S.C.	Pt 13	0.55	<b>0.22</b>	0.50	<b>0.23</b>	0.73	<b>0.26</b>
S.C.	Pt 14	0.55	0.42	0.44	<b>0.32</b>	0.72	<b>0.35</b>
S.C.	Pt 17	0.47	0.4	0.42	0.42	0.62	0.28
S.C.	Pt 18	0.58	0.32	0.45	0.34	0.73	0.29
S.C.	Pt 19	0.59	0.22	0.53	0.21	0.79	0.17
S.C.	Pt 20	0.60	0.25	0.57	0.27	0.77	0.22
S.C.	Pt 21	0.62	0.28	0.62	0.27	0.77	0.15
Avg.		0.53	0.32	0.51	0.29	0.72	0.24

S.C., simulated clinical samples;  $S_n$ , sensitivity; FPR, false prediction rate; and Avg., average result. The bold values denote outliers.

$$\hat{\mathbf{S}}^{(k)} = \Gamma_{\hat{\mathbf{P}}^{(k)} \rightarrow \mathbf{P}^{(k)}}(\mathbf{S}^{(k)}), \forall k, \quad (4)$$

$$\mathbf{S}^{(k)} = \text{Log}_{\hat{\mathbf{P}}^{(k)}}(\mathbf{P}^{(k)}), \quad (5)$$

where  $\hat{\mathbf{S}}^{(k)}$  denotes the generalized features,  $\Gamma_{B \rightarrow A}(\cdot)$  represents the parallel transport from  $B$  to  $A$ , and  $\mathbf{S}^{(k)}$  represents the projection of  $\mathbf{P}^{(k)}$  on the tangent space  $\mathcal{T}_{\hat{\mathbf{P}}^{(k)}}\mathcal{M}$  with the logarithm map (cite). In general, each domain feature is parallelly transported from  $\mathbf{P}^{(k)}$  to the global centroid  $\hat{\mathbf{P}}$ , and the transported point  $\hat{\mathbf{S}}^{(k)}$  is embedded in the  $\langle \cdot, \cdot \rangle_{\hat{\mathbf{P}}}$  inner product space to make the generalized features describable in the Euclidean space.

2) *PPDA*: The long short-term memory (LSTM) architecture and a peculiar learning strategy are evaluated by adding the PPDA module. PPDA divides the latent features into private portions specific to each subject and generalized components among all subjects. To leverage both the universal and private feature vectors, PPDA develops a learning procedure including the training phase, calibration phase, and test phase. Specially, the LSTM layer is adopted for encoding and decoding.

3) *MASF*: To assess the applicability of meta-learning and semantic features, the MASF module is employed in HIVE-CODAs. According to the study mentioned in reference [42], a model-agnostic learning paradigm is exploited to minimize the domain gap *via* using a global class alignment loss  $\mathcal{L}_{global}$  and a local sample clustering loss  $\mathcal{L}_{local}$ . The knowledge about interclass relationships and the domain-independent class-specific cohesion/separation is captured by  $\mathcal{L}_{global}$  and  $\mathcal{L}_{local}$ , respectively, which is given as follows:

$$(\psi, \theta) \leftarrow (\psi, \theta) - \eta \nabla_{\psi, \theta} (\mathcal{L}_{task} + \mathcal{L}_{meta}), \quad (6)$$

$$\mathcal{L}_{meta} \leftarrow \beta_1 \mathcal{L}_{global} + \beta_2 \mathcal{L}_{local}, \quad (7)$$

where  $\psi, \theta$  are the network parameters,  $\eta$  is the learning rate, and  $\beta_1, \beta_2$  denote the weighting coefficients.  $\mathcal{L}_{task}$  represents the loss function of the predictive task. By introducing both global and local information, the semantic structure regarding the EEG feature space is regularized explicitly.

### 3.6 Weighted Voting Scheme

To evaluate the contribution of each subnetwork, a weighted voting structure is introduced at the end of the network. We assumed that there are  $G$  modules for the classification with  $C$  classes. For an arbitrary class  $y = i$ , we denote by  $w_i^{(j)}$  the weight assigned to the  $j$ th module, where  $i \in [1, \dots, C]$  and  $j \in [1, \dots, G]$ . Then the collective probability  $p_i$  for the  $i$ th class is the normalized weighted sum over modules:

$$p_i = \frac{\sum_{j=1}^G w_i^{(j)} p_i^{(j)}}{\sum_{i=1}^C \sum_{j=1}^G w_i^{(j)} p_i^{(j)}}. \quad (8)$$

The prediction result  $\hat{y}$  can be given as follows:

$$\hat{y} = \arg \max_i p_i. \quad (9)$$

The applicability of each algorithm on epileptic EEG could be estimated *via* observing its weight unit. Besides, a more balanced and intuitive collective could be created as the subnetworks are trained adaptively.

