EpilepsyGAN: Synthetic Epileptic Brain Activities with Privacy Preservation

Damián Pascual†, Alireza Amirshahi†, Amir Aminifar, *Member, IEEE*, David Atienza, *Fellow, IEEE*, Philippe Ryvlin, Roger Wattenhofer

***Abstract*—Epilepsy is a chronic neurological disorder affecting more than 65 million people worldwide and manifested by recurrent unprovoked seizures. The unpredictability of seizures not only degrades the quality of life of the patients, but it can also be life-threatening. Modern systems monitoring electroencephalography (EEG) signals are being currently developed with the view to detect epileptic seizures in order to alert caregivers and reduce the impact of seizures on patients' quality of life. Such seizure detection systems employ state-of-the-art machine learning algorithms that require a large amount of labeled personal data for training. However, acquiring EEG signals during epileptic seizures is a costly and time-consuming process for medical experts and patients. Furthermore, this data often contains sensitive personal information, presenting privacy concerns. In this work, we generate synthetic seizure-like brain electrical activities, i.e., EEG signals, that can be used to train seizure detection algorithms, alleviating the need for sensitive recorded data. Our experiments show that the synthetic seizure data generated with our GAN model succeeds at preserving the privacy of the patients without producing any degradation in performance during seizure monitoring.**

***Index Terms*—Synthetic Brain Activities, Generative Adversarial Networks (GANs), Epilepsy Monitoring, Seizure Detection, Privacy.**

I. Introduction

Epilepsy is the fourth most common chronic neurological disorder worldwide [1], affecting over 65 million people. Epilepsy manifests itself by recurrent unprovoked seizures due to abnormal activity in the brain. One third of the epilepsy patients suffer from drug-resistant uncontrolled seizures, which time of occurrence is usually unpredictable. The length of the seizures can range from few seconds to several minutes with a large variety of symptoms, including sensory auras, loss of awareness, behavioral arrest, automatic movements and full body convulsions [2]. Epilepsy not only degrades the quality of life of the patients, but it is also associated with a mortality rate 5 times higher among patients with recurrent seizures than in the corresponding group of the general population [3], [4]. In epileptologists terminology, ictal samples are those extracted from the seizure segments, while inter-ictal samples are those extracted from non-seizure segments.

A promising solution to reduce mortality and to improve the living standard and independence of epilepsy patients is continuous real-time monitoring using wearable devices [5]–[10]. Wearable devices can continuously collect and process EEG signals from the patient in real time during extended periods of time in order to detect ictal periods. In this manner, upon occurrence of an epileptic seizure, an alert can be sent automatically to caregivers or family members.

However, a fundamental barrier in developing reliable epileptic seizure detection systems is the lack of sufficient volume of training data. Indeed, modern detection systems are driven by machine-learning-based algorithms [11], [12] that require a considerable amount of samples of recorded ictal periods in order to reliably detect future seizures. Collecting and labeling EEG data from epilepsy patients is a costly process that currently requires the patients to suffer seizures while being recorded in a monitoring unit. Such recordings are performed in clinical practice in a minority of patients and over short periods of time, typically a week, enabling to only record a few seizures per patient [13]. The privacy concerns that exist around sharing medical data exacerbates this problem. In particular, the possibility of patient re-identification from anonymized datasets [14] and the risk of data leakage hinders sharing medical data.

In this work, to address the aforementioned problems, we propose EpilepsyGAN, a Generative Adversarial Network (GAN) [15] that produces high quality synthetic epileptic seizure signals and we demonstrate the effectiveness of the proposed framework. To the best of our knowledge, this is the first time that seizure EEG samples are generated and used to train epilepsy detection algorithms. Then, we leverage the generative power of our model to address the privacy concerns for the case of epilepsy monitoring. To this end, first, we show that training an epilepsy monitoring system with synthetic data does not degrade the seizure detection performance compared to training based on real data. Further, we highlight the potential privacy concerns in the context of biomedical applications and, in particular, epilepsy, and demonstrate that the use of synthetic data hinders the re-identification of the patients. This constitutes a real-life application of GANs with a direct impact on healthcare and medical data privacy. Therefore, the main contributions of our work are summarized as follows:

1) A generative model capable of producing realistic synthetic seizure signals that can train an epilepsy monitoring system with a similar performance as the real seizure signals, in terms of detection of epileptic seizures.
2) Application of synthetic data for privacy preservation

† These authors contributed equally to this work.

D. Pascual and R. Wattenhofer are with the Swiss Federal Institute of Technology Zurich (ETHZ), A. Amirshahi and D. Atienza are with the Swiss Federal Institute of Technology Lausanne (EPFL), A. Aminifar is with Lund University (LU), and P. Ryvlin is with the Lausanne University Hospital (CHUV).

and a comparative study of the vulnerability of real and synthetic ictal data to patient re-identification, showing that synthetic data is up to 7.2 times less vulnerable when compared to the real data.

II. Related Work

The field of synthetic data generation has undergone extraordinary progress thanks to GANs. In recent years, GANs have attained outstanding results in a wide variety of challenging areas such as computer vision [16], [17], audio [18], [19] or natural language processing [20], [21]. However, their success in the medical domain has been more limited.

The generation of reliable synthetic data for medical applications has been extensively studied in the literature during the last years. Several studies have used synthetic data in areas such as medical imaging [22]–[24] and Intensive Care Unit (ICU) monitoring [25]–[27] to augment existing training sets in order to improve detection accuracy. Although this data augmentation approach has proved effective, previous attempts to train only with synthetic data have reported such a strong degradation in performance [23], [25] that it has not been possible so far to dispense with real training data. Therefore, the scenario where no real training data can be accessed and only a purely synthetic training set may be available remains unsolved. This is, however, a common scenario in several medical applications, including epilepsy, given the difficulties and privacy concerns associated with collecting and sharing medical data [28].

