Multichannel Synthetic Preictal EEG Signals to Enhance the Prediction of Epileptic Seizures

Yankun Xu ⊗, Jie Yang ⊗, and Mohamad Sawan ⊗, Fellow, IEEE

Abstract—Epilepsy is one of the most common chronic neurological disorders affecting 1% of people worldwide. Deep learning (DL) algorithms-based electroencephalograph (EEG) analysis provides the possibility for accurate epileptic seizure (ES) prediction, thereby benefiting patients suffering from epilepsy. To identify the preictal region that precedes the onset of seizure, a large number of annotated EEG signals are required to train DL algorithms. However, the scarcity of seizure onsets leads to significant insufficiency of data for training the DL algorithms. To overcome this data insufficiency, in this paper, we propose a preictal artificial signal synthesis algorithm based on a generative adversarial network to generate synthetic multichannel EEG preictal samples. A high-quality single-channel architecture, determined by visual and statistical evaluations, is used to train the generators of multichannel samples. The effectiveness of the synthetic samples is evaluated by comparing the ES prediction performances without and with synthetic preictal sample augmentation. The leave-one-seizure-out cross validation ES prediction accuracy and corresponding area under the receiver operating characteristic curve improvement from 73.0% and 0.676 to 78.0% and 0.704 by 10× synthetic sample augmentation, respectively. The obtained results indicate that synthetic preictal samples are effective for enhancing ES prediction performance.

Index Terms—Epilepsy, preictal, seizure prediction, machine learning, deep learning, generative adversarial network, data augmentation, synthetic EEG.

I. INTRODUCTION

Epilepsy is one of the most common chronic neurological disorders characterized by recurrent epileptic seizures (ESs) triggered by abnormal discharge of neurons [1], [2]. Approximately 1% people worldwide suffer from unprovoked ESs, and the primary treatment for epileptic patients is long-term drug therapy. Moreover, approximately one-third of patients are resistant to anti-epileptic medication [3], [4]. Epileptic patients can benefit from an accurate prediction system that enables early warning minutes before onset. This allows for precautional intervention or therapies, thereby alleviating the psychological burden on patients and the demand for healthcare resources.

Electroencephalography (EEG) is a prevalent neuroimaging approach widely used for epilepsy diagnosis and analysis. Recent research activities in the field of ES prediction are mainly based on either scalp EEG or invasive EEG signals [5]–[8]. Fig. 1 shows an annotated example of a single-channel EEG recording during a seizure occurrence. The recording of epileptic patients can be categorized into several periods: ictal (seizure occurrence), preictal (before seizure onset), postictal (shortly after the seizure), seizure prediction horizon (SPH, a short period before seizure onset in which precautional intervention is required) and interictal (between seizures excluding all other periods).

Over the past decades, various machine learning (ML) and deep learning (DL) based algorithms have been proposed to improve the performance of ES prediction [9]–[13]. To train a reliable classifier, either ML- or DL-based algorithms requires a large number of annotated data. However, the unavailability of sufficient annotated preictal samples is a major challenge for ES prediction studies. First, the acquisition of sufficient annotated seizure data is costly and time-consuming. Second, the preictal period is only observed if a seizure occurs. Third, only one or a few seizures usually occur per day for most patients, in extreme cases that patients with intractable epilepsy can suffer from multiple seizures per day. Therefore the duration of interictal periods is significantly longer than that of preictal periods. The natural difference between interictal and preictal durations leads to the numbers of the two types of samples for the training set to be significantly imbalanced.

Recently, several researchers have applied generative adversarial network (GAN) technique to ES studies [14]–[16].
However, they focus on the synthesis of either short-term (<5 s) single-channel raw EEG ictal signals or specific features extracted from raw EEG signals. These pioneering efforts cannot be used to synthesize long-term (>5 s) multichannel signals. This is primarily because this type of signals easily lead to mode collapse of GAN training. Additionally, the extracted features of preictal and ictal signals are significantly different even when they appear similar. Furthermore, the lack of an evaluation method for synthetic time-series data is always a challenge for GAN studies [17].

In this paper, we propose a data generation method based on GAN to deliver synthetic long-duration multichannel time-series EEG preictal samples to address the problem of scarcity of preictal data and imbalanced training sets. The main contributions of this study are as follows:

1) Generative model architectures based on convolutional neural network (CNN) and recurrent neural network (RNN) have been proposed to generate synthetic single-channel EEG preictal samples. Dedicated statistical evaluation metrics are also proposed to assess the quality of single-channel samples in order to choose the outperformed generative model.

2) A high-quality generative model is chosen to generate synthetic multichannel preictal samples. The effectiveness of the model is confirmed by comparing ES prediction performance without and with data augmentation (DA) techniques. The results reveal that synthetic preictal sample augmentation can improve ES prediction performance.

The remaining content of this paper is organized as follows. Section II describes previous studies applying the GAN technique to the field of epilepsy study. Section III elaborates on the methodology of our work including generative model training, performance evaluation and implementation of ES prediction experiments without and with DA. The performance and comparison are presented in Section IV. Sections V and VI include a discussion of our results and the main conclusions, respectively.

II. RELATED WORK

A. Literature Review

Since proposed in 2014 [18], the GAN technique has attracted significant attention from research community owing to its success in the fields of computer vision and natural language processing. Over the past three years, numerous studies that apply the GANs to the time-series biomedical signal research are emerging. In this section, we present a concise literature review that focuses on the application of GANs on EEG signal synthesis.