**TABLE 5** | Results compared with DA methods on the Freiburg Hospital database.

Source	Target	MIDA	MASF	PPDA	MMD-AAE	SAN	C-DCGANs	SIDA	CMDA	HIVE-CODAs
S.C.	Pt 1	0.59	0.63	0.74	0.78	0.78	0.80	0.81	0.80	0.86
S.C.	Pt 2	0.56	0.62	0.71	0.77	0.78	0.77	0.83	0.82	0.86
S.C.	Pt 3	0.55	0.62	0.67	0.74	0.75	0.80	0.77	0.83	0.84
S.C.	Pt 4	0.52	0.57	0.56	0.57	0.61	0.62	0.62	0.64	0.75
S.C.	Pt 5	0.60	0.63	0.72	0.79	0.79	0.80	0.80	0.82	0.85
S.C.	Pt 6	0.54	0.53	0.55	0.65	0.68	0.66	0.73	0.74	0.79
S.C.*	Pt 8	0.48	0.51	0.55	0.54	0.56	0.57	0.57	0.61	0.68
S.C.	Pt 9	0.50	0.55	0.61	0.69	0.68	0.70	0.70	0.73	0.80
S.C.	Pt 10	0.53	0.63	0.60	0.69	0.68	0.69	0.70	0.74	0.77
S.C.	Pt 11	0.62	0.65	0.70	0.77	0.81	0.73	0.83	0.82	0.87
S.C.	Pt 12	0.63	0.63	0.70	0.68	0.71	0.75	0.78	0.77	0.85
S.C.*	Pt 13	0.46	0.52	0.60	0.64	0.64	0.68	0.67	0.68	0.74
S.C.	Pt 14	0.48	0.55	0.63	0.65	0.66	0.69	0.68	0.71	0.78
S.C.	Pt 15	0.59	0.69	0.69	0.73	0.75	0.78	0.77	0.82	0.86
S.C.	Pt 16	0.45	0.48	0.58	0.62	0.67	0.69	0.67	0.70	0.76
S.C.	Pt 17	0.46	0.48	0.52	0.54	0.54	0.55	0.55	0.56	0.68
S.C.	Pt 18	0.61	0.64	0.68	0.75	0.77	0.79	0.78	0.81	0.86
S.C.	Pt 19	0.46	0.47	0.53	0.54	0.53	0.55	0.58	0.60	0.69
S.C.	Pt 20	0.50	0.55	0.60	0.67	0.69	0.68	0.71	0.70	0.78
S.C.	Pt 21	0.52	0.58	0.60	0.62	0.66	0.55	0.69	0.72	0.83
Avg.		0.53	0.58	0.63	0.64	0.65	0.69	0.71	0.73	0.80

S.C., simulated clinical samples. Note that S.C.\* uses NO samples of the predictor user.

## 4 RESULTS AND DISCUSSION

In this section, the comparison results and weight matrix are provided to verify the generalization ability and evaluate the DA algorithms. HIVE-CODA is assessed on both intracranial and scalp EEGs. We adopted three common measures for evaluation: sensitivity, false alarm rate per hour (FPR), and area under the receiver operating characteristic curve (AUC).

### 4.1 Generalization Ability Analysis

The comparison experiments are conducted to demonstrate the advantages of HIVE-CODAs over other conventional methods. Many time/frequency domain-based approaches have been applied to predict upcoming seizures. Two classic deep neural networks, CNN and LSTM, are selected to assess the generalization ability of our method, which have achieved success in patient-specific forecast. We also attempted to find a generic algorithm across different subjects for comparison. However, little existing research considers the similarity of data acquisition to clinical situation and uses plenty of “unseen” patient’s samples for training. The implementation details of CNN and LSTM refer to references [17] and [22], and the experimental results are listed in **Tables 3, 4**.

The experiment regarding the intracranial EEG is performed based on the widely used Freiburg Hospital database. **Table 3** illustrates that HIVE-CODAs achieve a sensitivity of 80% and an FPR of 0.18/h on average, which outperforms other forecast models. For the outlier-like Pt 17, HIVE-CODAs do not produce the desired effectiveness, which might be caused by a larger domain gap existing in the sample space.

Evidently, performances of all these prediction approaches show a significant decline compared with the patient-specific results in their literature. This phenomenon is reasonable since

the training and testing samples are collected from one same subject in prior studies, which consider little about the generalization ability. Conversely, our method is implemented based on the existing database and small amount of “unseen” patient’s data, which is more coincident with the real clinical situations. Though the precision is not relative high, the model performance is sufficient for the daily needs of patients, as it approximates to the first-in-man trial [5].

In terms of scalp EEG, the experiment is conducted based on the public CHB-MIT database, produced by the Massachusetts Institute of Technology. As shown in **Table 4**, HIVE-CODAs achieve a sensitivity of 72% and an FPR of 0.24/h on average. Since the conventional algorithms consider little about the domain shift among different patients, HIVE-CODAs exhibit obvious advantages over other prediction models. Still, for several outliers like Pt 2, Pt 9, Pt 10, and Pt 17, the sensitivity of our approach is slightly higher than the lower bound of a random binary classifier. HIVE-CODA is a variation over deep learning models. As such, it carries with it the uncertainties associated to deep neural networks, in particular a lack of formal convergence guarantees.

Furthermore, experiments compared with DA algorithms are conducted. Results of AUC value are listed in **Tables 5, 6**. For the Freiburg Hospital database, results indicate that HIVE-CODAs achieve higher generalization ability than the conventional algorithms. It also testifies to the application potential of integrated DA modules on processing epileptic EEG. To be specific, the interindividual variability could be alleviated, and the existing forecast systems could be transferred to the clinic due to the emerging technologies in DA.