In the specific case of brain signals, the application of GANs to the generation of realistic synthetic signals has obtained very limited success so far: [29] generated EEG-like signals, without demonstrating the quality of the synthetic data in any specific task or pathology detection. [30] generated synthetic EEG data to augment existing real training sets for Brain-Computer Interfaces, but they did not evaluate on fully synthetic training sets. [31] used a GAN to upsample the spatial resolution of EEG signals and, despite the improvement in visual quality, the resulting training set produced a degradation of 4–9% of accuracy in a mental imagery classification task in comparison to the original training set. Nevertheless, despite targeting the generation of synthetic EEG, the current literature has not addressed epilepsy, and specifically the generation of ictal samples.

Furthermore, in the last years, GAN models have attracted attention because of their ability to generate realistic synthetic data in privacy-sensitive applications. In the context of medical applications with sensitive data, DPGAN [32] and PATE-GAN [33] propose differentially private GAN models, where the privacy is obtained by adding noise to the gradients of the model. Their evaluation indicates that both PATE-GAN and DPGAN suffer from a substantial quality decrease in high dimensional datasets, such as, the UCI Epileptic Seizure Recognition dataset [34]. Differential privacy techniques introduce a well-known trade-off between privacy level (i.e., the magnitude of added noise) and performance [35]. That is, as we increase the noise magnitude, the synthetic data becomes more differentially private at the cost of a gradual loss of utility.

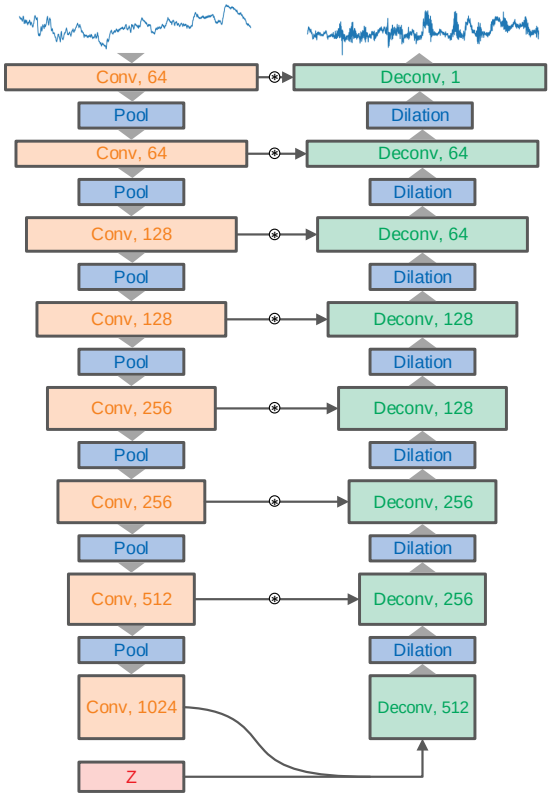


Fig. 1: Encoder and decoder blocks of the EpilepsyGAN generator. The encoder consists of convolution and pooling layers. Gaussian noise $\mathbf{Z}$ is added to the encoder's output. The decoder has de-convolution and dilation layers to increase their outputs length, as well as decreasing the depth. The weighted skip connections between the encoder and decoder are also shown in this figure.

A second class of privacy preservation techniques, such as MedGAN [36], generates high-dimensional synthetic discrete variables by a combination of an autoencoder and a GAN. Their results show that the 1-to-1 mapping between the generated data and the corresponding training records is weak, which means that the synthethic data preserve patients' privacy. MedGAN obtains impressive results on electronic health record (EHR) data, however, this data is discrete in nature and hence this type of models cannot be used for generating continuous biomedical signals such as EEG.

In this article, we propose the use of GANs to generate synthetic ictal data, which are rare events in the EEG recordings, and we evaluate the quality and utility of the generated samples on the task of seizure detection. Moreover, we show the possibility of patient re-identification and demonstrate that using synthetic signals produced by our EpilepsyGAN model alleviates the privacy concerns associated with sharing sensitive medical data in the epileptic seizure detection problem.

III. Generative Model

GANs are a class of deep generative models in which two neural networks are trained simultaneously, while competing

in a two-player minimax game. One network is a discriminator that estimates whether a sample is real or synthetic. The other network is a generator whose task is to generate realistic synthetic samples that maximize the probability of the discriminator making a mistake. During training, the discriminator improves its ability to recognize synthetic samples while the generator learns to produce increasingly realistic samples to deceive the discriminator. In this adversarial setting, the equilibrium is reached when the generator produces realistic samples such that the discriminator cannot distinguish whether they are real or synthetic.

EpilepsyGAN is a conditional GAN [37] that, given inter-ictal EEG samples at the input, it generates EEG samples of epileptic seizures. The rationale behind our design is that, while epileptic seizures are very costly to record due to their unpredictability and low frequency of occurrence, inter-ictal signals can be easily recorded in any moment. As a result, we condition the network on inter-ictal samples from the target patient in order to provide additional information to the generator that can be exploited to produce more realistic seizure samples. In this way, we can use an already existing database to train our GAN and then use the GAN to generate seizure samples for a new patient.

The architecture of our GAN is modeled after the SEGAN from [38], [39]. Our generator, depicted in Fig. 1, is a U-net [40] convolutional autoencoder network with weighted skip connections. Generally, an autoencoder consists of two symmetric parts, an encoder that processes the input sample and generates a latent code, and a decoder that restores the original sample by decoding the latent code. However, in our case, the decoder does not restore the original inter-ictal sample but translates the latent code into an ictal sample. In order to obtain sample variety, stochasticity is introduced into the model by concatenating Gaussian noise with mean 0 and standard deviation 1 to the latent code. The skip connections multiply the feature maps at each layer of the encoder with a weight which is learnt during training and then, the result of that operation is added to the corresponding feature map of the decoder. Thus, the weights of the skip connections regulate the amount of information that is fed from the encoder into the decoder. The discriminator of our GAN has the same structure as the encoder of the generator, but it includes an additional fully connected layer at the output. This way, the discriminator outputs a single value between 0 and 1, where 1 represents the real class and 0 the synthetic class.