Hartmann et al. [14] described an EEG-GAN framework based on deep convolution GAN (DCGAN) to synthesize naturalistic single-channel hand movement EEG signals. They compared several upsampling methods within the CNN-based generator and concluded that the cubic interpolation method outperforms other interpolation methods. Fahimi et al. [19] introduced a framework based on DCGAN to generate synthetic EEG signals to improve the accuracy of motor task experiment with diverted and focused attention conditions by 7.32% and 5.45%, respectively. Aznan et al. [20] utilized DCGAN, Wasserstein GAN (WGAN), and a variational autoencoder to create synthetic EEG data for improving steady-state visual evoked potential classification tasks. They obtained slightly better performance on single-subject task, and 35% improved accuracy on the across-subject generalization task. Luo et al. [21] and Jiao et al. [22] proposed a conditional WGAN to augment EEG signals to improve the classification performance in emotion recognition and driver sleepiness detection applications. In addition to CNN-based GAN, researchers [23], [24] utilized RNN-based GAN to augment different bio-signals. Abdelfattah et al. proposed a novel GAN model and demonstrated their model can enhance classification performance on small EEG dataset. Haradal et al. confirmed their RNN-based GAN model is capable of generating synthetic bio-signals using the electrocardiogram and EEG datasets. Authors in [25], they used bidirectional RNN neurons to generate artificial motor imagery EEG signals, and Short-term Fourier Transform (STFT) and Welch’s power spectral density to evaluate the time-frequency characteristics of artificial signals. The results indicated that GAN can capture important information of beta band. Hwang et al. [22] proposed a scheme for zero-shot EEG signal classification using RNN-based GAN. Their EEG encoder achieves an accuracy of 95.89% and they reached an accuracy of 39.65% for classification task with ten untrained EEG classes. In [22], authors applied RNN-based GAN model to drive sleepiness detection study based on EEG and electrooculography signals. They addressed the change in alpha waves and differentiate the two alpha-related phenomena. In addition, they adopted conditional WGAN to augment dataset and boost classifier performance.

There have also been instances of application of generative models to ES studies. Truong et al. [27] proposed a semi-supervised learning approach based on DCGAN to forecast the ESs. They took advantage of features generated from the STFT of EEG signals as real data input to generate synthetic two-dimensional feature samples. Subsequently, a discriminator was used to distinguish real and generated samples. When the training reached equilibrium, the trained discriminator was used as a feature extractor for supervised learning. However, this innovative trial did not outperform the CNN model, which had previously been proposed by the same authors for three different databases [28]. Rasheed et al. [15] proposed a generative model based on DCGAN to synthesize the spectrogram of the STFT of EEG signals for ES prediction, which utilized one-class support vector machine algorithm with radial basis function for data validation. Subsequently, the transfer learning technique was used to efficiently train the ES prediction classification model. They found that synthetic spectrogram sample augmentation increased the area under the receiver operating characteristic (ROC) curve (AUC) and sensitivity on two different databases respectively. Pascual et al. [29] proposed a conditional GAN [30] approach, namely EpilepsyGAN, to synthesize 4 s ictal samples from two selected channels for a seizure detection task. They compared the similarity between real and synthetic...
ictal samples using a spectral cosine metric, and obtained an average 1.3% improved sensitivity for 24 patients in the ES detection task. In [31], authors took advantage of GAN as a strong candidate for data enhancement intended for tackling imbalanced EEG data distribution. After data augmentation by the GAN, they designed one-dimensional CNN to achieve better performance on seizure detection classification task. You et al. [32] performed automatic seizure detection with the trained GAN as an anomaly detector based on behind-the-ear EEG signals. They eventually achieved AUC of 0.939 and sensitivity of 96.3%. In [33], authors presented a WGAN-based model to generate multichannel seizure data, and this approach is demonstrated to reduce the false positive rate by 72.72%, meanwhile slightly increasing detection latency and maintaining sensitivity.

III. METHODOLOGY OF PREICTAL SIGNAL SYNTHESIS

A. GAN Theory

The GAN framework consists of two opposing networks that are trained simultaneously to achieve an equilibrium. The first network, which is a generator, attempts to generate artificial realistic samples based on the latent noise input, and the second is a discriminator, which is used to distinguish between artificial and real samples. This adversarial learning aims to train the generator to capture the hidden structure and distribution of real samples to generate artificial samples with the same distribution. The loss function of GAN is defined as:

$$\min_G \max_D \mathbb{E}_{x \sim P_r}[\log D(x)] + \mathbb{E}_{\tilde{x} \sim P_g}[\log(1 - D(\tilde{x}))]$$  \hspace{1cm} (1)

where $P_r$ is the real data distribution and $P_g$ represents the model distribution defined by

$$\tilde{x} = G(z)$$  \hspace{1cm} (2)

where $z$ is a random vector from the noise distribution.