For the CHB-MIT database, the conventional studies show a lower performance in a clear margin compared with their patient-specific results, which is consistent with the experiment

**TABLE 6** | Results compared with DA methods on the CHB-MIT database.

Source	Target	MIDA	MASF	PPDA	MMD-AAE	SAN	C-DCGANs	SIDA	CMDA	HIVE-CODAs
S.C.	Pt 1	0.60	0.64	0.65	0.73	0.75	0.77	0.77	0.80	0.86
S.C.	Pt 2	0.48	0.52	0.46	0.64	0.66	0.70	0.69	0.82	0.86
S.C.	Pt 3	0.54	0.53	0.59	0.65	0.67	0.68	0.71	0.83	0.84
S.C.	Pt 5	0.52	0.56	0.62	0.70	0.75	0.74	0.74	0.82	0.85
S.C.	Pt 6	0.52	0.55	0.61	0.66	0.74	0.75	0.73	0.74	0.79
S.C.	Pt 7	0.59	0.61	0.64	0.70	0.72	0.75	0.76	0.74	0.79
S.C.	Pt 8	0.51	0.58	0.61	0.67	0.70	0.71	0.71	0.61	0.68
S.C.	Pt 9	0.47	0.49	0.52	0.58	0.60	0.61	0.64	0.73	0.80
S.C.	Pt 10	0.46	0.50	0.49	0.51	0.55	0.54	0.60	0.74	0.77
S.C.	Pt 13	0.49	0.53	0.46	0.58	0.62	0.63	0.62	0.68	0.74
S.C.	Pt 14	0.47	0.53	0.55	0.63	0.66	0.68	0.70	0.71	0.78
S.C.	Pt 17	0.51	0.54	0.52	0.61	0.63	0.62	0.64	0.56	0.68
S.C.	Pt 18	0.50	0.51	0.53	0.58	0.61	0.62	0.66	0.81	0.86
S.C.	Pt 19	0.51	0.53	0.56	0.63	0.66	0.66	0.69	0.60	0.69
S.C.	Pt 20	0.55	0.56	0.59	0.65	0.68	0.72	0.74	0.70	0.78
S.C.	Pt 21	0.51	0.53	0.60	0.66	0.71	0.73	0.77	0.72	0.83
Avg.		0.51	0.54	0.56	0.64	0.66	0.67	0.69	0.70	0.74

S.C., simulated clinical samples.

conducted on intracranial EEG. Moreover, all these model performances drop to a varying degree compared with the precisions on the Freiburg test set. It might be caused by the advantages of ensemble learning in analyzing low spatial resolution recordings, namely, the scalp EEG signals. In other words, intracranial EEG recordings have the high spatial resolution and SNR, and the artifacts are typically seen in scalp EEG [46, 47]. This result also illustrates that HIVE-CODAs have superiority for processing complex time series due to the diverse inner pattern of the collective structure.

## 4.2 Module Performance Analysis

As few studies evaluate the applicability of different DA algorithms for seizure prediction, this study provides an analysis based on the adaptively trained weight matrices. HIVE-CODAs introduce several successful machine learning models from related fields and assess their performance. The subnetworks are conducted *via* a statistical analysis of the weighted voting layer. The weight distributions are presented in **Figure 5**. The greater the normalized weight, the greater the contribution of the corresponding DA module. DA methods with high contributions are considered to have larger potential. This study also tests the predictive precision of each module running alone (with the other modules' weights reset to 0). The results of AUC values are illustrated in **Figure 6**. A detailed discussion regarding these DA techniques is provided as follows.

1) *CMDA*: CMDA relies on Riemannian manifold-based features to capture the characteristic scale of the neuronal events, which was proposed for motor imagery. As shown in **Figure 5**, CMDA surpasses the other approaches on both intracranial and scalp EEG datasets. We conjectured that the inner pattern of EEG sequences may obey a compact distribution in the embedding space, such that the manifold-based methods that capture continuous subspace might be applicable to such task. The experimental result indicates that the analytic Riemannian manifold can potentially be used to develop a robust seizure predictor.

