Check for updates

Seizure Prediction in EEG Signals Using STFT and Domain Adaptation

Peizhen Peng[1], Yang Song[2], Lu Yang[3] and Haikun Wei[1]*

[1] Key Laboratory of Measurement and Control of Control Science and Engineering (CSE), Ministry of Education, School of Automation, Southeast University, Nanjing, China, [2] State Grid Nanjing Power Supply Company, Nanjing, China, [3] Epilepsy Center, the Affiliated Brain Hospital of Nanjing Medical University, Nanjing, China
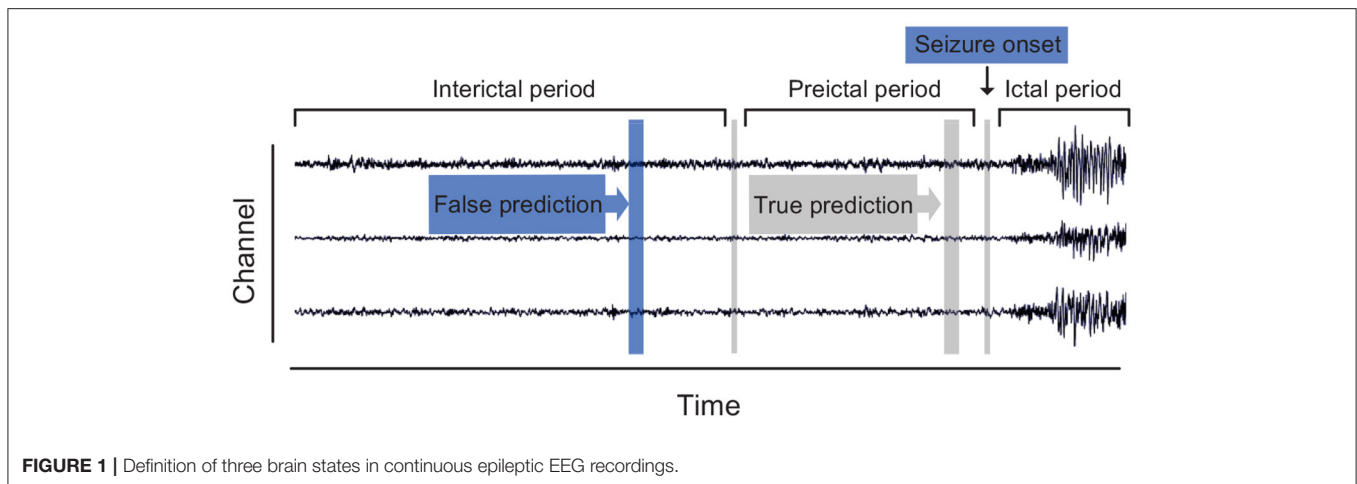
Epileptic seizure prediction is one of the most used therapeutic adjuvant strategies for drug-resistant epilepsy. Conventional approaches commonly collect training and testing samples from the same patient due to inter-individual variability. However, the challenging problem of domain shift between various subjects remains unsolved, resulting in a low conversion rate to the clinic. In this work, a domain adaptation (DA)-based model is proposed to circumvent this issue. The short-time Fourier transform (STFT) is employed to extract the time-frequency features from raw EEG data, and an autoencoder is developed to map these features into high-dimensional space. By minimizing the inter-domain distance in the embedding space, this model learns the domain-invariant information, such that the generalization ability is improved by distribution alignment. Besides, to increase the feasibility of its application, this work mimics the data distribution under the clinical sampling situation and tests the model under this condition, which is the first study that adopts the assessment strategy. Experimental results on both intracranial and scalp EEG databases demonstrate that this method can minimize the domain gap effectively compared with previous approaches.

Keywords: seizure prediction, feature extraction, neuropsychiatric disorders, domain adaptation, STFT, EEG

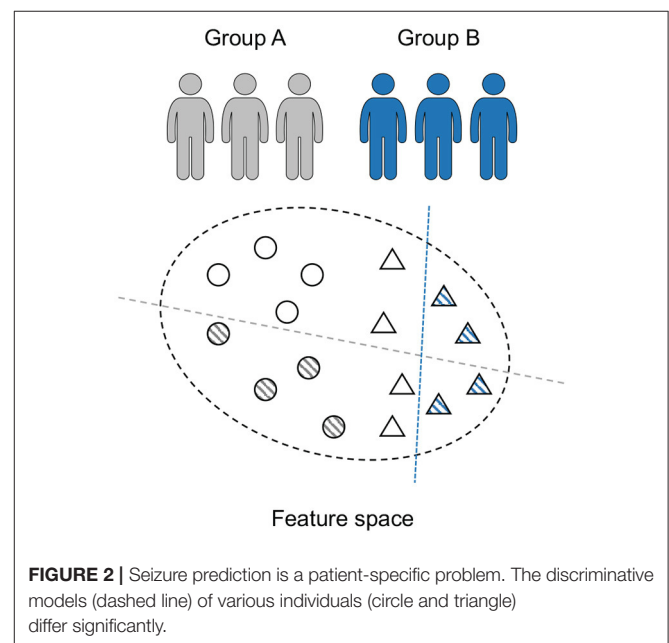## 1. INTRODUCTION

1.1. Epilepsy Background

Epilepsy is a brain disorder characterized by the transient occurrence of unexpected seizures, which stems from excessive, or hypersynchronous neuronal activities (Fisher et al., 2005). It affects approximately 1.0% of the world's population (Banerjee et al., 2009), and around half of them experience severe seizures. Besides, although the anti-epileptic drug (AED) administration is applied to patients, about 30% of them suffer from drug-resistant epilepsy (Kwan et al., 2011; Lin et al., 2014). These individuals might have seizures at any moment, such that their daily lives are influenced by unexpected behavioral changes, loss of muscular control and sudden faint. As a result, a reliable seizure prediction device is becoming an emerging and significant demand to prevent the injury of epileptic coma, or even death. A successful seizure forecast commonly adopts data-driven techniques to monitor the electroencephalography (EEG) signals of the epileptic brain, since such data records rhythmic information induced by coordinated neuronal. The first-in-man study that proves the predictability of seizure has been reported in 2013 (Cook et al., 2013). Since then, many EEG-based studies regarding seizure prediction have been proposed.

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

1.2. Related Work

At present, there are two main streams for epileptic seizure prediction. The first stream is a binary classification framework trained to discriminate preictal samples from interictal samples. The ictal and postictal samples are deserted during the training procedure for the uselessness of their contribution to forecast. This stream is widely adopted among researchers in the area of seizure prediction. The second stream assumes that a specific index that fluctuates with changes of seizure stage exists in EEG recordings. These methods attempt to describe this index explicitly and monitor it with a threshold. For instance, spike rate (Li et al., 2013; Karoly et al., 2016; Guo et al., 2017), zero-crossing intervals (Zandi et al., 2013), and phase/amplitude locking value (Myers et al., 2016) have been reported as the indicators. Since a universal preictal biomarker has not been defined explicitly, we also follow the binary scheme of the first stream, which is depicted in **Figure 1**.

Studies adopting the binary classifier usually combine the features extraction algorithms with machine learning techniques. To be specific, the features extraction algorithms are commonly used in data preprocessing due to the complexity and diversity of EEG signals, and then the machine learning techniques can analyze these features and give their categories. Features extraction approaches like wavelet transform (Vahabi et al., 2015; Moctezuma and Molinas, 2020), Q-factor wavelet transform (Al Ghayab et al., 2019), Fourier neural network (Peng et al., 2021), and fractional Fourier transform (Fei et al., 2017), are employed to learn the high-dimensional representations of samples. Machine learning techniques like support vector machines (Mirowski et al., 2009; Direito et al., 2017; Sun et al., 2019), random forests (Brinkmann et al., 2016), *k*-nearest neighbor (Zhang et al., 2018), and ensemble learning (Peng et al., 2020) are utilized to learn the spatial and temporal representations of seizures. Besides, recently most authors apply deep learning frameworks for seizure prediction. Convolutional neural network (CNN) (Zhang et al., 2019; Lin et al., 2020; Liu et al., 2020), 3D CNN (Ozcan and Erturk, 2019), Long Short-Term Memory (LSTM) Network (Tsiouris et al., 2018; Daoud and

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

Bayoumi, 2019; Li et al., 2020), and cascades of DNN (Özcan and Ertürk, 2017), are introduced to process continuously acquired EEG signals.

1.3. Significance

Although conventional studies report high precision (< 85% on average) for the epileptic seizure prediction task, their translation to the practical application is still a challenging issue. The major reason for this situation is that most of these studies only provide patient-specific results. For these patient-specific models, both training and testing sets are recorded from one subject, which leads to limited domain adaptability of previous approaches between different patients (each patient is considered as a domain). As shown in **Figure 2**, for the patients with epilepsy, the internal patterns vary significantly among various subjects

(Jirsa et al., 2017; Elger and Hoppe, 2018; Kuhlmann et al., 2018), which learn totally different discriminative hyperplanes. Therefore, these patient-specific models with good performance might obtain undesired results in real life, although they are significant to personalized medicine. It is obvious that how to develop a predictor that is universal to various patients is the key problem. However, this issue remains unsolved and thus the previous models are not yet in widespread use.