An improved GAN known as Wasserstein GAN is proposed to overcome the training stability issue of conventional GAN [34], [35]. One drawback of the conventional GAN framework is the training instability issue in which vanishing gradients occur in the generator if the discriminator were trained to optimality during each training iteration. WGAN is intended to minimize the Wasserstein distance (WD) instead of the Jensen-Shannon divergence, which is measured by the original GAN, between real and artificial data distributions. WD, also known as the Earth-Mover distance, is used to calculate the cost of transporting mass from one distribution to another. In the WGAN, discriminator (also called critic according to [34]) and generator are trained to optimize the loss function:

$$\min_G \max_D \mathbb{E}_{x \sim P_r}[D(x)] + \mathbb{E}_{\tilde{x} \sim P_g}[D(\tilde{x})]$$  \hspace{1cm} (3)

where $D$ is the set of 1-Lipschitz functions that aim to introduce Lipschitz continuity to improve the discriminator [35]. For practical employment of WGAN, the weights of the discriminator are clipped to a compact interval $[-c, c]$ to enforce the Lipschitz constraint.

B. Data Description

We employed and evaluated our proposed preictal signal synthesis method based on the CHB-MIT EEG dataset [36]. The CHB-MIT dataset contains scalp EEG signals from 23 cases collected from 22 pediatric patients, and all signals were sampled at 256 Hz. The electrode position system used in this dataset was under the International 10-20 system standard. Fig. 3 depicts the 10-20 system electrode placement [37]. However, there were electrode configuration differences among patients in terms of practical signal acquisition. We only considered patients with the same electrode configuration and no less than three leading seizures, hence 7 subjects with the same 22 electrode bipolar placement were selected for the experiments. The characteristics of selected patient subjects are shown in Table I that contains the number of leading and total seizures, preictal hours and detailed electrode placement settings, where FP, F, T, P, C, and O represent frontopolar, frontal, temporal, parietal, central, and occipital respectively, odd number, even number and Z represent left side, right side and midline of the brain.

| Patient ID | No. of leading seizures / total seizures | Preictal hours | Electrode placement |
|------------|----------------------------------------|----------------|-------------------|
| chb01      | 3/7                                    | 1.5            | FP-P1, P1-T1, T1-P1 |
| chb08      | 3/5                                    | 1.5            | P7-O1, FP1-F3, F3-C3 |
| chb09      | 3/4                                    | 1.5            | C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4, P4-O2 |
| chb10      | 6/7                                    | 9              | FP2-P8, P8-T8, T8-P8 |
| chb11      | 3/3                                    | 1.5            | P8-O2, FZ-CZ, CZ-FZ, P7-T7, T7-PO |
| chb19      | 3/3                                    | 1.5            | FT9-FT10, FT10-T8 |

C. Data Preparation

Conventionally, preprocessing and preparation of the data is required prior to the training phase. SPH is a very short period prior to the seizure onset providing intervention, it should last a few minutes, 5 min is a common choice according to previous studies [16], [28], [38]. Postictal period usually lasts 5-30 min, we expect the postictal signal would not disturb the interictal signal, so that the longest duration 30 min is chosen in this work [39]. The duration of an assumed preictal period varies across seizures and individuals, state-of-the-art studies usually choose the duration of preictal from 15 min to 60 min, and in terms of the CHB-MIT dataset, the most popular choice is 30 min. Therefore, we defined the duration of the SPH, preictal and postictal periods as 5 min, 30 min and 30 min respectively. The preictal and interictal samples are segmented from raw EEG recordings by a time-window without overlapping, the length of the time-window is chosen as 20 s because the seizure prediction algorithm used in this work is based on the work published in [40], and the algorithm of the later work required the training and testing samples segmented in 20 s. A segment of 20 s was used considering much longer continuous EEG signals than 1-5 s, meanwhile we considered the issues of computational cost, so that 20 s is a good option. In addition, 20 s duration provides...
flexibility for various needs of seizure prediction algorithm design because numerous state-of-the-art algorithms adopted the segments with less than 20 s. The length of the single sample should be 5120 (20 s × 256 Hz); however, in practical experiments, we down-sampled the real preictal samples from the raw sampling rate to 100 Hz intended for training convenience [41]. Thus, the length of real preictal sample for GAN training was 2000 (20 s × 100 Hz). Furthermore, we aimed to train the channel-wise generators independently, so that the shape of the real preictal samples was 1 × 2000.

D. Procedure for Preictal EEG Samples Synthesis

Fig. 2 shows the flow of the proposed preictal EEG sample synthesis method. It consists of three stages, i.e., adversarial learning, data evaluation and preictal sample synthesis. In the adversarial learning stage, we consider CNN- and RNN-based generative model architectures for both GAN and WGAN training. Thus, there were four candidates — DCGAN, deep convolution WGAN (DCWGAN), recurrent GAN (RGAN) and recurrent WGAN (RWGAN) for single-channel EEG sample synthesis. In the data evaluation stage, visual inspection and proposed statistical metrics were employed to evaluate the quality of synthetic single-channel samples generated by the four alternatives, and the best performing one was chosen for the latter stage. In the final stage, the high-quality generative model was used to train the channel-wise generators of all channels independently, then identical noise was added to all trained generators to synthesize signal-channel samples for all channels, the synthesized multi-channel preictal EEG samples are generated by concatenating these single-channel samples. Subsequently, the effectiveness was evaluated.