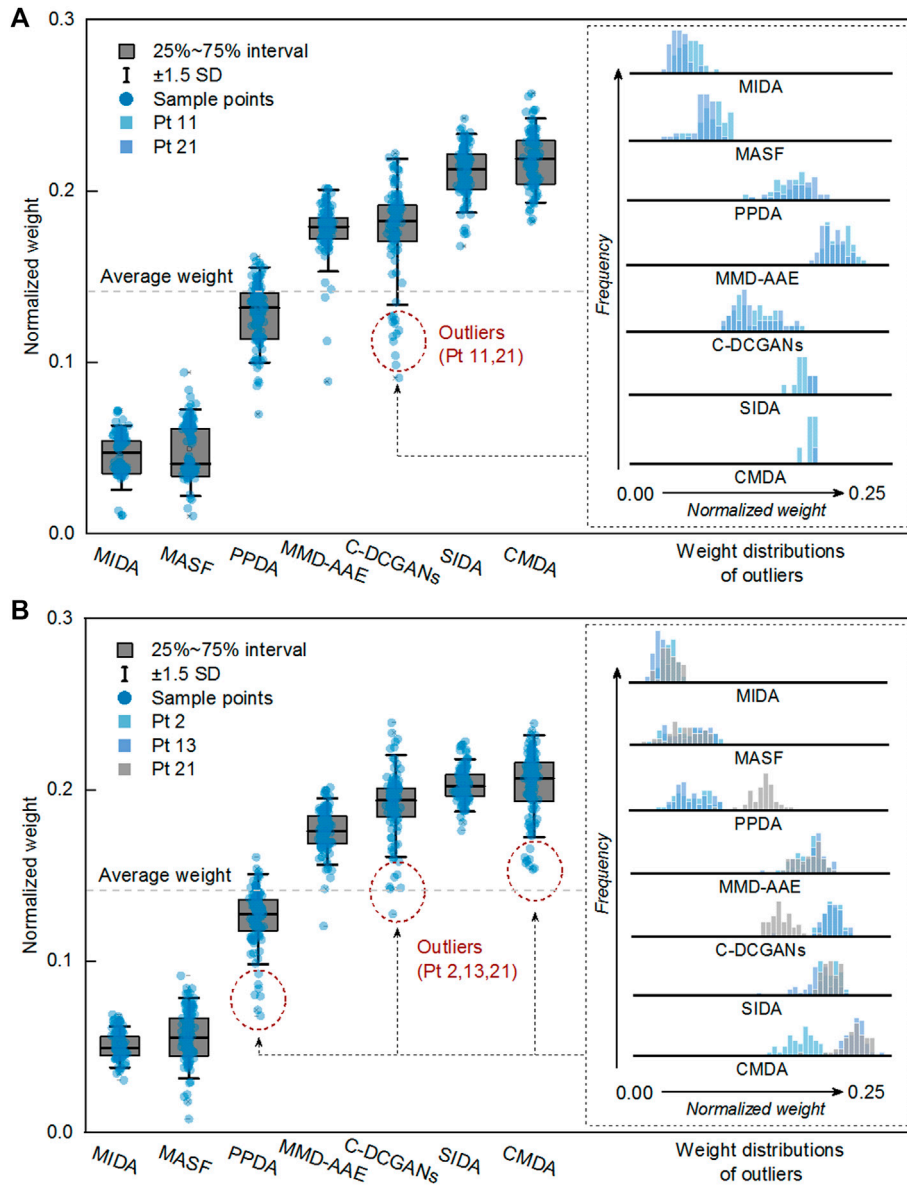
2) *SIDA*: The SIDA module is an adversarial neural network from the area of emotion recognition. It uses EEG spectra as input to learn a new representation, minimizing loss of emotion recognition and subject confusion. As we can see, SIDA makes relatively larger contribution compared with other modules. It might be due to the combination of CNN and generative adversarial network (GAN), which have been exploited to extract invariant latent features successfully. The weight unit of SIDA module may suggest the potential effect of adversarial learning on generalization ability, since SIDA exploited the architecture of GAN. However, this conjecture needs to be further verified for the SIDA module as it adopts the power spectral density (PSD) features as inputs.

3) *C-DCGANs*: C-DCGANs use conditional GANs to generate EEG artificially, which is developed for the detection of subject's movement intention (MI). We noticed that the performance of data augmentation-based module is not desired compared with the specific features and adversarial learning-based subnetworks. The degradation might be caused by the limitation of EEG data augmentation. The fake data usually involve more artifacts [48] that may contaminate EEG data. Still, C-DCGANs provide a decent accuracy, which suggests that data augmentation still has potential in developing a generic seizure forecast model.

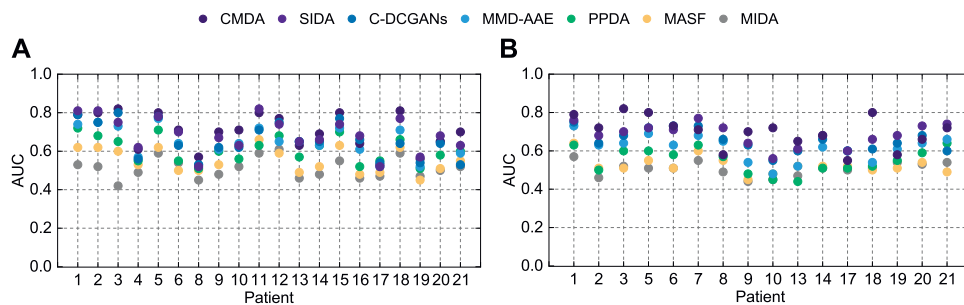
4) *MMD-AAE*: By matching the aggregated posterior with a prior distribution, the MMD-AAE module extracts the cross-domain features with adversarial learning. This scheme was originally used for image recognition. On both intracranial and scalp EEG datasets, MMD-AAE outperforms MIDA, MASFF, and PPDA modules and exhibits a slight decrease compared with the C-DCGAN module. Due to the above par performance of MMD-AAE, the conjecture in 2) about the superiority of adversarial learning is verified to some extent. This superiority may derive from the variational inference process of MMD-AAE, which alleviates the overfitting to the source domains effectively.

5) *PPDA*: PPDA is a technology applied to EEG-based emotion recognition. It uses LSTM-based encoder to