This work attempts to develop a seizure prediction model without the precondition of patient specificity. However, since the underlying patterns and dynamics of epilepsy are not well-understood in neuroscience, the complete desertion of the "target" samples (data of the patient to be tested) is impossible. The training set in existing studies is composed of the "target" data entirely. We attempt to reduce the reliance on the "target" data as much as possible until it reaches a clinically acceptable amount of EEG recordings. To be specific, due to the risk of infection in invasive surgery and the right of privacy, the training set in real life mainly consist of signals of previous patients. And only a small amount of "target" samples can be used for training. We try to simulate this sampling situation to train and test our model.

To achieve a higher generalizationa ability, this study introduces the strategy of domain adaptation (DA) methods for seizure prediction. It is a machine learning technique that can reduce the domain gap. In some successful DA models like maximum mean discrepancy-adversarial autoencoders (MMD-AAE) (Li et al., 2018) and cone manifold domain adaptation (CMDA) (Yair et al., 2019), to minimize the inter-domain distance in a high-dimensional space is the major optimization objective. Inspired by these researches, we hope to develop a generic seizure prediction model based on minimizing the inter-domain distance. To encode the raw EEG data into a high-dimensional space, we design a novel autoencoder using the short-time Fourier transform (STFT) (Cordes et al., 2021) and the MMD-AAE. The main contributions of this study are summarized as follow

- A general seizure prediction model for different patients is proposed based on the MMD-AAE model and STFT technique. It is tested under simulated clinical sampling conditions, making it feasible in practice.
- A domain adaptation framework is developed based on inter-domain distance. This algorithm can improve the generalization ability since it minimizes the domain gap between different patients.
- It is the first study to provide a comparison results of different domain adaptation algorithms on seizure forecast, which is important to follow-up researches.

Based on the MMD-AAE model and STFT technique, the proposed method obtains an above-par generalization ability. Experiments on two open datasets, the Freiburg Hospital EEG database and the CHB-MIT EEG database (Goldberger et al., 2000; Zhou et al., 2018), are conducted for model assessment. Compared with other methods, experimental results indicate that the proposed model achieves high robustness while preserving a decent precision.

2. DATA ACQUISITION AND PREPROCESSING

2.1. Patients
Two open EEG databases, the Freiburg Hospital intracranial EEG database (Zhou et al., 2018) and the CHB-MIT scalp EEG database (Goldberger et al., 2000), are selected to assess the model performance of our method.

The Freiburg Hospital database includes time series of 87 seizures of 21 subjects with medically intractable focal epilepsy, aged from 10 to 50 years old (8 males and 13 females). EEG signals are recorded invasively by 6 electrodes (3 near the epileptic focus and 3 away from the epileptogenic zone). The sampling rate is 256 Hz for all patients(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 5 males aged from 3 to 22 years and 17 females aged from 1.5 to 19 years. The EEG signals are recorded at a 256 Hz sampling rate using 16-bit analog-to-digital converters. Most samples are collected from 23 channel surface electrodes following the 10-20 standard system for electrode placement (Rojas et al., 2018). Each individual has a subfolder that has 9 to 42 recordings.

2.2. Data Selection and Labeling
We use the power line noise removal to denoise the raw EEG recordings. In the intracranial EEG test set, the frequency bands of 47–53 and 97–103 Hz are deserted and in the scalp EEG test set, the frequency bands of 57–63 and 117–123 are discarded. This is because that the noise of the Freiburg database usually occurs at 50 Hz and noise of the CHB-MIT database occurs at 60 Hz. In addition, we perform a patient selection. Only patients who had at least 2 seizures but fewer than 15 per day are chosen, since less than 2 seizures would not be sufficient to support training, and more than 15 would render the prediction meaningless. The subjects chosen in this work are presented in **Tables 1**, **2**.

The seizure occurrence period (SPO) is set to 0. Only the seizure prediction horizon (SPH) is considered in this study. Thirty minutes before seizure occurrence is set as the SPH. This parameter is given by empirical evidence of comparison experiments applying multiple preictal lengths. If a seizure occurs within 30 min, the forecast model then returns a positive. The raw EEG recordings are split into continuous, non-overlapping segments over a 5-s time window. The sample number for each case is sufficient (> 38,400) to support training. In addition, to obtain equal amounts of preictal and interictal samples, a random subsample on the interictal data is implemented, which circumvents the imbalance of different kinds of training samples.

3. METHODS

To reduce the impact of inter-individual variability, we propose a generic seizure prediction model. The core idea of our method is to minimize the domain distance between different subjects in the high-dimensional space. Such that domain-invariant features can be extracted during domain distribution alignment. The maximum mean discrepancy (MMD) measure (Zhang

TABLE 1 | Details of the Freiburg Hospital test set.

Patient	Gender	Age	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; GTC, generalized tonic-clonic.

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

Patient	Gender	Age	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; GTC, generalized tonic-clonic.

et al., 2020) is selected as the distance measure and the high-dimensional space is established by the adversarial autoencoders (AAE) (Makhzani et al., 2015).

3.1. Clinical Situation Simulation

The training set of conventional studies is not consistent with the sampling situation in real life. During clinical treatment, it is almost impossible to record a large number of EEG samples from a specific patient over a long period of time. Thus the traditional patient-specific learning strategy can not be performed because the data size is unable to support training. To tackle this issue, we propose a novel predictor that can use other patients' data for training.

To mimic the sampling situation in the clinic, we adopt a particular training and testing strategy, which is illustrated in **Figure 3**. To be specific, the training and validation set includes previous patient' data and one seizure from the "target" subject, while the remaining seizures of the "target" subject served as the test set. This strategy refers to the idea of the Leave-one-out cross-validation (LOOCV) approach (Peng et al., 2018). Moreover, the training and validation set are partitioned into 5-folds, and 80% of the data are assigned to the training set while the remaining 20% are assigned to the validation set to prevent overfitting.

3.2. Modal Transformation With STFT

Due to the low signal-to-noise ratio (SNR) of EEG recordings, we attempt to transform the input information from time domain into time-frequency domain. Two preprocessing techniques, wavelet and Fourier transforms (Muralidharan et al., 2011; Zhao et al., 2019), are commonly employed to convert EEG segments into image shapes. Here we adopt the short-time Fourier transform (STFT) to produce feature maps from raw EEG sequences. The conversion transforms the EEG time series into matrices, which can meet the input requirement of the two-dimensional MMD-AAE. This procedure can also extract the significant features for seizure prediction.
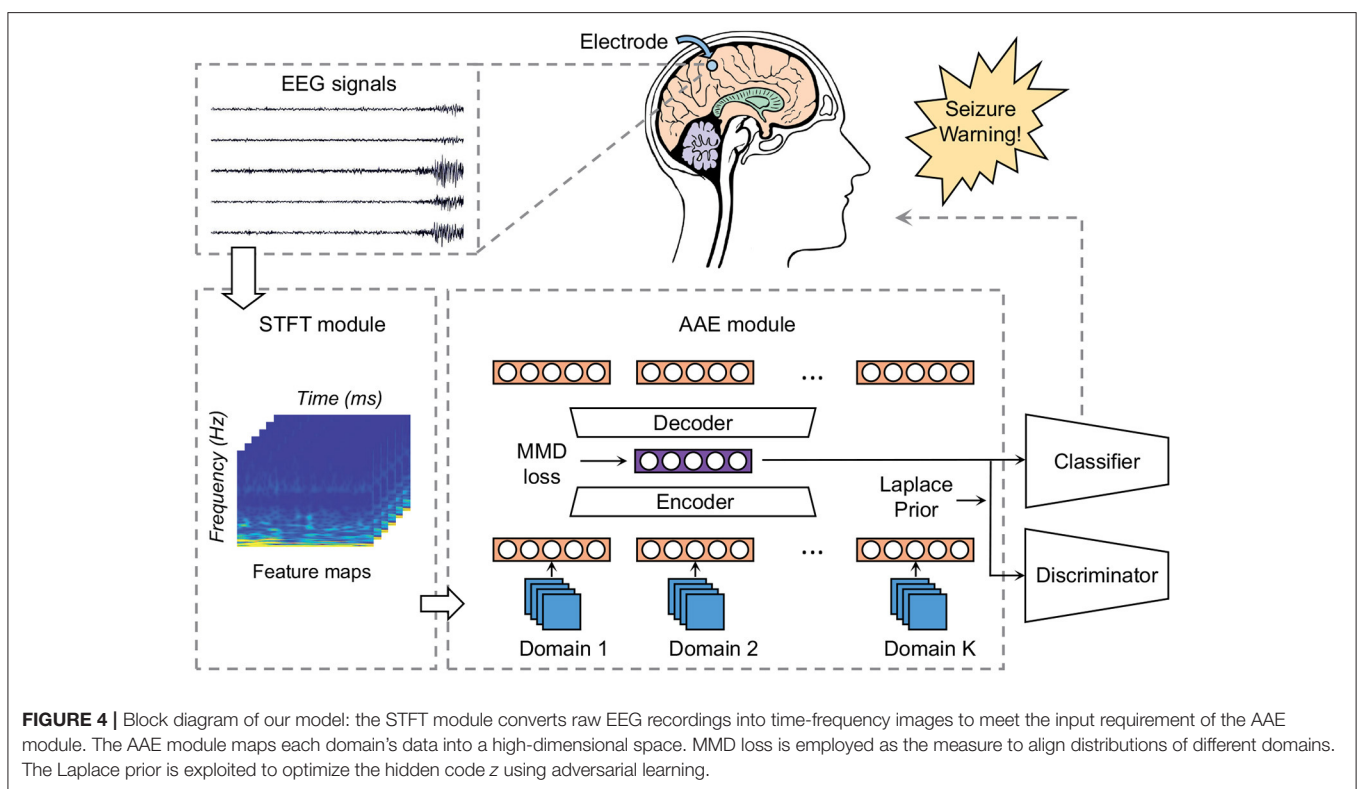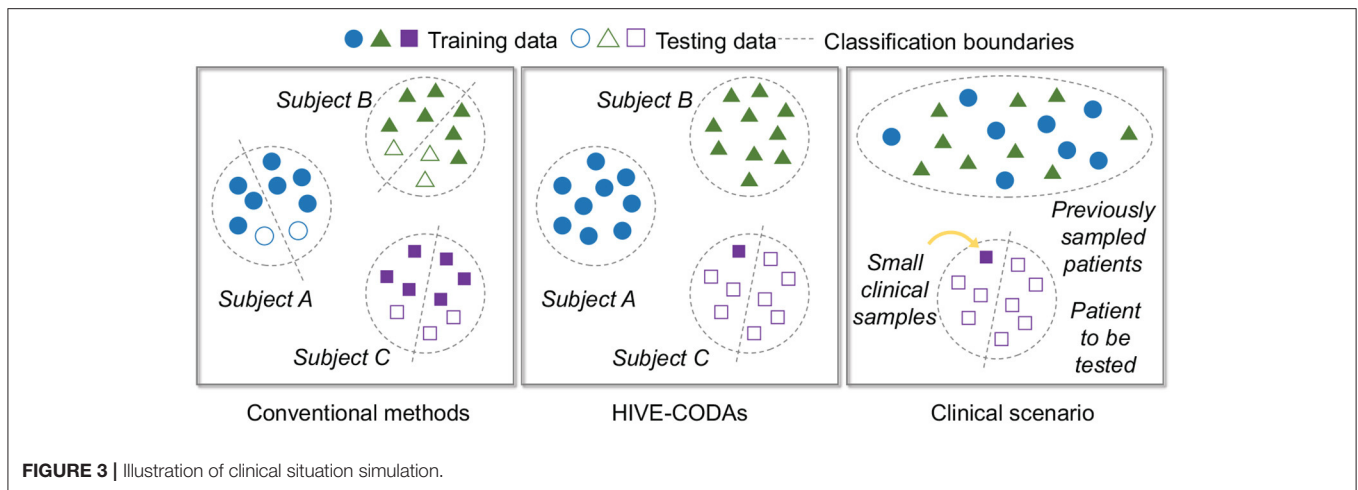
In the STFT module, the time-varying EEG fragment is converted to a two-dimensional matrix composed of frequency and time axes. Such that an insight in the time-evolution for each time window can be observed by the two-dimensional map. Suppose that there exist K domains (patient) in total. The input data from the K domains are denoted by $\bar{X} = [\bar{x}^{(1)}, \cdots, \bar{x}^{(K)}]^T \in \mathbb{R}^{K \times d}$, where $\bar{x} \in \mathbb{R}^{d \times 1}$. For an arbitrary domain, the segment with the time index t is given as $\bar{x}(t)$. By performing the STFT procedure, we get the time-frequency feature map of $\bar{x}(t)$, which is presented as:

$$x(\omega, u) = \sum_t \bar{x}(t)g(t-u)e^{j\omega t}, \tag{1}$$

where ω is the selected frequency band, $g(t - u)$ is the window function. For K domains, the STFT is implemented to each subject, and then the inputs are converts to $X = [x^{(1)}, \cdots, x^{(K)}]^T \in \mathbb{R}^{K \times d}$. The samples of each case are represented by spectrograms. These time-frequency feature maps are then sent to the AAE for invariant feature extraction.

3.3. Construction of High-Dimensional Space

This module attempts to establish a high-dimensional space with an encoder and a decoder. The model is illustrated in **Figure 4**. By using an encoder, we can map the time-frequency images of raw EEG samples into an embedding subspace. And by using a decoder, these hidden layers are then mapped back

FIGURE 3 | Illustration of clinical situation simulation.

FIGURE 4 | Block diagram of our model: the STFT module converts raw EEG recordings into time-frequency images to meet the input requirement of the AAE module. The AAE module maps each domain's data into a high-dimensional space. MMD loss is employed as the measure to align distributions of different domains. The Laplace prior is exploited to optimize the hidden code *z* using adversarial learning.

to a "fake" input matrix. The hidden space is high-dimensional and therefore contains more information. The MMD measure is then utilized to align the distributions of high-dimensional feature vectors between different domains. Thus the optimized hidden code contains sharable information of various patients. We then extract these latent characteristics that are universal among patients for classification.

There are two procedures during the construction of embedding space: the reconstruction process and the distribution alignment process. In the reconstruction process, the autoencoder attempts to recover the time-frequency image

from the high-dimensional vectors. The architecture of the autoencoder refers to the structure in MMD-AAE (Li et al., 2018). We set an optimization objective $\mathcal{L}_{rec}$ to guide the generated feature map $\tilde{x}$ to match the input map x. The loss function of the reconstruction procedure $\mathcal{L}_{rec}$ is defined as:

$$\mathcal{L}_{rec} = \left\| \tilde{x} - x \right\|_2^2. \tag{2}$$

Now, we specify the form of the inter-domain distance metric. The maximum mean discrepancy (MMD) measure (Jia et al., 2017) is exploited to align the distributions of different domains. Like the reconstruction process, we

also give a MMD-based regularization term to optimize the hyperparameters in the neural network (Li et al., 2018). Suppose that $\mathbf{Z} = [\mathbf{z}^{(1)}, \cdots, \mathbf{z}^{(K)}]^T \in \mathbb{R}^{K \times l}$ represents the learned high-dimensional features of K domains, where $\mathbf{z} \in \mathbb{R}^{l \times 1}$. For two arbitrary hidden vector $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(j)}$, we assume that they belong to two unseen probability distribution $\mathbb{P}^{(i)}$ and $\mathbb{P}^{(j)}$, respectively. By adopting the kernel embedding technique (Smola et al., 2007), the instance is mapped to a reproducing kernel Hilbert space (RKHS) $\mathcal{H}$. The corresponding mean value in RKHS is given as:

$$\mu(\mathbb{P}) = \mathbb{E}_{\mathbf{z} \sim \mathbb{P}}[h(\mathbf{z}, \cdot)], \qquad (3)$$

where is $\mu(\cdot)$ the mean map operation. $h(\mathbf{z}, \cdot)$ is the kernel function induced by the feature map in $\mathcal{H}$. In this work, we adopt the RBF kernel following the MMD-AAE model (Li et al., 2018). The inter-domain distance between the latent codes $\mathbf{z}^{(i)}$ and $\mathbf{z}^{(j)}$ can be described as:

$$dis(\mathbf{z}^{(i)}, \mathbf{z}^{(j)}) = \left\| \mu_{\mathbb{P}^{(i)}} - \mu_{\mathbb{P}^{(j)}} \right\|_{\mathcal{H}}. \qquad (4)$$

Then, it is obvious that the regularization item of the entire latent space can be defined as:

$$\mathcal{R}_{dis}(\mathbf{z}^{(1)}, \cdots, \mathbf{z}^{(K)}) = \frac{1}{K^2} \sum_{1 \le i,j \le K} dis(\mathbf{z}^{(i)}, \mathbf{z}^{(j)}). \qquad (5)$$

With the distance error above, the extracted high-dimensional features can generalize well across all the domains, since the neural network learns their common code by aligning their distributions.

3.4. Optimization Using Adversarial Learning

To further optimize the learned features in section 3.3, we introduce an adversarial learning-based module according to AAE (Makhzani et al., 2015). Adversarial learning is an emerging machine learning approach in recent years, which has been successfully applied in the area of epileptic EEG signal processing. It usually contains a generator network G with parameters Θ_G and a discriminator network D with parameters Θ_D. The generator network G will produce some fake version of the inputs. These fake data are sent to the discriminator network D together with the real input data. Then the discriminator network D will tell whether the input sample is artificially generated. During the training procedure, the neural network finds a Nash equilibrium between the generator and the discriminator, and the fake data are gradually approaching the real one. This "zero-sum game" can be described as:

$$\min_{G} \max_{D} \mathbb{E}_{\mathbf{x}_r \sim p_r}\left[\log D_{\Theta_D}(\mathbf{x}_r) \right] - \mathbb{E}_{\mathbf{x}_f \sim p_f}\left[\log D_{\Theta_D}(\mathbf{x}_f) \right], \quad (6)$$

where $\mathbf{x}_r$ and p_r are the real data and the corresponding distribution, $\mathbf{x}_f$ and p_f are the fake version. After the optimization with the loss $\mathcal{J}_{gan}$, the adversarial subnetwork can align p_f to the constant prior p_r.

We hope to utilize the aforementioned principle in the embedding space. Therefore, we assume that the "true" universal

features among different patients comes from an arbitrary prior distribution $p(\mathbf{z})$. The adversarial module draws samples from the prior distribution $p(\mathbf{z})$ and considers these samples as the real data $\mathbf{z}_r$. Accordingly, the learned latent information $\mathbf{z}$ is considered as the fake data $\mathbf{z}_f$, where the autoencoder represents the generator G. A discriminator D is also implemented in the adversarial module, which distinguishes the produced vector $\mathbf{z}$ from the samples of the prior. In this study, the prior distribution is selected as the Laplace distribution $\mathbf{z} \sim Laplace(\eta)$, where η denotes the hyperparameter.