To obtain the network parameters of the model, an optimization problem is solved iteratively in the training phase, during which the loss functions of the generator and the discriminator are alternatively minimized. In our model, these losses are based on the Least Squares GAN (LSGAN) [41]. Consequently, the minimization objective of the discriminator is given by:

$$\min_{\theta_D} \mathcal{L}_D(\theta_D) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[(D(\mathbf{x};\theta_D)-1)^2] + \mathbb{E}_{\hat{\mathbf{x}}\sim p_{data}(\hat{\mathbf{x}})}[(D(G(\hat{\mathbf{x}});\theta_D))^2], \quad (1)$$

In this objective, the function D corresponds to the discriminator and G to the generator. As mentioned before, the output of the discriminator lies between 0 and 1. The θ_D are the network parameters of the discriminator. The input data $\mathbf{x}$ is a real ictal signal sampled from the real data distribution $p_{data}(\mathbf{x})$. In the second term, $G(\hat{\mathbf{x}})$ is a synthetic ictal sample generated from the inter-ictal input $\hat{\mathbf{x}}$. The first term of the loss function pushes the discriminator to output 1 when the input is a real ictal sample $\mathbf{x}$, whereas the second term is minimized when the discriminator outputs a 0 given a synthetic sample.

On the other hand, the loss function of the generator, with network parameters θ_G, is given by:

$$\min_{\theta_G} \mathcal{L}_G(\theta_G) = \mathbb{E}_{\hat{\mathbf{x}}\sim p_{data}(\hat{\mathbf{x}})}[(D(G(\hat{\mathbf{x}};\theta_G))-1)^2] + \lambda||G(\hat{\mathbf{x}};\theta_G)-\mathbf{y}||_1, \quad (2)$$

Where $\mathbf{y}$ denotes the reference ictal signal paired with the inter-ictal sample $\hat{\mathbf{x}}$. To make the training more stable, the generator's loss includes a weighted L_1 regularization term. This term ensures that the generated signal is similar to the reference ictal signal $\mathbf{y}$. We observed that without this regularization term the parameters of the network quickly saturate during training, generating meaningless outputs. In Equation 2, λ is a hyperparameter that we tuned to 100 using the validation set in order to scale both terms of the loss function to a comparable magnitude, preventing the regularization term from dominating the optimization problem. Furthermore, the first term of the generator loss encourages the generator to produce synthetic samples that are classified as 1, i.e., real, by the discriminator, which is adversarial with respect to the discriminator's loss function. Hence, the competing interests of the generator and the discriminator during training drive the generator to produce more and more realistic samples.

A. Architecture Details

The input to the generator are samples of length 2048 points. The encoder consists of eight blocks that alternate a convolutional layer with a max-pooling layer of size 2 and stride 2. The 1D filters used in the convolutions have size 31 and stride 1. As shown in Fig. 1, the number of filters increases gradually in the encoder. The encoder's input has shape of 2048x1, while its output is 8x1024. The Gaussian noise $\mathbf{Z}$ concatenated to the latent code has that same shape, i.e., 8x1024. In the decoder, the filter size of dilations and deconvolutions are the same as the poolings and convolutions in the encoder. This way, the U-shaped encoder-decoder structure of the generator is symmetric, with an output shape of 2048x1.

The activation used is the leakyReLu function [42], except for the last block of the decoder, where we use the hyperbolic tangent function. All convolutions and deconvolutions are unbiased and spectral normalization [43] is applied before each block in both the generator and the discriminator. On top of that, in the discriminator we apply virtual batch normalization [44].

To train the model, we use the Adam [45] optimizer with 0 and 0.9 for the values of β_1 and β_2, respectively, and learning rates 0.0001 for the generator and 0.0004 for the discriminator. We found through a hyperparameter search that these values,

which coincide with those in [38], perform the best also in our application. The size of the minibatches of data employed during training is 100 samples.

B. Real-World Scenario

Generation of synthetic seizures is of special importance in the scenario where a new patient needs to be monitored and no seizure data is available to train the monitoring system. Instead of bringing the patient into an in-hospital recording unit until she suffers enough epileptic seizures to train the system, the hospital could use a generative model trained on data from other patients to generate synthetic data for the new patient. This is the scenario considered in this work, which is divided in two parts: training the generative model (Section III-C); and generation of synthetic ictal data and real-time seizure monitoring using the generated samples for training (Section V).

A high-level picture of this setting is depicted in Fig. 2. Given N patients, we consider that in the real-world scenario we have ictal data for $N-1$ patients and that the i-th patient not included in the $N-1$ patients is a new patient that requires monitoring and whose ictal data is not available. However, given that inter-ictal data is straightforward to record and can be done efficiently at any moment, we assume that we have inter-ictal data also for the i-th patient. Then, (1) we train our generative model using the data from the $N-1$ patients, and (2) we generate ictal data from inter-ictal data for the new patient i. Note that the model does not need to be retrained for every new patient; after training it once, synthetic data for any number of new patients can be generated. Once we have trained the EpilepsyGAN and produced synthetic ictal data for the target patient i, we use this synthetic data to train a classifier (e.g., Random Forest) for the task of seizure detection. Specifically, the task is to detect the real ictal samples of patient i training on the real inter-ictal and synthetic ictal samples of patient i. The performance of this classifier gives us an understanding of the quality of the synthetic data, i.e., how suitable it is to train monitoring algorithms.

C. Generative Model Training

To train EpilepsyGAN, we use the data from the EPILEPSIAE project database [46], which is one of the world's largest public databases for seizure detection. The dataset contains recordings from 30 different epilepsy patients with a total of 277 epileptic seizures that sum up to a duration of 21,001 seconds altogether. The EEG data is collected at a sampling frequency of 256 Hz and it is divided into recordings of one hour, each one corresponding to one recording session. The number of one-hour recordings varies for each patient in a range between 96 and 281 sessions.

In this work, we target the setup of real-world and stigma-free wearable monitoring devices [47] and thus we consider only the electrodes F7T3 and F8T4 in the standard 10–20 system [48], which can be easily hidden in glasses [7].amir These electrodes are most likely to capture the temporal lobe epilepsy. In the EPILEPSIAE database, 25 patients out of 30 experienced at least one seizure session localized in the temporal lobe. We use samples of four seconds of duration, since this length is effective to detect epileptic seizures. Given that the data was recorded at a frequency of 256 Hz, this results in samples of length 2048, i.e., 1024 per electrode. The four-second long ictal samples are collected with three seconds of overlap in order to augment the amount of training ictal samples, while the inter-ictal samples are collected with no overlap from recordings where no seizure occurred. Moreover, we do not apply any filtering or preprocessing step to the data and we work directly with the raw EEG signals.