E. Generative Model Architecture

The CNN- and RNN-based generator and discriminator architectures are shown in Figs. 4 and 5, respectively. These figures illustrate the operation and shape of each layer; these architectures are utilized for both GAN and WGAN training. In the CNN-based GAN, the vector from the noise latent space with a shape of 1 × 100 was fed to the generator which comprised one fully connected (FC) layer and five convolution blocks. These blocks were expected to maintain the network deep enough and the number of FC layer nodes was determined by the shape of the subsequently designed 3D tensor with a depth of 256, which was expected to contain a complex combination of patterns. Except for the last convolution block which used pixel-wise convolution to generate the output with required shape, the remaining blocks implemented the same operations in order: convolution with the kernel shape of 1 × 5, and a leaky ReLU active function. For the CNN-based discriminator, we utilized four convolution blocks containing the same operations in order: convolution with the kernel shape of 1 × 5, max pooling with the kernel shape of 1 × 2, and two FC layers with 64 nodes and 1 node, respectively.
In the RNN-based GAN, noise vectors were input to the generator which consisted of bidirectional long short-term memory (LSTM) cells with 20 time steps which were determined with the aim of processing 20 s preictal samples. The output from each cell was fed to a FC layer with shared weights across time steps, and eventually a FC layer was used for generating output with the required shape [42]. The discriminator used an architecture similar to that of the generator, and there were only two differences that the input should be segmented with several time steps for later bidirectional LSTM cells, and a single neuron was added for sigmoid output generation with distinguishing objectives.

We randomly selected one channel for training different GAN architectures. When the training phase was complete, four different trained generators were used to synthesize single-channel preictal samples, and then we upsampled the synthetic preictal samples from the shape of $1 \times 2000$ to the original shape of $1 \times 5120$.

**F. Evaluation of Generative Models**

Evaluation of GAN models usually relies on the evaluation of the performance of artificial outputs generated by generator [43]. In this section, we elaborate on visual inspection and proposed statistical metrics for the evaluation of the quality of synthetic single-channel preictal samples generated by the four generators. The outperformed one is chosen to train the generators of all channels according to the evaluation approaches.

For visual inspection, we directly compared synthetic samples with real samples by vision. Although EEG signals are significantly different from images that can be easily recognized visually, we can roughly recognize the quality of EEG signals from the amplitude, frequency and period characteristics by visual inspection only provides us an intuitive quality of the synthetic EEG samples. Furthermore, reliable statistical evaluation metrics are required.

Numerous statistical evaluation metrics can be used to evaluate the synthetic EEG signal. In this study, we selected three evaluation metrics for our application – frequency domain root mean square error (FDRMSE), Fréchet inception distance (FID) and WD. The root mean square error is widely applied to measure the difference between two signals. However, due to the similarity of real and synthetic samples in the time domain, meaningful information cannot be obtained because significant differences exist among different real samples in the time domain. Thus,

![Convolution neural network-based architecture: (a) and (b) are the generator and discriminator, respectively.](image)

![Recurrent neural network-based architecture: (a) and (b) are the generator and discriminator, respectively.](image)

**Algorithm 1:** Algorithm for Multichannel Preictal Sample Synthesis Using GAN.

```plaintext
Require: c, the clipping parameter, m, the batch size, n_d, the number of iterations of the discriminator per generator iteration.
Initialization: \( \theta_d \), discriminator’s parameters, \( \theta_g \), generator’s parameters, \( c = 0.01, m = 32 \)
1: for channels = 1, 2, ..., 22 do
2: for epochs = 1, 2, ..., 30000 do
3: for n_d = 1, 2, ..., 5 do
4: Sample \( \{ z^{(i)} \}_{i=1}^m \sim p(z) \) a batch from noise prior.
5: Sample \( \{ x^{(i)} \}_{i=1}^m \sim P_r \), a batch from the real single-channel preictal samples.
6: Updating the discriminator by ascending its stochastic gradient: \( \nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m [\log D(x^{(i)}) + \log(1 - D(G(z^{(i)})))] \)
7: Clipping the parameters \( \theta_d \leftarrow \text{clip}(\theta_d, -c, c) \)
8: end for
9: Sample \( \{ z^{(i)} \}_{i=1}^m \sim p(z) \) a batch from noise prior.
10: Updating the generator by ascending its stochastic gradient:
11: \( \nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log(1 - D(G(z^{(i)}))) \)
12: Generate synthetic single-channel preictal samples based on trained generator per channel:
13: \( G_{\text{channels}}(z) \)
14: end for
15: Combination of each synthetic single-channel samples to generate multichannel synthetic preictal samples using identical noise vector \( z: [G_1(z), G_2(z), \ldots, G_{22}(z)]^T \).
```
the frequency domain information attracts our interest, and we compare the real and synthetic samples by calculating the RMSE of the frequency domain information. The FDRMSE is calculated as:

$$FDRMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (F_n - \hat{F}_n)^2}$$

(4)

where $F$ and $\hat{F}$ represent the coefficients of the real and synthetic samples in the frequency domain.

FID is a common method designed to evaluate the quality of generative models and synthetic samples [14], [44]. Similar to the inception score, FID utilizes a pre-trained classifier model to compare the information of real and synthetic samples representation in the embedding layer of the model. The Fréchet distance is used to calculate the distributions of two values in the embedding layer. Compared to the inception score, FID functions better because it is more robust to noise and sensitive to the quality of generative models.

WD was used as the third statistical metric for evaluation. As mentioned in Section II-A, WD is a criterion for adversarial learning and can be used to measure the similarity between any two distributions. In this application, WD can also be regarded as sliced WD because the single-channel sample is one-dimensional.