The training strategy of the adversarial module is a variational inference process. To be specific, first, the latent coding space has been established by the encoder explicitly. Then the distributions among different domains are aligned with the MMD regularization item to extract the domain-invariant feature vectors. These features are guided to approach a prior distribution $p(\mathbf{z})$. To match the hidden code with an arbitrary distribution can effectively alleviate the overfitting to a certain patient. After the optimization process, the aggregated posterior distribution $q(\mathbf{z})$ of the hidden layer is as follows:

$$q(\mathbf{z}) = \int_{\mathbf{x}} q(\mathbf{z} \mid \mathbf{x}) p_d(\mathbf{x}) d\mathbf{x}, \qquad (7)$$

where $q(\mathbf{z} \mid \mathbf{x})$ is the encoding function of the autoencoder, $p_d(\mathbf{x})$ is the marginal distribution of data. During the training phase, the probabilistic autoencoder is regularized with an adversarial loss function $\mathcal{J}_{adv}$, which is described as:

$$\begin{aligned} \mathcal{J}_{adv} = \mathbb{E}_{\mathbf{z} \sim p(\mathbf{z})}\left[\log D(\mathbf{z}) \right] \\ + \mathbb{E}_{\mathbf{x} \sim p_d}\left[\log(1 - D(G(\mathbf{x}))) \right]. \end{aligned} \qquad (8)$$

After training, a generative model is defined by imposing the prior $p(\mathbf{z})$ on the data distribution. A one-hot encoding vector $\mathbf{y}$ is used for supervised learning (Kumar et al., 2018; Saito et al., 2020). Then we use the learned domain-invariant features for seizure prediction. An SVM classifier is introduced to analyze the extracted features. The loss function of the classification procedure is denoted by $\mathcal{L}_{cla}$. The objective function of the entire model can be defined as:

$$\min_{G,C} \max_{D} \mathcal{L}_{cla} + \lambda_0 \mathcal{L}_{rec} + \lambda_1 \mathcal{J}_{adv} + \lambda_2 \mathcal{R}_{dis}, \qquad (9)$$

where λ_0, λ_1 and λ_2 represent the trade-off positive parameters, and C is the classifier. Our model is optimized jointly with these modules. In general, the MMD-based regularization term is designed to align the distributions among different patients. The AAE architecture is used to construct the latent feature space. The adversarial module is developed to match the hidden code with a prior distribution. Thus this model can circumvent the overfitting to a certain patient.

4. RESULTS AND DISCUSSION

In this section, comparison experiments are conducted to verify the generalization ability and evaluate the forecast precision. Our model is tested on both intracranial and scalp EEG signals.

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

Source	Target	FT-CNN		PLV		SBP		Wav-CNN		Our model	
		S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)
S.C.	Pt 1	0.64	0.21	0.66	0.20	0.69	0.19	0.70	0.17	0.79	0.16
S.C.	Pt 2	0.63	0.3	0.65	0.28	0.66	0.26	0.67	0.24	0.82	0.12
S.C.	Pt 3	0.58	0.24	0.59	0.23	0.62	0.22	0.64	0.22	0.74	0.20
S.C.	Pt 4	0.64	0.25	0.65	0.24	0.66	0.22	0.69	0.20	0.83	0.16
S.C.	Pt 5	0.56	0.4	0.58	0.39	0.59	0.38	0.60	0.38	0.57	0.30
S.C.	Pt 6	0.64	0.27	0.64	0.26	0.67	0.26	0.69	0.26	0.73	0.18
S.C.*	Pt 8	0.54	0.33	0.55	0.33	0.57	0.32	0.57	0.31	0.68	0.29
S.C.	Pt 9	0.70	0.18	0.72	0.17	0.75	0.15	0.77	0.13	0.77	0.19
S.C.	Pt 10	0.52	0.34	0.53	0.33	0.55	0.32	0.58	0.30	0.81	0.16
S.C.	Pt 11	0.50	0.32	0.5	0.30	0.52	0.29	0.52	0.28	0.68	0.29
S.C.	Pt 12	0.72	0.15	0.74	0.13	0.75	0.12	0.77	0.13	0.82	0.09
S.C.*	Pt 13	0.55	0.27	0.56	0.25	0.59	0.24	0.60	0.23	0.66	0.29
S.C.	Pt 14	0.56	0.46	0.57	0.46	0.58	0.44	0.60	0.43	0.75	0.22
S.C.	Pt 15	0.66	0.17	0.66	0.16	0.69	0.15	0.69	0.13	0.83	0.12
S.C.	Pt 16	0.59	0.33	0.6	0.32	0.63	0.31	0.65	0.30	0.85	0.12
S.C.	Pt 17	0.59	0.34	0.62	0.33	0.63	0.31	0.65	0.30	0.77	0.21
S.C.	Pt 18	0.76	0.14	0.78	0.13	0.80	0.11	0.83	0.12	0.84	0.09
S.C.	Pt 19	0.48	0.29	0.48	0.28	0.48	0.27	0.5	0.26	0.73	0.23
S.C.	Pt 20	0.45	0.33	0.47	0.33	0.48	0.33	0.51	0.32	0.82	0.15
S.C.	Pt 21	0.60	0.28	0.62	0.27	0.62	0.25	0.65	0.24	0.66	0.31
	Avg.	0.59	0.28	0.61	0.27	0.63	0.26	0.64	0.25	0.76	0.19

S.C., simulated clinical samples; S_n, sensitivity; FPR, false prediction rate; Avg., average result. S.C.* uses NO samples of the predictor user.

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

Source	Target	FT-CNN		PLV		SBP		Wav-CNN		Our model	
		S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)
S.C.	Pt 1	0.52	0.33	0.54	0.31	0.55	0.31	0.56	0.31	0.77	0.25
S.C.	Pt 2	0.46	0.37	0.47	0.37	0.48	0.34	0.49	0.32	0.56	0.32
S.C.	Pt 3	0.59	0.30	0.60	0.30	0.63	0.29	0.63	0.28	0.70	0.24
S.C.	Pt 5	0.48	0.39	0.48	0.37	0.49	0.35	0.51	0.34	0.74	0.23
S.C.	Pt 6	0.64	0.3	0.66	0.29	0.68	0.28	0.70	0.28	0.79	0.27
S.C.	Pt 7	0.53	0.21	0.56	0.21	0.56	0.29	0.57	0.26	0.71	0.15
S.C.	Pt 8	0.58	0.26	0.60	0.25	0.61	0.24	0.63	0.23	0.82	0.22
S.C.	Pt 9	0.51	0.34	0.54	0.33	0.55	0.33	0.56	0.32	0.78	0.20
S.C.	Pt 10	0.5	0.31	0.51	0.29	0.53	0.28	0.54	0.26	0.72	0.24
S.C.	Pt 13	0.46	0.21	0.47	0.20	0.50	0.28	0.50	0.27	0.54	0.37
S.C.	Pt 14	0.46	0.38	0.48	0.38	0.49	0.36	0.50	0.34	0.80	0.14
S.C.	Pt 17	0.42	0.37	0.43	0.35	0.44	0.35	0.44	0.35	0.75	0.3
S.C.	Pt 18	0.49	0.29	0.52	0.29	0.53	0.27	0.54	0.25	0.70	0.22
S.C.	Pt 19	0.56	0.28	0.58	0.27	0.60	0.25	0.63	0.23	0.73	0.19
S.C.	Pt 20	0.57	0.24	0.59	0.22	0.60	0.2	0.62	0.28	0.82	0.16
S.C.	Pt 21	0.63	0.25	0.66	0.24	0.67	0.22	0.70	0.20	0.68	0.28
	Avg.	0.51	0.30	0.54	0.29	0.56	0.29	0.57	0.28	0.73	0.24

S.C., simulated clinical samples; S_n, sensitivity; FPR, false prediction rate; Avg., average result.
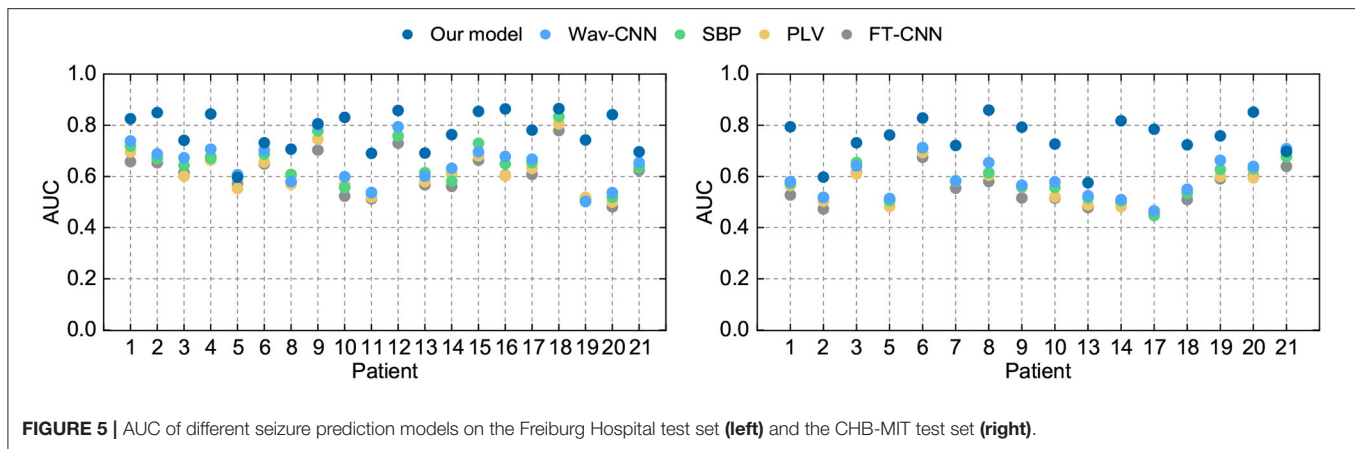
FIGURE 5 | AUC of different seizure prediction models on the Freiburg Hospital test set **(left)** and the CHB-MIT test set **(right)**.