To construct the training set we pair each ictal sample to an inter-ictal sample picked at random from the same patient. This means that each inter-ictal sample given at the input of the generator is associated to an ictal sample that is used as a reference to guide the training. In this manner, the generator learns to map inter-ictal samples to ictal samples of the corresponding patient. Since the ictal samples are overlapped by one second and the inter-ictal are not, similar ictal samples (overlapped) are paired to dissimilar inter-ictal samples (not overlapped), therefore, to learn the transformation the model needs to understand the relation between ictal and inter-ictal distributions rather than between arbitrarily paired portions of signal; this effectively acts as a regularizer.

As explained in Section III-B, in order to train the model, the leave-one-out strategy is followed: for each target patient i, EpilepsyGAN is trained using the ictal and inter-ictal data coming from all other patients ($N-1$), and thus, it can be used to generate synthetic ictal samples for the unseen patient. The exact number of training samples depends on the number of seconds of seizure recording available in the database for all patients except for the left-out patient i and, although it varies slightly, it is approximately 20,000 samples. Following this scheme, the GAN is trained independently leaving-out one patient i each time and thus, we obtain one model per patient, i.e., 30 models.

IV. Privacy Preservation

Seizure signals are considered sensitive data in privacy because they can disclose the health status of the patients. This includes the severity of the seizures, the type of epilepsy, and the very fact that the user suffers from epilepsy. This sensitive information may be exploited by a byzantine third-party and therefore, sharing and storing seizure signals can jeopardize the privacy of the patients. To address this privacy problem and protect patient confidentiality, we propose to use synthetic ictal signals instead of the real ictal signals. The synthetic data is produced by the EpilepsyGAN as presented in Section III.

In this section, we introduce an evaluation procedure to assess whether privacy is preserved when using synthetic EEG ictal samples. To compare the real ictal and synthetic ictal signals in the privacy context, we suppose that these signals are anonymized, i.e., there is no identity information. We use the identifiability of the ictal signals as a metric to show the existence of privacy concerns in epilepsy signals.

A. Patient Identification

We define *patient identification* as the problem of identifying the person in a dataset to whom an input signal belongs.

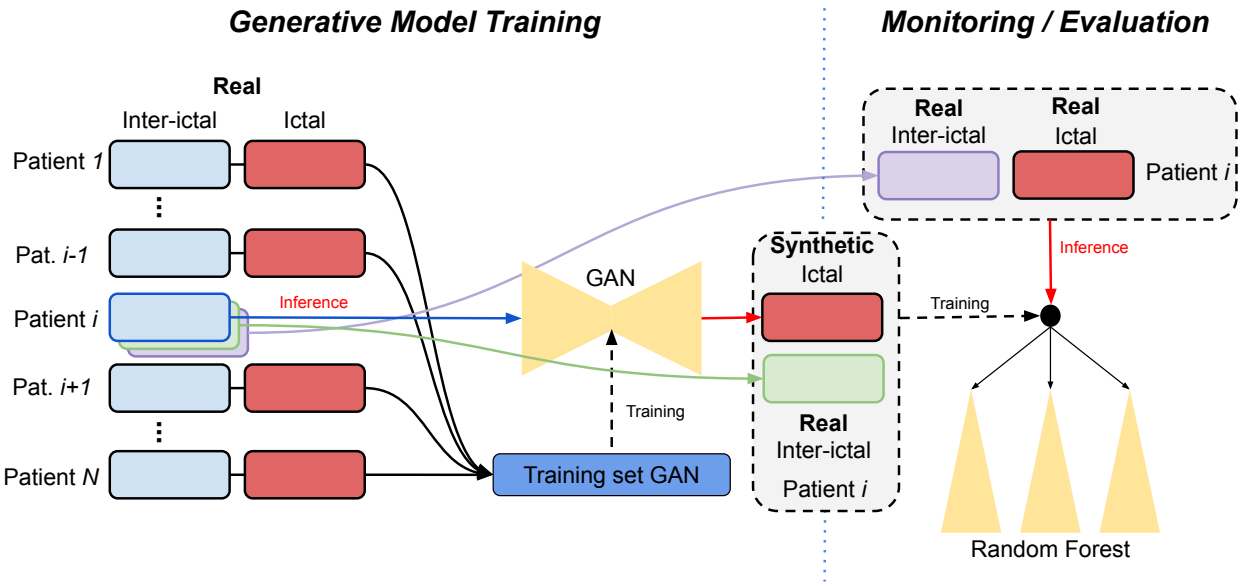


Fig. 2: Real-world scenario considered for training and evaluation of the model.

In this task, a classification model learns to map every input signal to its corresponding patient. For instance, the accuracy of random patient identification is $\frac{1}{N}$, where N is the total number of patients considered in the experiment. We consider this random patient identification, denoted by Exp_{rand}, as a baseline for reference.

Let us now focus on the possibility of patient identification based on the original ictal signals versus the synthetic ictal signals. To this end, we propose two experiments: Exp_{orig} and Exp_{synt}. In the first step of both experiments, a convolutional neural network (CNN) is trained for patient identification based only on the inter-ictal signals. Then, in Exp_{orig}, the model is tested on the original ictal signals, i.e., to map each input ictal signal to the patient to whom the ictal signal belongs. We follow the same procedure in Exp_{synt}, but the model is tested on the synthetic ictal signals.

B. Classification Model Architecture

The CNN model proposed for the patient identification task is inspired from ResNet [49] and is illustrated in Fig. 3.

The model takes a sample of normalized EEG signals as input and generates the output to identify to which patient the input signal belongs. The total input length is 4136, and it comprises four parts:

- 4-second EEG signal from the electrode F7T3,
- 4-second EEG signal from the electrode F8T4,
- Discrete Wavelet Transform (DWT) of the first part,
- DWT of the second part.