Thus, we considered the frequency coefficients and distributions of the raw data and embedding layer instead of directly comparing the differences from the time domain. These evaluation metrics were used to select the best generative model, that was utilized to train the generators of all channels. Eventually channel-wise trained generators were used to synthsize multichannel preictal samples for following multichannel preictal sample evaluation. The pseudo-code of proposed multichannel preictal sample synthesis method is presented in Algorithm 1.

G. Evaluation of Multichannel Samples

When the outperformed generative model is chosen, we used this model to train the generators of all channels thereby synthesizing multichannel preictal samples. After generation of a data pool containing a sufficient number of synthetic multichannel preictal samples, the effectiveness was evaluated by comparing ES prediction performance without and with synthetic samples DA. We only augmented real preictal samples with synthetic samples at different ratios, such as 3, 5 and 10 times, and the real interictal samples were kept unchanged. We utilized the classification algorithm proposed earlier by our group (Xu et al. [40]) as a baseline model for subsequent patient-specific ES prediction experiments. Fig. 6 shows the architecture of the model which consists of 3 one-dimensional convolution blocks, 2 two-dimensional convolution blocks and 2 FC layers. This model aims to address raw EEG signal and learn the time-domain representations of EEG signals. We expect that the DA condition can achieve the similar or improved performance compared to the all-real condition, in order to prove the effectiveness of synthetic preictal samples.

Several statistical metrics are considered to evaluate the performance of ES prediction in this study, such as classification accuracy, ROC curve, and AUC.

IV. RESULTS

A. Generative Model Selection

The experiments were carried out on Python and deep learning framework Keras with Tensorflow backend, and the hardware environment is a single NVIDIA 2080Ti GPU machine. The training details are illustrated as follows. The optimizer was Adam with momentum $\beta_1 = 0.9$, $\beta_2 = 0.999$, and learning rate $\alpha = 1e^{-5}$. The leaky ReLU action function was used a slope of 0.2. A batch size of 32 and 30000 training epochs was set for training the GAN to achieve equilibrium. In addition, we trained the discriminator once and five times per generator training iteration for the conventional GAN and WGAN respectively.

Each trained generator could generate a data pool containing any number of synthetic preictal samples. Fig. 7 shows the real and synthetic samples generated by the four different generators, and we display four samples randomly selected from each data pool. According to visual inspection, synthetic samples generated by DCGAN and DCWGAN are better than those generated by RGAN and RWGAN. This is because diversities in frequency, amplitude, and period are expected within the single real sample based on visual inspection, and samples should appear distinct among different real samples. However, the two RNN-based GANs both synthesize samples with similar appearances, and diversities cannot be observed within the single sample. Even though synthetic samples generated by DCGAN and DCWGAN
Fig. 7. Real and synthetic single-channel EEG samples. Samples with 3 s duration are shown and the amplitudes of all samples are normalized to the same range:

(a) Real samples, (b), (c), (d) and (e) Synthetic samples generated by DCGAN, DCWGAN, RGAN, RWGAN, respectively. The amplitude values are scaled to [-1,1].

have better appearance, we still cannot distinguish their performance by vision directly. Thus, we need statistical methods to evaluate the quality of samples generated by these two GANs.

Table II shows the scores of three statistical metrics achieved by the four GANs. We randomly selected 100 samples from the data pool generated by each trained generator for score calculation, and real and noise items were added as a control group for comparison. We compared two different sample sets where one is selected from real samples and the other one is selected from samples of corresponding items. In terms of real item, two different real sample sets, where each includes 100 real preictal samples randomly selected from real data pool, are compared. We accumulated total 100 scores, then averaged them to generate final score shown in Table II. According to the statistical scores, we can make unanimous conclusion from visual inspection that DCGAN and DCWGAN outperform RGAN and RWGAN. Furthermore, between the two CNN-based GANs, DCWGAN achieves better scores on all three metrics than DCGAN. Therefore, we selected the DCWGAN as a high-quality model to train the generators of all channels. In terms of implementation of experiments, there are 5,778,066 parameters among the proposed DCWGAN model, and training each channel generator would cost ∼4 h on a single GPU environment.

Table II

| Model    | FDRMSE | FID  | WD  |
|----------|--------|------|-----|
| DCGAN    | 7.10   | 8.8  | 0.66|
| DCWGAN   | 5.39   | 7.8  | 0.45|
| RGAN     | 7.17   | 17.4 | 0.72|
| RWGAN    | 9.32   | 18.5 | 0.92|
| Real     | 1.89   | 2.2  | 0.14|
| Noise    | 9.52   | 410  | 0.89|

*FDRMSE: Frequency domain root mean square error
**FID: Fréchet inception distance
***WD: Wasserstein distance

B. Synthetic Multichannel Preictal Samples

Once adversarial training of the chosen model for all channels was complete, we were able to generate synthetic complete preictal samples based on a channel-wise trained generator. With the availability of sufficient synthetic preictal samples, the effectiveness was evaluated by comparing performance of ES prediction experiments without and with augmented synthetic preictal samples. We implemented ES prediction experiments by patient-specific leave-one-seizure-out cross validation approach, preictal samples before one seizure are regarded as testing samples, and the rest preictal samples before all other seizures are used to train the model. We implemented 5 different runs and recorded averaged values of cross validation results for performance comparisons. Only preictal training set is augmented with synthetic samples at different ratios, and the number of interictal samples were always equal to the number of preictal samples for the training and testing phases.