Three commonly-used indicators about model performance are exploited in experiments: sensitivity, false alarm rate per hour (FPR), and area under the receiver operating characteristic curve (AUC). Noted that each EEG fragment represents an event so that the event-based indicators are used for evaluation (Temko et al., 2011).

4.1. Comparison With Conventional Methods

To demonstrate the advantages over conventional methods, we select four seizure prediction researches for comparison: FT-CNN (Truong et al., 2018), phase locking value (PLV) (Cho et al., 2016), spectral band power (SBP) (Ozcan and Erturk, 2019), and Wav-CNN (Khan et al., 2017). All these approaches have obtained good model performance when the training and testing processes are performed on the same subject. But data from previous cases are not used in their training phases. Here we train these models with EEG samples from multiple patients and test them with the "unseen" patient's data. The sensitivity and FPR are provided in **Tables 3**, **4**. The AUC value for each patient is illustrated in **Figure 5**.

The widely-used Freiburg Hospital database is employed in this work to evaluate our model on the intracranial EEG. **Table 3** illustrates that our model outperforms all the other conventional methods in a clear margin. It is reasonable since the prior studies adopt the patient-specific strategy and consider little about the domain adaptability. Conversely, our method exhibits obvious advantages in terms of generalization ability, which achieves a sensitivity of 76% and an FPR of 0.19/h on average. Although these results do not achieve the high accuracy of tests under the patient-specific conditions, such precision can still meet the daily needs of patients with epilepsy because they are similar to the warning frequency in the first-in-man trial (Cook et al., 2013).

Still, the simulated clinical sampling situation is "harsh" for the forecast task. It can be observed that the performances of all these models are not desired. Moreover, on several outliers like Pt 5, 11, and 21, the performance degradation is particularly noticeable. It might be caused by a more complex internal mode in the high-dimensional space. Note that even on these

outliers, the sensitivity of our model is slightly higher than other methods, which demonstrates that our approach achieves better robustness.

As for the scalp EEG recordings, we test these methods using the public CHB-MIT database, produced by the Massachusetts Institute of Technology. As shown in **Table 4**, our algorithm achieves a sensitivity of 73% and an FPR of 0.24/h on average. The results of our model show a significant improvement compared with the conventional methods, which is consistent with the results on intracranial EEG. However, all the model performances drop to a varying degree compared with the precisions on the Freiburg test set. It might be caused by the low spatial resolution of the scalp EEG signals for they are more susceptible to being contaminated by noises (Ramantani et al., 2016; Usman et al., 2019). In other words, intracranial EEG recordings have the higher SNR and the artifacts are typically seen in scalp EEG.

There are also some outliers in the tests on the scalp EEG signal. On Pt 2, 13, and 21, all these models obtain a subpar performance. Larger domain gaps might exist in the sample space of these outliers, which makes the hyperplane difficult to capture. Even on these outliers, the precision of our model is slightly higher than the lower bound of a random binary classifier. It gives us confidence in applying DA techniques to epileptic seizure prediction.

4.2. Comparison With DA Methods

We further compare our model with domain adaptation (DA) methods in the existing literature. However, few applications regarding DA approaches have been reported in the area of epileptic seizure prediction. Thus we have to employ DA methods from other fields. The maximum independence domain adaptation (MIDA) (Yan et al., 2017), model-agnostic learning of semantic features (MASF), conditional deep convolutional generative adversarial networks (C-DCGANs) (Zhang et al., 2021) and subject-invariant domain adaption (SIDA) (Rayatdoost et al., 2021) are introduced to verify the advantage of our model. The sensitivity and FPR are provided in **Tables 5**, **6**. The AUC value for each patient is illustrated in **Figure 6**.

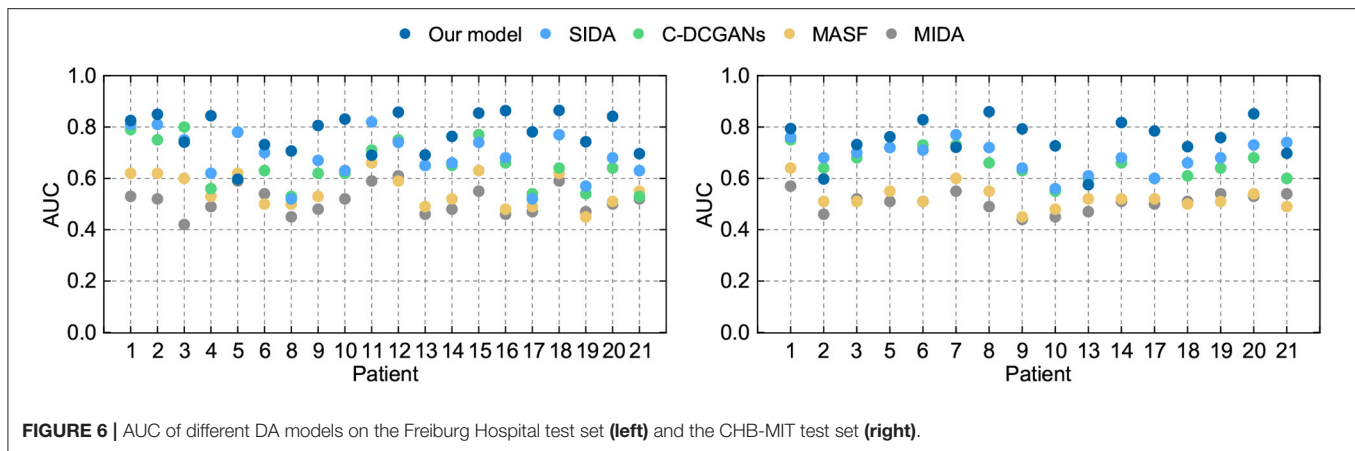**TABLE 5 |** Results compared with DA methods on the Freiburg Hospital database.

Source	Target	MIDA		MASF		C-DCGANs		SIDA		Our model	
		S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)
S.C.	Pt 1	0.57	0.22	0.62	0.21	0.78	0.19	0.83	0.18	0.79	0.16
S.C.	Pt 2	0.56	0.28	0.60	0.26	0.73	0.26	0.82	0.25	0.82	0.12
S.C.	Pt 3	0.52	0.23	0.62	0.22	0.77	0.22	0.78	0.20	0.74	0.2
S.C.	Pt 4	0.49	0.23	0.60	0.23	0.61	0.21	0.62	0.20	0.83	0.16
S.C.	Pt 5	0.57	0.37	0.53	0.35	0.78	0.35	0.79	0.33	0.57	0.30
S.C.	Pt 6	0.53	0.23	0.60	0.23	0.63	0.23	0.70	0.23	0.73	0.18
S.C.*	Pt 8	0.45	0.33	0.51	0.33	0.53	0.33	0.55	0.30	0.68	0.29
S.C.	Pt 9	0.49	0.37	0.51	0.36	0.68	0.26	0.70	0.24	0.77	0.19
S.C.	Pt 10	0.52	0.33	0.54	0.32	0.62	0.32	0.64	0.31	0.81	0.16
S.C.	Pt 11	0.59	0.33	0.57	0.32	0.67	0.30	0.79	0.31	0.68	0.29
S.C.	Pt 12	0.59	0.36	0.63	0.34	0.73	0.24	0.75	0.22	0.82	0.09
S.C.*	Pt 13	0.45	0.29	0.56	0.29	0.69	0.27	0.69	0.26	0.66	0.29
S.C.	Pt 14	0.45	0.46	0.52	0.45	0.65	0.44	0.60	0.44	0.75	0.22
S.C.	Pt 15	0.56	0.16	0.67	0.16	0.52	0.36	0.74	0.16	0.83	0.12
S.C.	Pt 16	0.44	0.35	0.48	0.33	0.76	0.33	0.64	0.31	0.85	0.12
S.C.	Pt 17	0.44	0.36	0.47	0.35	0.53	0.32	0.52	0.32	0.77	0.21
S.C.	Pt 18	0.58	0.36	0.61	0.35	0.77	0.22	0.77	0.21	0.84	0.09
S.C.	Pt 19	0.45	0.29	0.47	0.28	0.53	0.27	0.58	0.26	0.73	0.23
S.C.	Pt 20	0.47	0.34	0.53	0.34	0.62	0.31	0.66	0.31	0.82	0.15
S.C.	Pt 21	0.52	0.30	0.54	0.29	0.52	0.27	0.66	0.26	0.66	0.31
	Avg.	0.51	0.31	0.56	0.30	0.66	0.29	0.69	0.27	0.76	0.19

S.C., simulated clinical samples; S_n, sensitivity; FPR, false prediction rate; Avg., average result. S.C.* uses NO samples of the predictor user.