The first and second parts have 2048 points in total. The DWT used in the third and fourth parts are Daubechies-4 and performed up to three levels. We found empirically that adding the DWT to the input information significantly helps the model in the identification task. We conjecture that the power in the different frequency bands are representative features of the patient identity and hence, providing the DWT of the input signal eases the task of the classifier.

As shown in Fig. 3, the first layer in the model is a convolutional layer with 32 filters. Each filter has a length of 5 and stride 2. The model includes four residual blocks: $B1$ to $B4$. All blocks have two convolutional layers and a shortcut path (residual connection). Also, batch normalization [50] and rectified linear unit (ReLU) [51] are applied before each convolution in the blocks. Thus, the overall structure of blocks follows the pre-activation block design in [52]. In each block, the first and second convolutional layers have strides of 2 and 1, respectively. Hence, the output of each block is subsampled by a factor of 2 in comparison to its input. To ensure that the output of the convolutional blocks is compatible to the output of the residual connection, i.e., of the same size, we alternate two downsampling strategies. In blocks $B2$ and $B4$ the max pooling operator is applied in the residual connection, while in blocks $B1$ and $B3$ a convolutional layer with a filter size of 1 and stride 2 is used. This way, we provide a reasonable number of additional parameters, which determines the expressive power of the model.

The number of filters gradually increases from 32 in the first convolution layer to 128 in the convolutions in $B4$. In this manner, deeper layers can detect a more complex combination of patterns. After the residual blocks, the model has two fully connected (FC) layers and a softmax activation layer. The final fully connected layer has as many neurons as the number of patients in the training set N. Finally, the softmax layer returns a probability distribution over patients.

V. Evaluation Procedure

For certain types of data, such as images or audio, a person can naturally perceive the quality of a synthetic sample. Therefore, metrics that correlate well with human perception can be used in these fields to evaluate synthetically generated data. However, EEG signals are abstract and so, a metric

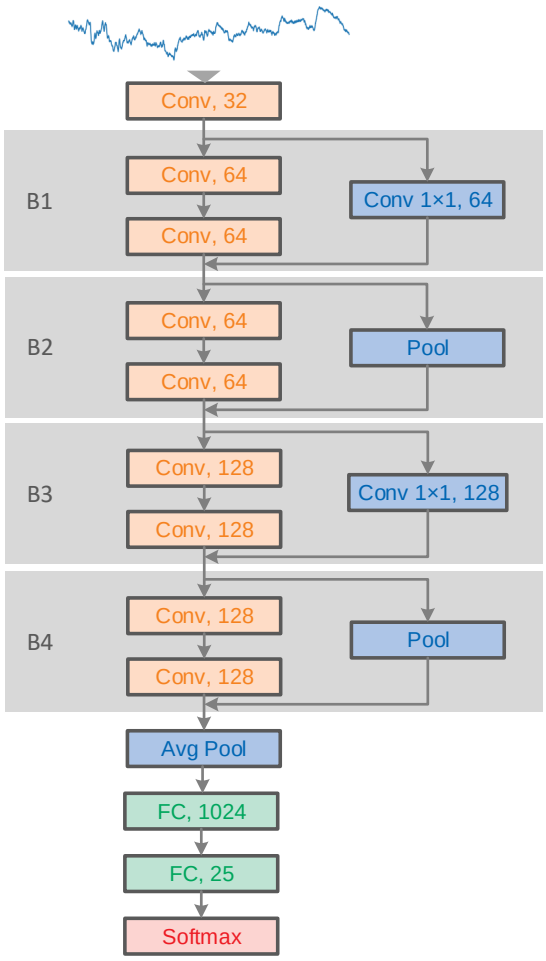


Fig. 3: Model architecture for the patient identification. In the model, $B1$, $B2$, $B3$, and $B4$ are blocks with residual skip connections. Each block has two convolution layers and a skip connection. Each convolution layer is followed by a batch normalization and a ReLU activation, which are not shown in this figure. The number of filters is indicated inside each layer.

derived from human perception is not informative. An alternative is to let medical experts evaluate the quality of the generated samples. Apart from the logistic difficulties derived from getting access to medical experts, this approach makes the reproduction of the results and comparison with new techniques difficult.

In this work, our goal is to generate synthetic ictal samples for seizure detection. Hence, we propose to directly quantify the quality of synthetic EEG ictal samples in the seizure detection problem. We design an evaluation procedure based on this task that allows us to systematically evaluate synthetic ictal samples and follows a similar philosophy as the "Train on Synthetic, Test on Real" method from [25]. This procedure, corresponds to the second part of the real world scenario described in Section III-B.

Our evaluation is based on a state-of-the-art classifier that uses the random forest algorithm [53] to determine whether an incoming four-second sample is an ictal or an inter-ictal sample. This classifier uses 500 trees and follows that of [7], which is tailored to a stigma-free wearable device for epilepsy monitoring. Therefore, using that same classifier evaluates the feasibility of using our synthetic samples not only in a medical environment, but also in the more restrictive setting of continuous monitoring using wearable technologies.

Each patient i is targeted independently. We divide the real inter-ictal samples from the target patient i in three *non-overlapping* sets that we denote as $S_i^{GAN}, S_i^{Train}, S_i^{Test}$. Our procedure considers the scenario where the only data available for training the seizure detection algorithm are real inter-ictal samples S_i^{Train} and *synthetic ictal* samples of target patient i generated from S_i^{GAN}. As a baseline for comparison, we consider the case where *real* ictal samples from all other $N-1$ patients and inter-ictal samples S_i^{Train} from the target patient are available. The evaluation procedure can be structured in four steps:

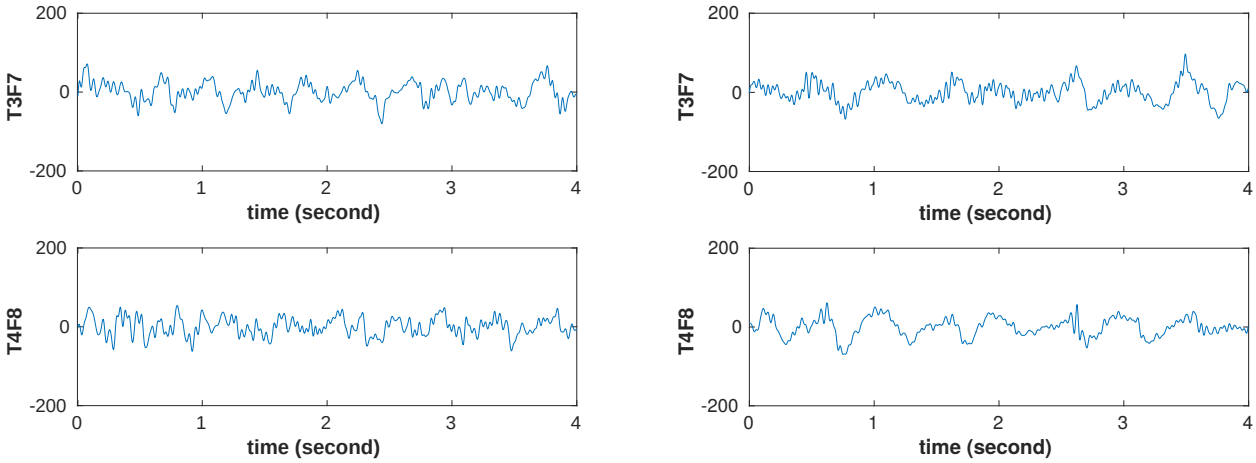
1) Construction of target training set T_i, baseline training set B_i, and test set E_i.
2) Identification and extraction of relevant domain features.
3) Training the classifier for seizure detection with the target training set T_i and evaluation on the test set E_i.
4) Training the classifier for seizure detection with the baseline training set B_i and evaluation on the test set E_i.

Therefore, we first build the target training set T_i, baseline training set B_i, and the test set E_i in the following way:

- Target training set, T_i: consists of 2,000 *synthetic ictal* samples (generated using real inter-ictal samples from the target patient S_i^{GAN}) and 2,000 real inter-ictal samples (S_i^{Train}) from the target patient i.
- Baseline training set, B_i: consists of 2,000 samples of real seizures randomly selected from all the $N-1$ patients in the database except for the target patient i, as well as 2,000 inter-ictal samples from the target patient (S_i^{Train}).
- Test set, E_i: consists of all the ictal samples of the target patient i without overlap and twice as many inter-ictal samples from the same patient, i.e., S_i^{Test}.

Note that, the synthetic ictal samples are the only aspect that differs between the target training set T_i and the baseline training set B_i. This ensures that the performance difference between both training sets is due exclusively to the quality of the ictal samples and not to the real inter-ictal samples S_i^{Train}. On the other hand, we build an unbalanced test set E_i with twice as many inter-ictal as ictal samples; nevertheless, the metrics we use to measure the performance of our system (sensitivity and specificity) are invariant to the ratio of inter-ictal and ictal in the dataset.The size of the test set changes slightly among patients, given that a different number of seizures are recorded for each patient.

Once the data is split in training and test sets, the second step performed in our procedure is identification and extraction of relevant features. In this stage, we follow again [7] and extract 54 features of power and non-linearity per electrode, and, since we consider two electrodes, the total number of features is 108. These features are subsequently extracted for all the samples in both the training and the test sets. To



(a) 4 seconds of a real EEG ictal sample

(b) 4 seconds of a synthetic EEG ictal sample

Fig. 4: Comparison of a real and a synthetic ictal sample for electrodes T3F7 and T4F8. It is visible how in both cases the delta-theta rhythm are present in the signals.

calculate the non-linear features, the signal is decomposed using the discrete wavelet transform down to level seven. The nonlinear features extracted are: sixth and seventh level sample entropy [54] for $k = 0.2$ and $k = 0.35$; third, fourth, fifth, sixth and seventh level permutation entropy [55] for $n = 3$, $n = 5$ and $n = 7$; third, fourth, fifth, sixth and seventh level, as well as raw signal, Shannon, Renyi and Tsallis entropies. The power features are: total power, total and relative band power in the bands delta [0.5,4] Hz, theta [4,8] Hz, alpha [8,12] Hz, beta [13,30] Hz, gamma [30,45] Hz as well as in the bands [0,0.1] Hz, [0.1,0.5] Hz, [12,13] Hz.

After the features are extracted, the target training set is used to train the random forest classifier and the resulting classifier is evaluated against the test set. The same procedure is then repeated for the baseline training set, which results are used as reference. We repeat these experiments 15 times for each patient, shuffling the data each time, in order to make our results robust against different splits of data, as well as against different learnt configurations of the random forest.

The proposed procedure provides a framework to easily compare different generative models by simply using in the target training set the synthetic ictal samples generated by a given model. Furthermore, using the downstream task of seizure detection in wearable devices as a proxy for signal quality allows us to directly draw conclusions on the utility of a set of synthetic ictal samples. Therefore, subsequent work could rely on this procedure to compare the quality and utility of synthetically generated ictal data.

VI. Results

A. Model Convergence

To determine the convergence point of EpilepsyGAN, we randomly select five patients (patients 2, 8, 9, 17 and 30) as validation set and evaluate the data generated by the model at different training epochs; we also use this validation set to determine the other hyper-parameters of the model, described in Section III. Patient 8 suffers from frontal lobe epilepsy, while the other patients in the validation set have temporal lobe epilepsy. We calculate the performance score of each model as the geometric mean of sensitivity and specificity, where sensitivity and specificity are the true-positive rate and the true-negative rate, respectively. We report the geometric mean because it is the only correct average of normalized values [56]. Table I shows the mean performance over the five patients for different number of epochs. We observe how training longer than 100 epochs hurts performance, and thus, we train EpilepsyGAN for 100 epochs.

TABLE I: Mean performance on the validation set for different training epochs.