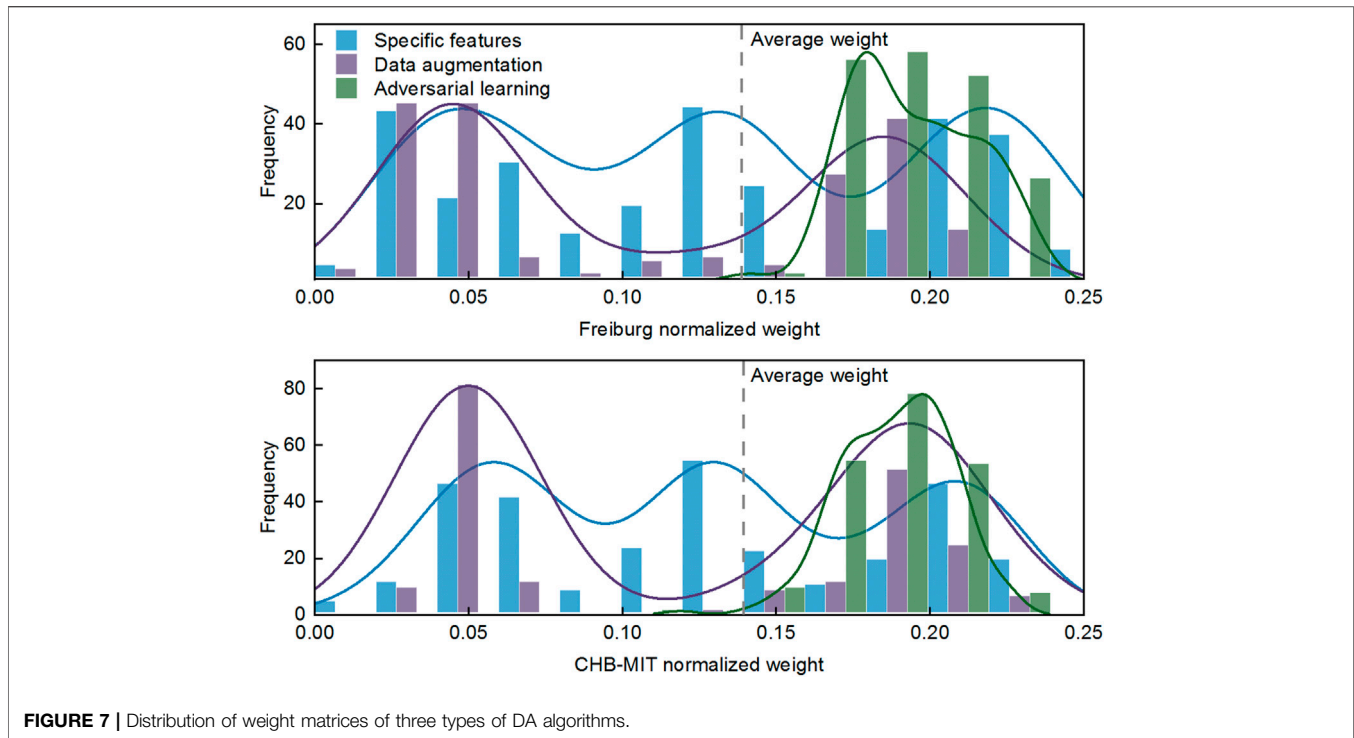




**FIGURE 5 |** Module performance analysis for the Freiburg Hospital (A) and CHB-MIT (B) database. It is calculated based on the learned weight vector of each DA subnetwork.



**FIGURE 6 |** AUC of different modules on the Freiburg Hospital test set (A) and the CHB-MIT test set (B).



**FIGURE 7** | Distribution of weight matrices of three types of DA algorithms.

decompose the features into general characteristics applicable to all individuals and personalized characteristics. Dividing the raw EEG data into subject-specific information and generalized information is a commonly-adopted strategy for domain adaptation. However, for seizure prediction, PPDA displays subpar performance, which we did not expect. Feature decomposition and the adoption of LSTM seem reasonable in these tasks. However, due to the few reports of relevant models, it cannot be absolutely determined that the decomposed features and LSTM are not suitable for epileptic signals.

6) *MASF*: MASF exploits semantic features and gradient-based meta-learning to establish a model-agnostic learning paradigm. In the field of image processing, successful application of MASF has been reported. Notably, the performance of the semantic feature-based method is unsatisfactory. We conjectured that the discriminant hyperplane in the feature space may be too complex to be adapted by the explicit semantic features. Moreover, the limitation of the initial neural architecture for meta-learning might also be a constraint of the search space.

7) *MIDA*: MIDA is originally applied in the emotion recognition field. The purpose of this model is to reduce differences in domain distributions by learning a subspace with maximum independence. **Figure 5** indicates that all the DA methods can outperform the MIDA module. This result was expected, given the limited background information on epileptic data. Obviously, the background-specific features are not valid characteristics.

Based on these results, we observed that adversarial learning and manifolds may achieve good performance in epilepsy prediction. In addition, CNN and PSD features may also have the potential to process epileptic signals. In the domain generalization field, CNN has gradually become one of the

most popular algorithms. This also further echoes the conjecture about CNN in this experiment. Note that the module performance may be variable on some special cases, since several outliers (in the Freiburg dataset, Pt 11, 21 for the C-DCGAN module; in the CHB-MIT dataset, Pt 2, 13 for the PPDA module, Pt 21 for C-DCGANs module, and Pt 2 for CMDA module) have been observed.

### 4.3 Model Applicability Analysis

Here, we attempt to summarize the universal characteristics and architectures based on the observations of DA algorithms in **Section 4.2**. The weight vectors of three types of DA methods (specific features, data augmentation, and adversarial learning) are quantified in a statistical analysis, which is depicted in **Figure 7**.

As shown in **Figure 7**, the adversarial learning-based approaches exhibit obvious advantages over the other DA methods on both intracranial and scalp EEG. Meanwhile, the weight distribution indicates that model performance regarding the specific features and data augmentation is volatile. Comparing adversarial learning to specific features, we reckon that the amount of above par weights is about 64% up. Comparing adversarial learning to data augmentation, a further 61% benefit is obtained, for a total of about 125% margin over data augmentation-based methods. These observations give us confidence in the efficacy of adversarial learning for processing epileptic signals, and we conjecture that data augmentation is relatively inferior for alleviating individual variability.