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

Source	Target	MIDA		MASF		C-DCGANs		SIDA		Our model	
		S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)	S_n	FPR (/h)
S.C.	Pt 1	0.55	0.34	0.61	0.33	0.74	0.30	0.74	0.28	0.77	0.25
S.C.	Pt 2	0.43	0.38	0.49	0.38	0.66	0.37	0.64	0.35	0.56	0.32
S.C.	Pt 3	0.51	0.28	0.50	0.27	0.65	0.25	0.67	0.24	0.70	0.24
S.C.	Pt 5	0.48	0.42	0.51	0.40	0.69	0.37	0.69	0.36	0.74	0.23
S.C.	Pt 6	0.46	0.29	0.52	0.27	0.72	0.27	0.72	0.25	0.79	0.27
S.C.	Pt 7	0.54	0.25	0.56	0.24	0.73	0.21	0.73	0.19	0.71	0.15
S.C.	Pt 8	0.48	0.27	0.60	0.26	0.67	0.25	0.66	0.24	0.82	0.22
S.C.	Pt 9	0.46	0.31	0.51	0.29	0.58	0.27	0.57	0.25	0.78	0.20
S.C.	Pt 10	0.45	0.28	0.46	0.27	0.52	0.25	0.52	0.24	0.72	0.24
S.C.	Pt 13	0.48	0.21	0.51	0.39	0.61	0.38	0.62	0.26	0.54	0.37
S.C.	Pt 14	0.47	0.39	0.48	0.39	0.64	0.36	0.65	0.35	0.80	0.14
S.C.	Pt 17	0.49	0.38	0.50	0.37	0.61	0.37	0.59	0.35	0.75	0.30
S.C.	Pt 18	0.50	0.30	0.49	0.30	0.57	0.28	0.61	0.28	0.70	0.22
S.C.	Pt 19	0.51	0.39	0.51	0.36	0.62	0.34	0.62	0.24	0.73	0.19
S.C.	Pt 20	0.53	0.25	0.55	0.23	0.70	0.23	0.66	0.21	0.82	0.16
S.C.	Pt 21	0.50	0.27	0.52	0.26	0.63	0.24	0.68	0.22	0.68	0.28
	Avg.	0.49	0.31	0.52	0.31	0.65	0.30	0.65	0.27	0.73	0.24

S.C., simulated clinical samples.

FIGURE 6 | AUC of different DA models on the Freiburg Hospital test set **(left)** and the CHB-MIT test set **(right)**.

For the intracranial EEG samples, we still exploit the Freiburg Hospital database as the test set. Evidently, compared with other DA methods, our model achieves the best performance with a sensitivity of 76% and an FPR of 0.19/h on average. Then the SIDA method exhibits a slight advantage over other methods. Comparing this model to SIDA, we see a benefit of approximately 7% is obtained. Comparing our method to C-DCGANs, we remark that a further 3% benefit is obtained. As for the MASF, a benefit of 20% is observed, for a total of 25% margin over MIDA.

In terms of the scalp EEG data, the open CHB-MIT dataset is applied in the experiment. The order of precision of these DA algorithms is consistent with the performance of the intracranial EEG data. Our model has obvious advantages over other approaches with a sensitivity of 73% and an FPR of 0.24/h on average. The high model performance of our model on both intracranial and scalp EEG testifies to the application potential on seizure forecast.

Evidently, SIDA achieves the best performance except for the proposed method. The reason for this superiority might credit the combination of CNN and generative adversarial network (GAN). SIDA is the deep neural network from the field of emotion recognition. Raw EEG data are converted to spectrum in EEG. By minimizing loss of emotion recognition and subject confusion, SIDA extracts the invariant features among different domains. We conjecture that the architecture of GAN in SIDA might have advantages in this task, which needs to be further proved.

The C-DCGANs makes a relatively larger contribution compared with other modules. An adversarial learning-based structure is also employed by C-DCGANs. It also uses the data augmentation technique, which generates EEG recordings artificially. By increasing the data diversity, C-DCGANs hopes to improve the domain shift robustness. Nevertheless, the man-made data may involve more artifacts (Fahimi et al., 2020) that contaminate EEG samples. Besides, C-DCGANs is a variation of deep learning-based frameworks. As such, it has uncertainties associated with DNN, in particular a lack of formal convergence guarantees.

Not surprisingly, the results of MASF and MIDA are not satisfactory. The core idea of MASF is to establish a model-agnostic learning paradigm using semantic features and gradient-based meta-learning. However, the discriminant hyperplane in the high-dimensional space might be too complex to be described with semantic features. MIDA reduces differences between distributions of domains by learning a subspace with background information. It is obvious that the background-specific features are not valid characteristics.

Based on the aforementioned observations, we conjecture that adversarial learning-based techniques are relatively superior for alleviating individual variability, since all the adversarial learning-based models achieve a decent model performance and generalization ability for seizure prediction. Experiments on both intracranial and scalp EEG datasets suggest that adversarial structure has potential in developing a generic seizure forecast model.

4.3. Impact on Different Components

In this section, we conduct experiments to understand the impact of different modules of the proposed model on the final forecast results. To calculate the contribution of each component quantitatively, we adjust the corresponding trade-off positive parameters and observe the variation tendency. The experiment results are listed in **Tables 7**, **8**. Here we discuss three components in this model: the reconstruction module with a loss $\mathcal{L}_{rec}$, the adversarial module with a loss $\mathcal{J}_{adv}$, and the inter-domain distance regularization term $\mathcal{R}_{dis}$.

As shown in **Tables 7**, **8**, we observe that removing the inter-domain distance regularization item, the adversarial subnetwork, or the classification component causes performance drop on both intracranial and scalp EEG databases. Such results indicate that these modules can effectively improve the model performance: (1) AAE is suitable for epileptic EEG signal processing, and the embedding space made by AAE is meaningful. (2) MMD is an appropriate distance measure to minimize the domain gaps in the seizure forecast task. (3) The reconstruction procedure can force the model to learn the significant features from the latent high-dimensional space.

**TABLE 7 |** Comparison results on the Freiburg Hospital database using various components.

Method	S_n	FPR (/h)	Acc	AUC
No $\mathcal{R}_{dis}$	0.60 ± 0.03	0.35 ± 0.03	0.63 ± 0.04	0.63 ± 0.03
No $\mathcal{J}_{adv}$	0.67 ± 0.03	0.31 ± 0.03	0.68 ± 0.03	0.69 ± 0.03
No $\mathcal{L}_{rec}$	0.71 ± 0.03	0.27 ± 0.03	0.72 ± 0.03	0.72 ± 0.03
Our model	0.76 ± 0.03	0.19 ± 0.03	0.78 ± 0.03	0.78 ± 0.03

S_n, sensitivity; FPR, false positive rate; Acc, accuracy.

TABLE 8 | Comparison results on the CHB-MIT database using various components.

Method	S_n	FPR (/h)	Acc	AUC
No $\mathcal{R}_{dis}$	0.57 ± 0.03	0.36 ± 0.03	0.59 ± 0.03	0.59 ± 0.03
No $\mathcal{J}_{adv}$	0.64 ± 0.04	0.33 ± 0.03	0.66 ± 0.04	0.66 ± 0.03
No $\mathcal{L}_{rec}$	0.68 ± 0.03	0.29 ± 0.03	0.69 ± 0.03	0.70 ± 0.03
Our model	0.73 ± 0.03	0.24 ± 0.03	0.75 ± 0.03	0.75 ± 0.03