Number of epochs	Mean Score
60	74.61
70	76.24
80	78.36
90	75.70
100	**78.72**
110	77.34
120	74.80
130	73.90
140	75.97
150	75.14

B. Similarity Between Real and Synthetic Samples

Before performing a quantitative evaluation of the quality and utility of the samples generated by our GAN model, we visually inspect the synthetic ictal samples and compare them with real ictal samples. In this way, we verify that the model generates samples that can not only train a detection system, but also are realistic and follow patterns associated with epileptic seizures. An example is shown in Fig. 4, where we compare 4 seconds of real ictal EEG signal for the two electrodes of interest with 4 seconds of synthetic ictal EEG generated for the same patient. We observe the presence of the

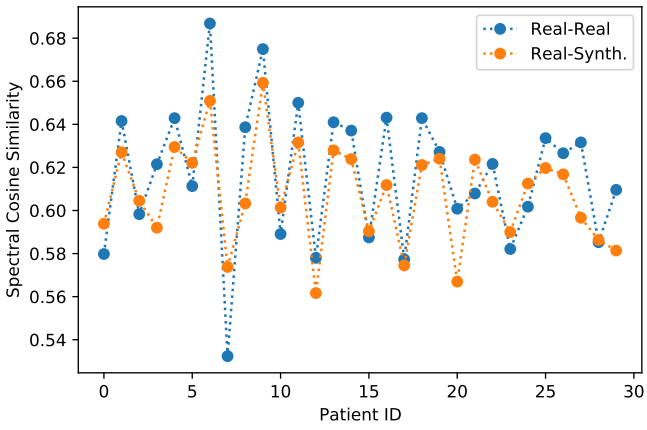


Fig. 5: Similarity between real ictal signals and between real and synthetic ictal signals for each patient.

well-known delta–theta rhythm, i.e., rhythmic slow activity with a frequency of oscillation in 0.5–4 or 4–7 hertz. This pattern is a clear indication of the correct generation of the ictal discharge and epileptic seizure segment in the synthetically-generated EEG signals [57].

Next, we study quantitatively how similar our synthetic ictal data is to real ictal data. In order to understand the similarity between synthetic and real ictal samples, we also calculate the similarity between real samples and use this as a reference. This way, for each patient we obtain two values: real-to-real similarity and real-to-synthetic-similarity. To calculate each of these similarity values we do the following:

1) We select 2000 random pairs of signals from the same patient. In the real-to-synthetic case, in each pair one of the signals is real and the other one synthetic; in the real-to-real case both signals are real and different.
2) For each pair of signals, we calculate the FFT for each electrode separately; this yields two FFTs per signal, i.e., four FFTs per pair.
3) We calculate the cosine similarity between the FFTs of the corresponding electrodes of each pair; then, average over both electrodes.
4) Finally, we average over the 2000 pairs, obtaining one number per patient.

The justification for measuring the similarity between FFTs instead of between raw signals is that the epileptic seizures are manifested as changes in certain frequency bands; thus, the FFT better represents the information we are interested in. The choice of the cosine similarity is reasonable since this metric is sensitive to shifts in the data. This way, if two signals present high values in different frequency bands the similarity value will be small. The results of this similarity study are shown in Figure 5. Note that the range of the vertical axis is narrow, between 0.54 and 0.68.

The results show that in both cases, real-to-real and real-to-synthetic, signal similarity is in the same range. Furthermore, it is clear from the figure that in both cases the similarity values follows the same pattern across patients, i.e., patients with high real-to-real similarity tend to have high real-to-synthetic similarity and viceversa. This experiment gives a first evidence of the validity of our generative approach.

TABLE II: Geometric mean of sensitivity and specificity per patient of our evaluation.

Patient ID	EEG foci	Baseline (%)	Synthetic (%)	Difference (%)
1	Central - Parietal	73.49	80.06	**+6.57**
3	Temporal	77.59	82.84	**+5.25**
4	Temporal	76.34	78.33	**+1.99**
5	Temporal - Frontal	64.86	68.19	**+3.33**
6	Temporal	74.10	74.74	**+0.64**
7	Temporal	68.11	68.59	**+0.48**
10	Temporal - Frontal	66.84	65.87	-0.97
11	Temporal - Central	81.03	83.66	**+2.63**
12	Temporal	63.00	66.56	**+3.56**
13	Temporal - Frontal	77.20	78.54	**+1.34**
14	Temporal	74.32	76.51	**+2.19**
15	Temporal	74.25	74.07	-0.18
16	Temporal	78.11	80.64	**+2.53**
18	Temporal	66.20	71.62	**+5.42**
19	Temporal	76.95	78.13	**+1.18**
20	Temporal	73.42	68.67	-4.75
21	Occipital	79.18	71.61	-7.57
~~*22*~~	~~*Parietal*~~	~~*26.88*~~	~~*12.62*~~	~~*-14.26*~~
23	Temporal	77.05	78.62	**+1.57**
24	Temporal	78.28	77.87	-0.41
25	Central	77.02	75.65	-1.37
26	Temporal	74.36	76.15	**+1.79**
27	Temporal	76.00	78.00	**+2.0**
28	Temporal - Parietal	81.97	83.07	**+1.10**
29	Temporal	75.73	78.41	**+2.68**
Total	-	74.39	75.69	**+1.3**

C. Synthetic Signal Performance

After visual examination and the similarity analysis, we run the experiments according to our evaluation procedure for the 25 remaining patients of the EPILEPSIAE dataset. We highlight here that the patient evaluated at each time corresponds to the patient left-out during the training of the GAN and therefore there is no information leakage between the GAN and the test set. The detailed results of our experiments for each patient are reported in Table II. Again, the reported values are the geometric mean of sensitivity and specificity.