In particular, the manifold feature of CMDA surpasses all the other methods, such that the effectiveness of manifold feature

**TABLE 7** |  $\rho$  values on the Freiburg Hospital dataset.

Patient	Pt 1	Pt 2	Pt 3	Pt 4	Pt 5	Pt 6
Patient	Pt 1	Pt 2	Pt 3	Pt 4	Pt 5	Pt 6
$\rho$ value	0.032	0.041	0.034	0.034	<b>0.052</b>	0.044
Patient	Pt 8	Pt 9	Pt 10	Pt 11	Pt 12	Pt 13
$\rho$ value	0.037	0.032	0.038	0.028	0.032	0.032
Patient	Pt 14	Pt 15	Pt 16	Pt 17	Pt 18	Pt 19
$\rho$ value	0.034	0.014	0.045	<b>0.062</b>	0.025	<b>0.055</b>
Patient	Pt 20	Pt 21	—	—	—	—
$\rho$ value	0.047	0.041	—	—	—	—

The bold values denote outliers.

**TABLE 8** |  $\rho$  values on the CHB-MIT dataset.

Patient	Pt 1	Pt 2	Pt 3	Pt 5	Pt 6	Pt 7
Patient	Pt 1	Pt 2	Pt 3	Pt 5	Pt 6	Pt 7
$\rho$ value	0.046	<b>0.068</b>	0.038	<b>0.055</b>	0.028	0.033
Patient	Pt 8	Pt 9	Pt 10	Pt 13	Pt 14	Pt 17
$\rho$ value	0.032	<b>0.056</b>	0.048	0.040	0.044	0.038
Patient	Pt 18	Pt 19	Pt 20	Pt 21	—	—
$\rho$ value	0.035	0.025	0.028	0.028	—	—

The bold values denote outliers.

requires a further demonstration. The statistical significance of the manifold feature is assessed for discriminating preictal and interictal stages across different patients. The two-sample Kolmogorov–Smirnov test [49] at a 5% significance level ( $p < 0.05$ ) is implemented during the evaluation.

The significance analysis for each patient is provided in **Tables 7, 8**. The unqualified performance index is marked in bold format. For the manifold feature, 17 of 20 subjects in the Freiburg dataset and 13 of 16 subjects in the CHB-MIT dataset present an adequately distinguished ability. According to this observation, the manifold-based methods might be the promising techniques in developing a robust seizure predictor.

## 5 CONCLUSION

This study proposes a universal approach to alleviate the problem of individual variability in epileptic seizure prediction. By combining the DA and ensemble learning techniques, the

## REFERENCES

- Fisher RS, Boas WVE, Blume W, Elger C, Genton P, Lee P, et al. Epileptic Seizures and Epilepsy: Definitions Proposed by the International League against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* (2005) 46(4):470–2. doi:10.1111/j.0013-9580.2005.66104.x
- Banerjee PN, Filippi D, Allen Hauser W. The Descriptive Epidemiology of Epilepsy—A Review. *Epilepsy Res* (2009) 85(1):31–45. doi:10.1016/j.eplepsyres.2009.03.003
- Kwan P, Schachter SC, Brodie MJ. Drug-resistant Epilepsy. *N Engl J Med* (2011) 365(10):919–26. doi:10.1056/nejmra1004418

proposed HIVE-CODA model mitigates the effects of epileptic individual variance and increases the generalization ability. Besides, a simulated clinical sampling scenario is adopted during training and testing periods, which is the first attempt to adopt this evaluating strategy. Compared with the patient-specific scheme in conventional studies, such an assessment model is relatively demanding and challenging. Nonetheless, HIVE-CODAs achieve high domain shift robustness and precision, which demonstrates its feasibility of real-world applications.

By analyzing the contributions of each module, the experimental results also demonstrate the effectiveness of adversarial learning and manifolds in epileptic seizure prediction. The underlying causes of this phenomenon remain unclear because there is no definitive explanation of the dynamics of the manifold module in the existing literature. However, the success of the manifold module in this experiment brings new inspiration. We speculate that the mapping of EEG in the high-dimensional space may follow a compact distribution, so the kernel-based method for searching hyperplanes may have potential in this task. The search for more powerful DA algorithms and the underlying reasons will be considered as part of our future research extension to achieve higher performance.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. These data can be found here: <https://epilepsy.uni-freiburg.de/freiburg-seizure-predictionproject/eeg-database>.

## ETHICS STATEMENT

Written informed consent was obtained from the individual(s) and minor(s)' legal guardian/next of kin, for the publication of any potentially identifiable images or data included in this article.

## AUTHOR CONTRIBUTIONS

PP contributed to conception and design of the study, analysis and/or interpretation of data, and drafting the manuscript.

- Lin L-C, Ouyang C-S, Chiang C-T, Yang R-C, Wu R-C, Wu H-C. Early Prediction of Medication Refractoriness in Children with Idiopathic Epilepsy Based on Scalp EEG Analysis. *Int J Neur Syst* (2014) 24(07):1450023. doi:10.1142/s0129065714500233
- Cook MJ, O'Brien TJ, Berkovic SF, Murphy M, Morokoff A, Fabinyi G, et al. Prediction of Seizure Likelihood with a Long-Term, Implanted Seizure Advisory System in Patients with Drug-Resistant Epilepsy: a First-In-Man Study. *Lancet Neurol* (2013) 12(6):563–71. doi:10.1016/s1474-4422(13)70075-9
- Li S, Zhou W, Yuan Q, Liu Y. Seizure Prediction Using Spike Rate of Intracranial EEG. *IEEE Trans Neural Syst Rehabil Eng* (2013) 21(6):880–6. doi:10.1109/tnsre.2013.2282153