S_n, sensitivity; FPR, false positive rate; Acc, accuracy.
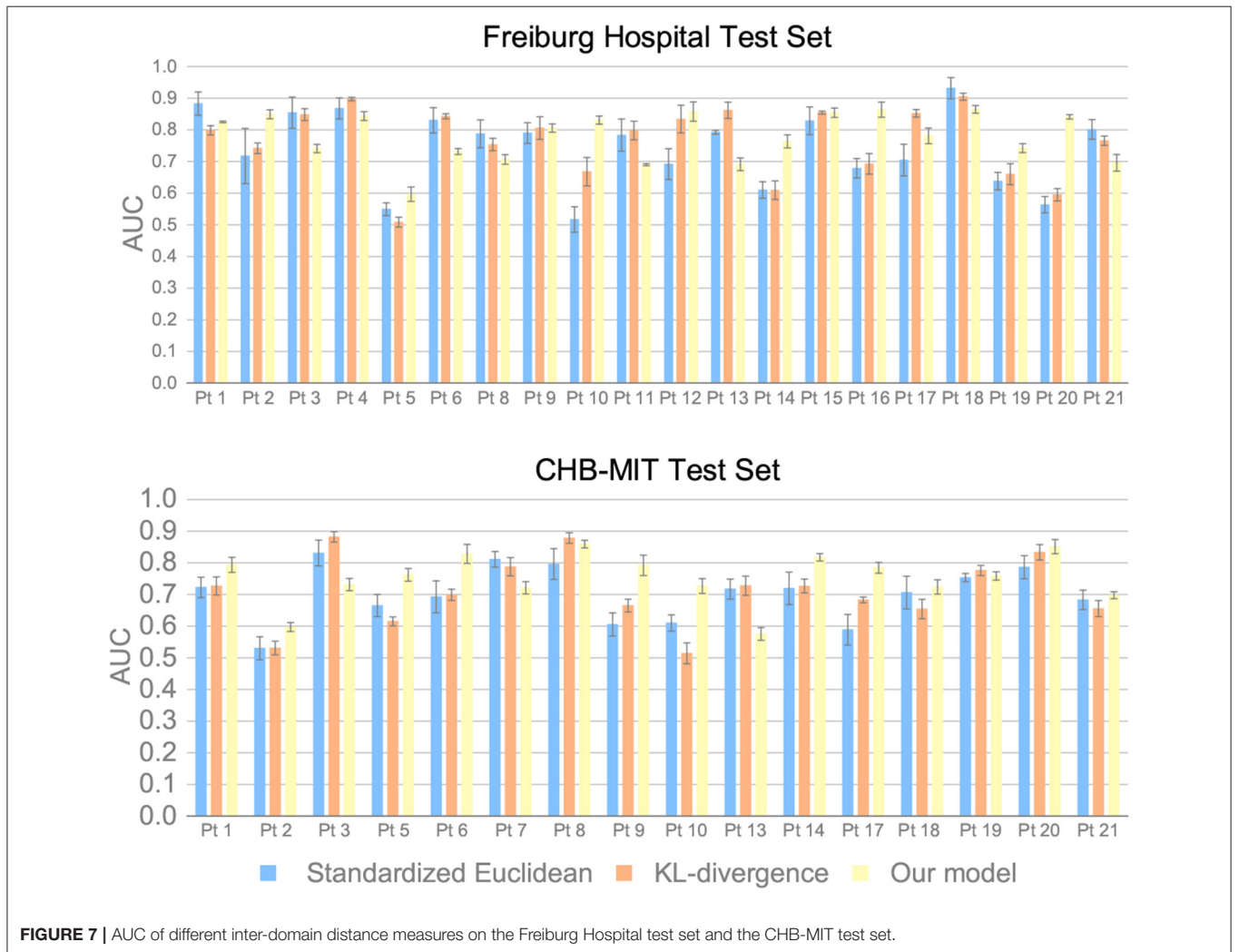
FIGURE 7 | AUC of different inter-domain distance measures on the Freiburg Hospital test set and the CHB-MIT test set.

By adjusting these trade-off items, a set of hyperparameters that are suitable for seizure prediction can be obtained. For the intracranial EEG data, the most appropriate trade-off parameters are set as $\lambda_0 = 1.05, \lambda_1 = 1.2e2, \lambda_2 = 0.7$. For the scalp EEG data, the most appropriate parameter are set as $\lambda_0 = 1, \lambda_1 = 1.1e2, \lambda_2 = 0.6$.

We also discuss the superiority of the MMD measure over other distance metrics. The standardized Euclidean distance and the KL-divergence are used for comparison. Experimental results are provided in **Figure 7**. The results suggest that the precision can increase by 3% for intracranial EEG and 4% for scalp EEG by applying the MMD measure, which demonstrates the advantage of MMD measure on the seizure prediction task.

5. CONCLUSION

This work proposes a generic seizure predictor to alleviate the impact of individual variability. By combining STFT with MMD-AAE, our model reduces the effects of epileptic domain variance and improves the generalization ability. Besides, a simulated clinical sampling scenario is used during training and testing periods, which is the first attempt to adopt this assessing strategy. Compared with the patient-specific strategy from previous researches, such a test approach is relatively challenging. Nonetheless, our method achieves high domain shift robustness and precision, which demonstrates its feasibility of real-world applications.

By analyzing the comparison results of DA methods, a conjecture about the effectiveness of adversarial learning in epileptic seizure prediction is obtained. The underlying causes of this phenomenon remain unclear because there is no definitive explanation of the dynamics of epilepsy in the existing literature. The search for more powerful DA algorithms and the underlying reasons will be considered as part of our future research.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/supplementary material.

ETHICS STATEMENT

Ethical review and approval were not required for the current study in accordance with the local legislation and institutional requirements. 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, LY, and HW designed the experiments. PP and HW analyzed the experimental results. All authors carried out the experiments, wrote the manuscript, contributed to the article, and approved the submitted version.

FUNDING

REFERENCES

Al Ghayab, H. R., Li, Y., Siuly, S., and Abdulla, S. (2019). A feature extraction technique based on tunable Q-factor wavelet transform for brain signal classification. *J. Neurosci. Methods* 312, 43–52. doi: 10.1016/j.jneumeth.2018.11.014

Banerjee, P. N., Filippi, D., and Hauser, W. A. (2009). The descriptive epidemiology of epilepsy–a review. *Epilepsy Res.* 85, 31–45. doi: 10.1016/j.eplepsyres.2009.03.003

Brinkmann, B. H., Wagenaar, J., Abbot, D., Adkins, P., Bosshard, S. C., Chen, M., et al. (2016). Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. *Brain* 139, 1713–1722. doi: 10.1093/brain/aww045

Cho, D., Min, B., Kim, J., and Lee, B. (2016). EEG-based prediction of epileptic seizures using phase synchronization elicited from noise-assisted multivariate empirical mode decomposition. *IEEE Trans. Neural Syst. Rehabil. Eng.* 25, 1309–1318. doi: 10.1109/TNSRE.2016.2618937

Cook, M. J., O'Brien, T. J., Berkovic, S. F., Murphy, M., Morokoff, A., Fabinyi, G., et al. (2013). 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.* 12, 563–571. doi: 10.1016/S1474-4422(13)70075-9

Cordes, D., Kaleem, M. F., Yang, Z., Zhuang, X., Curran, T., Sreenivasan, K. R., et al. (2021). Energy-period profiles of brain networks in group fMRI resting-state data: a comparison of empirical mode decomposition with the short-time Fourier transform and the discrete wavelet transform. *Front. Neurosci.* 15:663403. doi: 10.3389/fnins.2021.663403

Daoud, H., and Bayoumi, M. A. (2019). Efficient epileptic seizure prediction based on deep learning. *IEEE Trans. Biomed. Circ. Syst.* 13, 804–813. doi: 10.1109/TBCAS.2019.2929053

Direito, B., Teixeira, C. A., Sales, F., Castelo-Branco, M., and Dourado, A. (2017). A realistic seizure prediction study based on multiclass SVM. *Int. J. Neural Syst.* 27:1750006. doi: 10.1142/S012906571750006X

Elger, C. E., and Hoppe, C. (2018). Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection. *Lancet Neurol.* 17, 279–288. doi: 10.1016/S1474-4422(18)30038-3

Fahimi, F., Dosen, S., Ang, K. K., Mrachacz-Kersting, N., and Guan, C. (2020). Generative adversarial networks-based data augmentation for brain-computer interface. *IEEE Trans. Neural Netw. Learn. Syst.* doi: 10.1109/TNNLS.2020.3016666

Fei, K., Wang, W., Yang, Q., and Tang, S. (2017). Chaos feature study in fractional Fourier domain for preictal prediction of epileptic seizure. *Neurocomputing* 249, 290–298. doi: 10.1016/j.neucom.2017.04.019

Fisher, R. S., Boas, W. V. E., Blume, W., Elger, C., Genton, P., Lee, P., et al. (2005). Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* 46, 470–472. doi: 10.1111/j.0013-9580.2005.66104.x

Goldberger, A. L., Amaral, L. A., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101, e215–e220. doi: 10.1161/01.CIR.101.23.e215

Guo, L., Wang, Z., Cabrerizo, M., and Adjouadi, M. (2017). A cross-correlated delay shift supervised learning method for spiking neurons with application to interictal spike detection in epilepsy. *Int. J. Neural Syst.* 27:1750002. doi: 10.1142/S0129065717500022

Jia, X., Zhao, M., Di, Y., Yang, Q., and Lee, J. (2017). Assessment of data suitability for machine prognosis using maximum mean discrepancy. *IEEE Trans. Indus. Electron.* 65, 5872–5881. doi: 10.1109/TIE.2017.2777383

Jirsa, V. K., Proix, T., Perdikis, D., Woodman, M. M., Wang, H., Gonzalez-Martinez, J., et al. (2017). The virtual epileptic patient: individualized whole-brain models of epilepsy spread. *Neuroimage* 145, 377–388. doi: 10.1016/j.neuroimage.2016.04.049

Karoly, P. J., Freestone, D. R., Boston, R., Grayden, D. B., Himes, D., Leyde, K., et al. (2016). Interictal spikes and epileptic seizures: their relationship and underlying rhythmicity. *Brain* 139, 1066–1078. doi: 10.1093/brain/aww019

Khan, H., Marcuse, L., Fields, M., Swann, K., and Yener, B. (2017). Focal onset seizure prediction using convolutional networks. *IEEE Trans. Biomed. Eng.* 65, 2109–2118. doi: 10.1109/TBME.2017.2785401

Kuhlmann, L., Lehnertz, K., Richardson, M. P., Schelter, B., and Zaveri, H. P. (2018). Seizure prediction-ready for a new era. *Nat. Rev. Neurol.* 14, 618–630. doi: 10.1038/s41582-018-0055-2

Kumar, A., Sattigeri, P., Wadhawan, K., Karlinsky, L., Feris, R., Freeman, W. T., et al. (2018). Co regularized alignment for unsupervised domain adaptation. *arXiv [Preprint]*.arXiv:1811.05443

Kwan, P., Schachter, S. C., and Brodie, M. J. (2011). Drug-resistant epilepsy. *N. Engl. J. Med.* 365, 919–926. doi: 10.1056/NEJMra1004418