Table III shows leave-one-seizure-out cross validation performance in the ES prediction experiments under $3 \times$, $5 \times$, and $10 \times$ synthetic preictal sample DA conditions, all of them were used for comparison. From the overall results, all-real condition achieves accuracy of 73.0% and AUC of 0.676, and the three DA conditions achieve accuracies of 76.8%, 77.8%, 78.0% and AUC of 0.696, 0.703, 0.704, respectively. The results reveal that...
TABLE III
PERFORMANCE OF SEIZURE PREDICTION EXPERIMENTS VIA LEAVE-ONE-SEIZURE-OUT CROSS VALIDATION APPROACH

| Patient ID | All-real | 3× DA | 5× DA | 10× DA |
|------------|---------|-------|-------|--------|
|            | ACC     | AUC   | ACC   | AUC    | ACC   | AUC   |
| chb01      | 94.3    | 0.884 | 95.6  | 0.902  | 96.5  | 0.910 |
| chb08      | 86.6    | 0.879 | 88.1  | 0.881  | 88.5  | 0.899 |
| chb09      | 51.5    | 0.449 | 51.5  | 0.507  | 56.6  | 0.526 |
| chb10      | 66.7    | 0.587 | 66.7  | 0.613  | 69.3  | 0.616 |
| chb11      | 77.0    | 0.644 | 77.0  | 0.654  | 83.1  | 0.663 |
| chb19      | 44.4    | 0.483 | 44.4  | 0.501  | 60.7  | 0.497 |
| chb23      | 85.3    | 0.809 | 90.3  | 0.812  | 90.1  | 0.809 |
| Average    | 73.0    | 0.676 | 76.8  | 0.695  | 77.8  | 0.703 |

Synthetic Preictal Samples Augmentation With Three Different Ratios Conditions are Compared With All-Real Condition.

*DA: Data augmentation
**ACC: Accuracy (%)
***AUC: Area under curve

V. DISCUSSION

The results from Section IV-B indicate augmentation with synthetic samples can be implemented effectively on either few times (3×) or more times (10×) condition. However, we cannot confirm that which augmentation ratio can bring best performance according to Table III, because performances are similarly enhanced among three ratios. This is sufficient to demonstrate the goal of enhancing ES prediction by improved prediction performance metrics and addressing the unbalanced dataset issue to achieve robust results. Furthermore, the privacy of patient is an important issue to be considered when medical data are shared, hospitals cannot share private medical data without the permission of patients. Our proposed method that synthesis of effective artificial EEG signals is a solution enabling privacy preservation [45], [46].

Most state-of-the-art seizure prediction machine learning algorithms extracted <30 s segments from raw recordings as training samples, however, numerous recent studies raised that long-term EEG recordings in several minutes take an important role in overcoming limitations of epilepsy studies [47], [48]. Thus, we tested our proposed method to synthesize single-channel EEG recordings in 5 minutes. Because there is a scarcity of existing algorithm using recordings in several minutes as training samples, synthetic recordings will be evaluated only by visual inspection and statistical methods. Fig. 10 shows 3 randomly selected 5min synthetic and real recordings, we can see that they are quite similar visually. In addition, we evaluated them by 3 statistical methods proposed in Section III-F, FDRMSE of 6.58, FID of 8.2 and WD of 0.49 are achieved between synthetic and real recordings. Thus, this can demonstrate similarity between synthetic and real recordings.

In Table IV, we compared our work with prior-art publications related to the application GAN to epilepsy studies. We can see the merits and advantages of our work according to the table. Researchers would like to extract various features from raw EEG signals to train ES prediction classification models, few studies have processed raw EEG samples directly. In addition, researchers may need to determine features or raw signals extracted by different time windows, so that the synthesis of specific features and short-duration signals cannot be widely

Fig. 8. Accuracy histogram of seizure prediction task under the all-real and three data augmentation conditions. Average item represents averaged values of total 7 subjects.

Fig. 9. ROC of seizure prediction without and with data augmentation by synthetic preictal samples. Each line represents performance of each patient-specific model respectively: (a) The all-real condition without data augmentation, (b)–(d) The data augmentation condition with 3×, 5×, and 10× synthetic preictal samples, respectively.
applied to various needs for ES prediction studies. While our synthetic preictal signal is more flexible owing to its raw EEG format.

Mode collapse is one of the most challenging aspects of GAN training. According to the trials of training our proposed GAN, the different noises would converge to the same synthetic single-channel samples if mode collapse occurs. In addition, the hyperparameters, and epochs of adversarial training also have a crucial impact on the successful training of effective generative models. In the practical experiments, we focused on tuning the hyperparameters and the number of epochs to effectively train generative models, eventually parameters are determined as illustrated in Section IV-A. In terms of the proposed method, we trained channel-wise generative models to initially synthesize single-channel preictal samples rather than directly synthesize multichannel samples. This is because multichannel EEG signals containing a large number of data dimensions can easily fail to train the GAN. And only a few channels are still informative, it is proved that few channels or even 2 channels are also effective for the seizure prediction task [49]–[51]. Hence, our single-channel generative model is also flexible for the condition of channel reduction.