7. Karoly PJ, Freestone DR, Boston R, Grayden DB, Himes D, Leyde K, et al. Intercal Spikes and Epileptic Seizures: Their Relationship and Underlying Rhythmicity. *Brain* (2016) 139(4):1066–78. doi:10.1093/brain/aww019
8. Guo L, Wang Z, Cabrerizo M, Adjouadi M. A Cross-Correlated Delay Shift Supervised Learning Method for Spiking Neurons with Application to Intercal Spike Detection in Epilepsy. *Int J Neur Syst* (2017) 27(03):1750002. doi:10.1142/s0129065717500022
9. Shahidi Zandi A, Tafreshi R, Javidan M, Dumont GA. Predicting Epileptic Seizures in Scalp EEG Based on a Variational Bayesian Gaussian Mixture Model of Zero-Crossing Intervals. *IEEE Trans Biomed Eng* (2013) 60(5):1401–13. doi:10.1109/tbme.2012.2237399
10. Myers MH, Padmanabha A, Hossain G, de Jongh Curry AL, Blaha CD. Seizure Prediction and Detection via Phase and Amplitude Lock Values. *Front Hum Neurosci* (2016) 10:80. doi:10.3389/fnhum.2016.00080
11. Mirowski P, Madhavan D, LeCun Y, Kuzniecky R. Classification of Patterns of EEG Synchronization for Seizure Prediction. *Clin Neurophysiol* (2009) 120(11):1927–40. doi:10.1016/j.clinph.2009.09.002
12. Direito B, Teixeira CA, Sales F, Castelo-Branco M, Dourado A. A Realistic Seizure Prediction Study Based on Multiclass SVM. *Int J Neur Syst* (2017) 27(03):1750006. doi:10.1142/s012906571750006x
13. Sun C, Cui H, Zhou W, Nie W, Wang X, Yuan Q. Epileptic Seizure Detection with EEG Textural Features and Imbalanced Classification Based on Easyensemble Learning. *Int J Neur Syst* (2019) 29(10):1950021. doi:10.1142/s0129065719500217
14. Brinkmann BH, Wagenaar J, Abbot D, Adkins P, Bosshard SC, Chen M, et al. Crowdsourcing Reproducible Seizure Forecasting in Human and Canine Epilepsy. *Brain* (2016) 139(6):1713–22. doi:10.1093/brain/aww045
15. Zhang T, Chen W, Li M. Fuzzy Distribution Entropy and its Application in Automated Seizure Detection Technique. *Biomed Signal Process Control* (2018) 39:360–77. doi:10.1016/j.bspc.2017.08.013
16. Peng P, Xie L, Wei H. A Deep Fourier Neural Network for Seizure Prediction Using Convolutional Neural Network and Ratios of Spectral Power. *Int J Neur Syst* (2021) 31:2150022. doi:10.1142/s0129065721500222
17. Zhang Y, Guo Y, Yang P, Chen W, Lo B. Epilepsy Seizure Prediction on Eeg Using Common Spatial Pattern and Convolutional Neural Network. *IEEE J Biomed Health Inform* (2020) 24(2):465–74. doi:10.1109/JBHI.2019.2933046
18. Liu G, Zhou W, Geng M. Automatic Seizure Detection Based on S-Transform and Deep Convolutional Neural Network. *Int J Neur Syst* (2020) 30(04):1950024. doi:10.1142/s0129065719500242
19. Lin L-C, Ouyang C-S, Wu R-C, Yang R-C, Chiang C-T. Alternative Diagnosis of Epilepsy in Children without Epileptiform Discharges Using Deep Convolutional Neural Networks. *Int J Neur Syst* (2020) 30(05):1850060. doi:10.1142/s0129065718500600
20. Ozcan AR, Erturk S. Seizure Prediction in Scalp EEG Using 3D Convolutional Neural Networks with an Image-Based Approach. *IEEE Trans Neural Syst Rehabil Eng* (2019) 27(11):2284–93. doi:10.1109/tnsre.2019.2943707
21. Daoud H, Bayoumi MA. Efficient Epileptic Seizure Prediction Based on Deep Learning. *IEEE Trans Biomed Circuits Syst* (2019) 13(5):804–13. doi:10.1109/tbcas.2019.2929053
22. Tsiouris KM, Pezoulas VC, Zervakis M, Konitsiotis S, Koutsouris DD, Fotiadis DI. A Long Short-Term Memory Deep Learning Network for the Prediction of Epileptic Seizures Using EEG Signals. *Comput Biol Med* (2018) 99:24–37. doi:10.1016/j.combiomed.2018.05.019
23. Li Y, Yu Z, Chen Y, Yang C, Li Y, Allen Li X, et al. Automatic Seizure Detection Using Fully Convolutional Nested LSTM. *Int J Neural Syst* (2020) 30(4):2050019. doi:10.1142/s0129065720500197
24. Özcan AR, Ertürk S. Epileptic Seizure Prediction with Recurrent Convolutional Neural Networks. In: *Signal Processing and Communications Applications Conference* (2017). p. 1–4.
25. Peng P, Wei H, Xie L, Song Y. Epileptic Seizure Prediction in Scalp Eeg Using an Improved HIVE-COTE Model. In: *Chinese Control Conference*. IEEE (2020). p. 6450–7. doi:10.23919/ccc50068.2020.9188930
26. Jirsa VK, Proix T, Perdikis D, Woodman MM, Wang H, Gonzalez-Martinez J, et al. The Virtual Epileptic Patient: Individualized Whole-Brain Models of Epilepsy Spread. *Neuroimage* (2017) 145:377–88. doi:10.1016/j.neuroimage.2016.04.049