Li, H., Pan, S. J., Wang, S., and Kot, A. C. (2018). "Domain generalization with adversarial feature learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 5400–5409. doi: 10.1109/CVPR.2018.00566

Li, S., Zhou, W., Yuan, Q., and Liu, Y. (2013). Seizure prediction using spike rate of intracranial EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* 21, 880–886. doi: 10.1109/TNSRE.2013.2282153

Li, Y., Yu, Z., Chen, Y., Yang, C., Li, B., Li, X. A., et al. (2020). Automatic seizure detection using fully convolutional nested LSTM. *Int. J. Neural Syst.* 30:2050019. doi: 10.1142/S0129065720500197

Lin, L.-C., Ouyang, C.-S., Chiang, C.-T., Yang, R.-C., Wu, R.-C., and Wu, H.-C. (2014). Early prediction of medication refractoriness in children with idiopathic epilepsy based on scalp EEG analysis. *Int. J. Neural Syst.* 24:1450023. doi: 10.1142/S0129065714500233

Lin, L.-C., Ouyang, C.-S., Wu, R.-C., Yang, R.-C., and Chiang, C.-T. (2020). Alternative diagnosis of epilepsy in children without epileptiform discharges using deep convolutional neural networks. *Int. J. Neural Syst.* 30:1850060. doi: 10.1142/S0129065718500600

Liu, G., Zhou, W., and Geng, M. (2020). Automatic seizure detection based on s-transform and deep convolutional neural network. *Int. J. Neural Syst.* 30:1950024. doi: 10.1142/S0129065719500242

Makhzani, A., Shlens, J., Jaitly, N., Goodfellow, I., and Frey, B. (2015). Adversarial autoen coders. *arXiv [Preprint]*. arXiv:1511.05644

Mirowski, P., Madhavan, D., LeCun, Y., and Kuzniecky, R. (2009). Classification of patterns of EEG synchronization for seizure prediction. *Clin. Neurophysiol.* 120, 1927–1940. doi: 10.1016/j.clinph.2009.09.002

Moctezuma, L. A., and Molinas, M. (2020). EEG channel-selection method for epileptic-seizure classification based on multi-objective optimization. *Front. Neurosci.* 14:593. doi: 10.3389/fnins.2020.00593

Muralidharan, A., Chae, J., and Taylor, D. (2011). Extracting attempted hand movements from EEGs in people with complete hand paralysis following stroke. *Front. Neurosci.* 5:39. doi: 10.3389/fnins.2011.00039

Myers, M. H., Padmanabha, A., Hossain, G., de Jongh Curry, A. L., and Blaha, C. D. (2016). Seizure prediction and detection via phase and amplitude lock values. *Front. Hum. Neurosci.* 10:80. doi: 10.3389/fnhum.2016.00080

Özcan, A. R., and Ertürk, S. (2017) "Epileptic seizure prediction with recurrent convolutional neural networks," in *Signal Processing and Communications Applications Conference* (Antalya), 1–4. doi: 10.1109/SIU.2017.7960594

Ozcan, A. R., and Erturk, S. (2019). Seizure prediction in scalp EEG using 3D convolutional neural networks with an image-based approach. *IEEE Trans. Neural Syst. Rehabil. Eng.* 27, 2284–2293. doi: 10.1109/TNSRE.2019.2943707

Peng, L.-H., Yin, J., Zhou, L., Liu, M.-X., and Zhao, Y. (2018). Human microbe-disease association prediction based on adaptive boosting. *Front. Microbiol.* 9:2440. doi: 10.3389/fmicb.2018.02440

Peng, P., Wei, H., Xie, L., and Song, Y. (2020). "Epileptic seizure prediction in scalp EEG using an improved HIVE-COTE model," in *Chinese Control Conference* (Shenyang: IEEE), 6450–6457. doi: 10.23919/CCC50068.2020.9188930

Peng, P., Xie, L., and Wei, H. (2021). A deep Fourier neural network for seizure prediction using convolutional neural network and ratios of spectral power. *Int. J. Neural Syst.* 31:2150022. doi: 10.1142/S0129065721500222

Ramantani, G., Maillard, L., and Koessler, L. (2016). Correlation of invasive EEG and scalp EEG. *Seizure* 41, 196–200. doi: 10.1016/j.seizure.2016.05.018

Rayatdoost, S., Yin, Y., Rudrauf, D., and Soleymani, M. (2021). "Subject-invariant EEG representation learning for emotion recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing* (Barcelona), 3955–3959. doi: 10.1109/ICASSP39728.2021.9414496

Rojas, G. M., Alvarez, C., Montoya, C. E., de la Iglesia-Vayá, M., Cisternas, J. E., and Gálvez, M. (2018). Study of resting-state functional connectivity networks using EEG electrodes position as seed. *Front. Neurosci.* 12:235. doi: 10.3389/fnins.2018.00235

Saito, K., Kim, D., Sclaroff, S., and Saenko, K. (2020). Universal Domai n Adaptation through Self Supervision. *arXiv [Preprint]*.arXiv:2002.07953

Smola, A., Gretton, A., Song, L., and Schölkopf, B. (2007). "A Hilbert space embedding for distributions," in *International Conference on Algorithmic Learning Theory* (Sendai: Springer), 13–31. doi: 10.1007/978-3-540-75225-7_5

Sun, C., Cui, H., Zhou, W., Nie, W., Wang, X., and Yuan, Q. (2019). Epileptic seizure detection with EEG textural features and imbalanced classification based on easyensemble learning. *Int. J. Neural Syst.* 29:1950021. doi: 10.1142/S0129065719500217

Temko, A., Thomas, E., Marnane, W., Lightbody, G., and Boylan, G. (2011). Performance assessment for EEG-based neonatal seizure detectors. *Clin. Neurophysiol.* 122, 474–482. doi: 10.1016/j.clinph.2010.06.035

Truong, N. D., Nguyen, A. D., Kuhlmann, L., Bonyadi, M. R., Yang, J., Ippolito, S., et al. (2018). Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram. *Neural Netw.* 105, 104–111. doi: 10.1016/j.neunet.2018.04.018

Tsiouris, K. M., Pezoulas, V. C., Zervakis, M., Konitsiotis, S., Koutsouris, D. D., and Fotiadis, D. I. (2018). A long short-term memory deep learning network for the prediction of epileptic seizures using EEG signals. *Comput. Biol. Med.* 99, 24–37. doi: 10.1016/j.compbiomed.2018.05.019

Usman, S. M., Khalid, S., Akhtar, R., Bortolotto, Z., Bashir, Z., and Qiu, H. (2019). Using scalp EEG and intracranial EEG signals for predicting epileptic seizures: review of available methodologies. *Seizure* 71, 258–269. doi: 10.1016/j.seizure.2019.08.006

Vahabi, Z., Amirfattahi, R., Shayegh, F., and Ghassemi, F. (2015). Online epileptic seizure prediction using wavelet-based bi-phase correlation of electrical signals tomography. *Int. J. Neural Syst.* 25:1550028. doi: 10.1142/S0129065715500288

Yair, O., Ben-Chen, M., and Talmon, R. (2019). Parallel transport on the cone manifold of SPD matrices for domain adaptation. *IEEE Trans. Signal Process.* 67, 1797–1811. doi: 10.1109/TSP.2019.2894801

Yan, K., Kou, L., and Zhang, D. (2017). Learning domain-invariant subspace using domain features and independence maximization. *IEEE Trans. Cybernet.* 48, 288–299. doi: 10.1109/TCYB.2016.2633306

Zandi, A. S., Tafreshi, R., Javidan, M., and Dumont, G. A. (2013). Predicting epileptic seizures in scalp EEG based on a variational Bayesian Gaussian mixture model of zero-crossing intervals. *IEEE Trans. Biomed. Eng.* 60, 1401–1413. doi: 10.1109/TBME.2012.2237399

Zhang, T., Chen, W., and Li, M. (2018). Fuzzy distribution entropy and its application in automated seizure detection technique. *Biomed. Signal Process. Control* 39, 360–377. doi: 10.1016/j.bspc.2017.08.013

Zhang, W., Yan, F., Han, F., He, R., Li, E., Wu, Z., et al. (2021). Auto recognition of solar radio bursts using the C-DCGAN method. *Front. Phys.* 9:646556. doi: 10.3389/fphy.2021.646556

Zhang, Y., Guo, Y., Yang, P., Chen, W., and Lo, B. (2019). Epilepsy seizure prediction on EEG using common spatial pattern and convolutional neural network. *IEEE J. Biomed. Health Inform.* 24, 465–474. doi: 10.1109/JBHI.2019.2933046

Zhang, Y., Zhou, Z., Bai, H., Liu, W., and Wang, L. (2020). Seizure classification from EEG signals using an online selective transfer TSK fuzzy classifier with joint distribution adaption and manifold regularization. *Front. Neurosc.* 14:496. doi: 10.3389/fnins.2020.00496

Zhao, J., Lin, C.-M., and Chao, F. (2019). Wavelet fuzzy brain emotional learning control system design for MIMO uncertain nonlinear systems. *Front. Neurosci.* 12:918. doi: 10.3389/fnins.2018.00918

Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., et al. (2018). Epileptic seizure detection based on EEG signals and CNN. *Front. Neuroinform.* 12:95. doi: 10.3389/fninf.2018.00095