VI. CONCLUSION

We proposed a generative model based on DCWGAN to successfully synthesize effective multichannel EEG preictal samples for ES prediction. We initially generated synthetic signal-channel EEG samples and compared the quality for determining the best GAN architecture from four alternatives. Then, complete multichannel samples were synthesized by channel-wise generators. According to the comparisons of ES prediction performance, synthetic preictal sample augmentation can reach or improve the results based on the all-real data condition, which demonstrates the effectiveness of synthetic complete multichannel EEG preictal samples.

To the best of our knowledge, this is the first study to apply a GAN to synthesize artificial multichannel EEG preictal samples for ES prediction. This method provides a solution for the scarcity of a sufficient number of preictal samples as well as privacy concerns. Furthermore, ES prediction is one of the most difficult biomedical applications that relies on EEG signals, and we believe that our method can be applied in further research efforts to synthesize EEG signals for other scenarios.

REFERENCES

[1] World Health Organization, Neurological Disorders: Public Health Challenges. Geneva, Switzerland, World Health Organization, 2006.
[2] J. Yang and M. Sawan, “From seizure detection to smart and fully embedded seizure prediction engine: A review,” IEEE Trans. Biomed. Circuits Syst., vol. 14, no. 5, pp. 1008–1023, Oct. 2020.
[3] E. B. Assi et al., “Towards accurate prediction of epileptic seizures: A review,” Biomed. Signal Process. Control, vol. 34, pp. 144–157, 2017.
[4] Y. Xu, J. Yang, and M. Sawan, “Trends and challenges of processing measurements from wearable devices intended for epileptic seizure prediction,” J. Signal Process. Syst., pp. 1–16, 2021, doi: 10.1007/s11265-021-01659-x.
[5] K. Rasheed et al., “Machine learning for predicting epileptic seizures using EEG signals: A review,” IEEE Rev. Biomed. Eng., vol. 14, pp. 139–155, Jul. 2021, doi: 10.1109/RBME.2020.3008792.
[6] L. Kuhlmann et al., “Seizure prediction-ready for a new era,” Nature Rev. Neurol., vol. 14, no. 10, pp. 618–630, 2018.
[7] D. Hu et al., “Epileptic state classification by fusing hand-crafted and deep learning EEG features,” IEEE Trans. Circuits Syst. II: Exp. Briefs, vol. 68, no. 4, pp. 1542–1546, Apr. 2021.
[8] L. Isenmiedis et al., “Adaptive epileptic seizure prediction system,” IEEE Trans. Biomed. Eng., vol. 50, no. 5, pp. 616–627, May 2003.
[9] K. Fujiwara et al., “Epileptic seizure prediction based on multivariate statistical process control of heart rate variability features,” IEEE Trans. Biomed. Eng., vol. 63, no. 6, pp. 1321–1332, Jun. 2016.
[10] H.-T. Shiao et al., “SVM-based system for prediction of epileptic seizures from iEEG signal,” IEEE Trans. Biomed. Eng., vol. 64, no. 5, pp. 1011–1022, May 2017.
[11] H. Khan et al., “Focal onset seizure prediction using convolutional networks,” IEEE Trans. Biomed. Eng., vol. 65, no. 9, pp. 2109–2118, Sep. 2018.
[12] S. Zhao et al., “Binary single-dimensional convolutional neural network for seizure prediction,” in Proc. IEEE Int. Symp. Circuits Syst., 2020, pp. 1–5.
[13] L. Chisci et al., “Real-time epileptic seizure prediction using ar models and support vector machines,” IEEE Trans. Biomed. Eng., vol. 57, no. 5, pp. 1124–1132, May 2010.
XU et al.: MULTICHANNEL SYNTHETIC PREICTAL EEG SIGNALS TO ENHANCE THE PREDICTION OF EPILEPTIC SEIZURES 3525

[14] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, “EEG-GAN: Generative adversarial networks for electroencephalographic (EEG) brain signals,” 2018, arXiv:1806.01875.

[15] K. Rasheed et al., “A generative model to synthesize EEG data for epileptic seizure prediction,” 2020, arXiv:2012.00430.

[16] N. D. Truong et al., “Epileptic seizure forecasting with generative adversarial networks,” IEEE Access, vol. 7, pp. 143999–144009, 2019.

[17] C. Esteban, S. L. Hyland, and G. Rätsch, “Real-valued (medical) time series generation with recurrent conditional GANs,” 2017, arXiv:1706.02633.

[18] I. J. Goodfellow et al., “Generative adversarial networks,” 2014, arXiv:1406.2661.

[19] F. Fahimi et al., “Generative adversarial networks-based data augmentation for brain–computer interface,” IEEE Trans. Neural Netw. Learn. Syst., vol. 32, no. 9, pp. 4039–4051, Sep. 2021.

[20] N. K. N. Aznan et al., “Simulating brain signals: Creating synthetic EEG data via neural-based generative models for improved SSVEP classification,” in Proc. IEEE Int. Joint Conf. Neural Netw., 2019, pp. 1–8.

[21] Y. Luo and B.-L. Lu, “EEG data augmentation for emotion recognition using a conditional wasserstein GAN,” in Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2018, pp. 2535–2538.