27. Kuhlmann L, Lehnertz K, Richardson MP, Schelter B, Zaveri HP. Seizure Prediction - Ready for a new era. *Nat Rev Neurol* (2018) 14(10):618–30. doi:10.1038/s41582-018-0055-2
28. Elger CE, Hoppe C. Diagnostic Challenges in Epilepsy: Seizure Under-reporting and Seizure Detection. *Lancet Neurol* (2018) 17(3):279–88. doi:10.1016/s1474-4422(18)30038-3
29. Samek W, Meinecke FC, Muller K-R. Transferring Subspaces between Subjects in Brain-Computer Interfacing. *IEEE Trans Biomed Eng* (2013) 60(8):2289–98. doi:10.1109/tbme.2013.2253608
30. Zhang L, Wang X, Yang D, Sanford T, Harmon S, Turkbey B, et al. Generalizing Deep Learning for Medical Image Segmentation to Unseen Domains via Deep Stacked Transformation. *IEEE Trans Med Imaging* (2020) 39(7):2531–40. doi:10.1109/tmi.2020.2973595
31. Pan SJ, Yang Q. A Survey on Transfer Learning. *IEEE Trans Knowledge Data Eng* (2009) 22(10):1345–59.
32. Long M, Cao Z, Wang J, Jordan MI, Conditional Adversarial Domain Adaptation. *Advances in Neural Information Processing Systems* (2018).
33. Combes RTd., Zhao H, Wang Y-X, Gordon G, Domain Adaptation with Conditional Distribution Matching and Generalized Label Shift. *Advances in Neural Information Processing Systems* (2020).
34. Zhou M, Tian C, Cao R, Wang B, Niu Y, Hu T, et al. Epileptic Seizure Detection Based on EEG Signals and CNN. *Front Neuroinform* (2018) 12:95. doi:10.3389/fninf.2018.00095
35. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* (2000) 101(23):e215–20.E215. doi:10.1161/01.cir.101.23.e215
36. Peng L-H, Yin J, Zhou L, Liu M-X, Zhao Y. Human Microbe-Disease Association Prediction Based on Adaptive Boosting. *Front Microbiol* (2018) 9:2440. doi:10.3389/fmicb.2018.02440
37. Rayatdoost S, Yin Y, Rudrauf D, Soleymani M. Subject-invariant EEG Representation Learning for Emotion Recognition. In: *IEEE International Conference on Acoustics, Speech and Signal Processing* (2021). p. 3955–9. doi:10.1109/icassp39728.2021.9414496
38. Zhang W, Yan F, Han F, He R, Li E, Wu Z, et al. Auto Recognition of Solar Radio Bursts Using the C-DCGAN Method. *Front Microbiol* (2021) 9:646556. doi:10.3389/fphy.2021.646556
39. Zhao L-M, Yan X, Lyu B. Plug-and-play Domain Adaptation for Cross-Subject EEG-Based Emotion Recognition. In: *Proceedings of the AAAI Conference on Artificial Intelligence* (2021).
40. Yan K, Kou L, Zhang D. Learning Domain-Invariant Subspace Using Domain Features and Independence Maximization. *IEEE Trans Cybern* (2018) 48(1):288–99. doi:10.1109/TCYB.2016.2633306
41. Li H, Pan SJ, Wang S, Kot AC. Domain Generalization with Adversarial Feature Learning. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018). p. 5400–9. doi:10.1109/cvpr.2018.00566
42. Dou Q, Castro DC, Kamnitsas K, Glocker B. Domain Generalization via Model-Agnostic Learning of Semantic Features” in *Advances In Neural Information Processing Systems* (2019).
43. Yair O, Ben-Chen M, Talmon R. Parallel Transport on the Cone Manifold of SPD Matrices for Domain Adaptation. *IEEE Trans Signal Process* (2019) 67(7):1797–811. doi:10.1109/tsp.2019.2894801
44. Gill V, Singh J, Singh Y. Analytical Solution of Generalized Space-Time Fractional Advection-Dispersion Equation via Coupling of Sumudu and Fourier Transforms. *Front Phys* (2019) 6:151. doi:10.3389/fphy.2018.00151
46. Usman SM, Khalid S, Akhtar R, Bortolotto Z, Bashir Z, Qiu H. Using Scalp EEG and Intracranial EEG Signals for Predicting Epileptic Seizures: Review of Available Methodologies. *Seizure* (2019) 71:258–69. doi:10.1016/j.seizure.2019.08.006
47. Ramantani G, Maillard L, Koessler L. Correlation of Invasive Eeg and Scalp EEG. *Seizure* (2016) 41:196–200. doi:10.1016/j.seizure.2016.05.018
48. Fahimi F, Dosen S, Ang KK, Mrachacz-Kersting N, Guan C. Generative Adversarial Networks-Based Data Augmentation for Brain-Computer Interface. *IEEE Trans Neural Networks Learn Syst* (2020).

49. Xiao Y. A Fast Algorithm for Two-Dimensional Kolmogorov-Smirnov Two Sample Tests. *Comput Stat Data Anal* (2017) 105:53–8. doi:10.1016/j.csda.2016.07.014

**Conflict of Interest:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of

the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

*Copyright © 2022 Peng. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.*