[22] Y. Jiao et al., “Driver sleepiness detection from EEG and EOG signals using GAN and LSTM networks,” Neurocomputing, vol. 408, pp. 100–111, 2020.

[23] S. M. Abdelfattah, G. M. Abdelrahman, and M. Wang, “Augmenting the size of EEG datasets using generative adversarial networks,” in Proc. IEEE Int. Joint Conf. Neural Netw., 2018, pp. 1–6.

[24] S. Haradal, H. Hayashi, and S. Uchida, “Biosignal data augmentation based on generative adversarial networks,” in Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2018, pp. 368–371.

[25] S. Roy et al., “MIEEG-GAN: Generating artificial motor imagery electroencephalography signals,” in Proc. Int. Joint Conf. Neural Netw., 2020, pp. 1–8.

[26] Y. Jiao et al., “Driver sleepiness detection from EEG and EOG signals using GAN and LSTM networks,” Neurocomputing, vol. 408, pp. 100–111, 2020.

[27] N. D. Truong, L. Zhou, and O. Kavehei, “Semi-supervised seizure prediction with generative adversarial networks,” in Proc. IEEE Int. Joint Conf. Neural Netw., 2019, pp. 2369–2372.

[28] N. D. Truong et al., “Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram,” Neural Netw., vol. 105, pp. 104–111, 2018.

[29] D. Pascual et al., “EpilepsyGAN: Synthetic epileptic brain activities with privacy preservation,” IEEE Trans. Biomed. Eng., vol. 68, no. 8, pp. 2435–2446, Aug. 2021.

[30] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” 2014, arXiv:1411.1784.

[31] B. Gao et al., “Generative adversarial network and convolutional neural network-based EEG imbalanced classification model for seizure detection,” Biocybernetics Biomed. Eng., vol. 42, no. 1, pp. 1–15, 2022.

[32] S. You et al., “Unsupervised automatic seizure detection for focal-onset seizure recorded with behind-the-ear EEG using an anomaly-detecting generative adversarial network,” Comput. Methods Programs Biomed., vol. 193, 2020, Art. no 105472.

[33] Y. Guan et al., “Generative adversarial network-based synthetic seizure dataset augmentation,” in Proc. 10th Int. IEEE/EMBS Conf. Neural Eng., 2021, pp. 797–800.

[34] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in Proc. Int. Conf. Mach. Learn., 2017, pp. 214–223.

[35] I. Gulrajani et al., “Improved training of wasserstein GANs,” 2017, arXiv:1704.00028.

[36] A. H. Shoeb, “Application of machine learning to epileptic seizure onset detection and treatment,” Ph.D. dissertation, Massachusetts Inst. Technol., Boston, USA: 2009.

[37] G. H. Klem, “The ten-twenty electrode system of the international federation, the international federation of clinical neurophysiology,” Electroencephalography Clin. Neurophysiol. Suppl., vol. 52, pp. 3–6, 1999.

[38] H. Daoud and M. A. Bayoumi, “Efficient epileptic seizure prediction based on deep learning,” IEEE Trans. Biomed. Circuits Syst., vol. 13, no. 5, pp. 804–813, Oct. 2019.

[39] R. S. Fisher and S. C. Schacter, “The postictal state: A neglected entity in the management of epilepsy,” Epilepsy Behav., vol. 1, no. 1, pp. 52–59, 2000.

[40] Y. Xu et al., “An end-to-end deep learning approach for epileptic seizure prediction,” in Proc. 2nd IEEE Int. Conf. Artif. Intell. Circuits Syst., 2020, pp. 266–270.

[41] R. Hussein et al., “Human intracranial EEG quantitative analysis and automatic feature learning for epileptic seizure prediction,” 2019, arXiv:1904.03603.

[42] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.

[43] A. Borji, “Pros and cons of gan evaluation measures,” Comput. Vis. Image Understanding, vol. 179, pp. 41–65, 2019.

[44] S. Barratt and R. Sharma, “A note on the inception score,” 2018, arXiv:1801.01973.

[45] W. G. Van Panhuis et al., “A systematic review of barriers to data sharing in public health,” BMC Public Health, vol. 14, no. 1, pp. 1–9, 2014.

[46] C. Clifton et al., “Privacy-preserving data integration and sharing,” in Proc. 9th ACM SIGMOD Workshop Res. Issues Data Mining Knowl. Discov., 2004, pp. 19–26.

[47] J. Duun-Henriksen et al., “A new era in electroencephalographic monitoring? subspace devices for ultra-long-term recordings,” Epilepsia, vol. 61, no. 9, pp. 1805–1817, 2020.

[48] T. Xinghua et al., “The clinical value of long-term electroencephalogram (EEG) in seizure-free populations: Implications from a cross-sectional study,” BMC Neurol., vol. 20, no. 1, pp. 1–7, 2020.

[49] R. Jana and I. Mukherjee, “Deep learning based efficient epileptic seizure prediction with EEG channel optimization,” Biomed. Signal Process. Control, vol. 68, 2021, Art. no. 102767.

[50] Y. Salant, I. Gath, and O. Henriksen, “Prediction of epileptic seizures from two-channel EEG,” Med. Biol. Eng. Comput., vol. 36, no. 5, pp. 549–556, 1998.

[51] N.-F. Chang et al., “Channel selection for epilepsy seizure prediction method based on machine learning,” in Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc., 2012, pp. 5162–5165.