Intuitively, we would expect that training a seizure detection algorithm with synthetic ictal samples would produce a degradation in performance. However, our experiments show that training with synthetic samples yields a 1.3% improvement overall compared to training only with real samples from a generic database. An explanation for the performance improvement when using synthetic data is that, since our GAN generates ictal samples given inter-ictal samples from the same patient, the model generates synthetic seizures that retain a number of personal features. This total difference in performance is calculated as the difference between the geometric mean of the scores of all patients in the synthetic case and the geometric mean of the scores of all patients in the baseline case. These results pass the Wilcoxon statistical significance test with a p-value of 0.011, when Patient 22 is

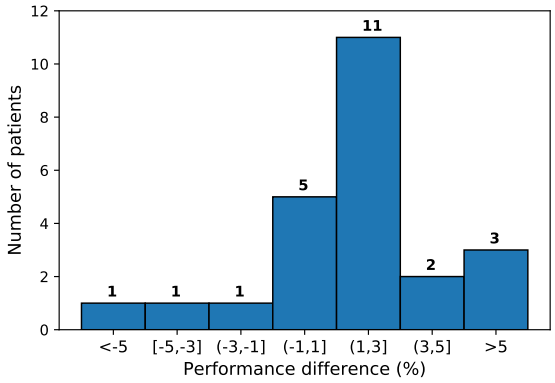


Fig. 6: Performance difference in the classification task between synthetic and real training sets. The vertical axis represents the number of patients and the horizontal one the total difference, where larger is better. As seen, seizure detection for 67% of the patients improves by 1.3%.

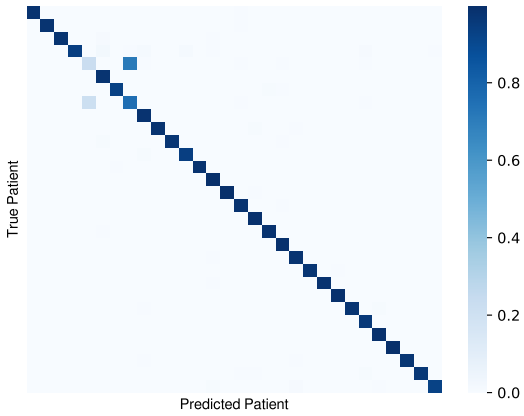


Fig. 7: The normalized confusion matrix for the patient identification among all the synthetic ictal signals. Each row corresponds to one patient and the diagonal elements show the proportion of correctly identified patients.

already excluded[1], which indicates that the difference between the results obtained for the baseline and synthetic training sets is statistically significant. On top of that, as detailed in Fig. 6, 16 out of 24 patients, i.e., 67%, improve by more than 1%, while only for three out of 24 patients the performance decreases by more than 1%.

Regarding the patients for whom the performance degrades most significantly, the seizures from Patient 21 arise in the occipital lobe and are dominated by repetitive spiking. This pattern is relatively rare and is not well represented in the dataset (only 10.5% of the seizures). Therefore, our GAN does not model this behavior with as much precision as it does capture other patterns. Moreover, Patient 20 has only 4 seizures in this dataset, which is the lowest number of seizures in the dataset and hinders robust evaluation of our synthetic data for this patient.

These results demonstrate the high quality of the synthetic ictal data that we generate as well as its utility for the real-world task of seizure detection. In our experiments, we seek to evaluate the performance of our data in a comparable manner, and therefore we do not introduce any prior knowledge in the evaluation. However, in a real-world setting where a patient needs to be monitored, prior knowledge regarding the patterns of the seizures suffered by the target patient can be used to adapt the training of the GAN. This way, EpilepsyGAN could be trained only with those ictal samples that follow the same pattern as the target patient and this would potentially improve the performance of the detection algorithm. We leave this extension of our experiments for future work.

D. Mode Collapse Evaluation

The adversarial training of GANs may fail, resulting in Mode Collapse, i.e., the generator collapses to a few modes or sample types that systematically trick the discriminator [58]. Therefore, if the model suffers from this undesired failure, the variety of the data generated is limited and the input conditioning is ignored.

If our model suffered from mode collapse, the data generated for different patients would converge to the same synthetic samples. Here, we study if mode collapse occurs during the training of our GAN model and if so, to what extent. In this experiment, we use only the *synthetic ictal* signals from all the patients and we randomly split these signals into a training set (70%) and a test set (30%). We consider our CNN architecture and train the model to map each *synthetic ictal* signal to its corresponding patient, i.e., the patient for whom this signal was generated.

The results of this classification task are normalized and presented as a confusion matrix in Fig. 7, with 28 out of 30 patients classified with an accuracy of more than 93.5%. Only two patients present lower accuracy values, 75.9% and 23.3%, while the average accuracy is as high as 95.4%. The results clearly show that the synthetic ictal samples generated for each patient are different and distinguishable, which in turn indicates that mode collapse does not occur across patients.

E. Patient Identification

To evaluate the patient identification task, we use the same five patients for validation as above (Patients 2, 8, 9, 17, and 30). The remaining 25 patients conform the test set, from which we pick different subsets in our experiments. In particular, we pick subsets of $N = 2, 4, 8, 16,$ and 25 patients, randomly chosen from the 25 patients in the test set. On these subsets, we perform the identification task described in Section IV, using all the inter-ictal, real ictal, and synthetic ictal signals in the dataset. The results are shown in Fig. 8. In this figure, the five groups show the results for the experiments with $N = 2, 4, 8, 16,$ and 25 patients in the test set. Each group includes three bars, which represent the identification accuracy for Exp_{orig}, Exp_{synt}, and Exp_{rand}, respectively. We note here

[1]We observe that Patient 22 performs extremely poorly for both the baseline and the synthetic training sets and, therefore, it is not a relevant indicator of the sample quality. This patient suffers from parietal lobe epilepsy, and the mentioned electrodes hardly sense this lobe. Consequently, this patient has been removed from the calculation of the total difference in performance.

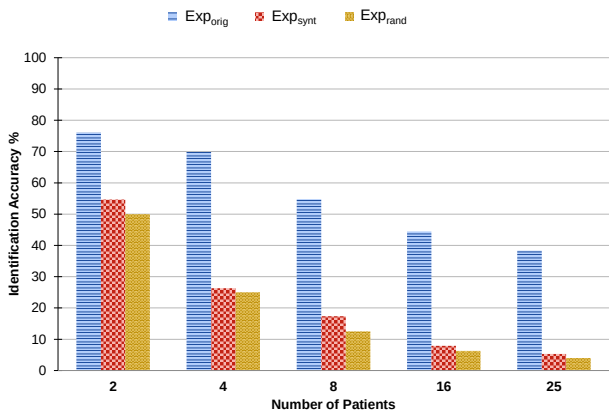


Fig. 8: Identification accuracy results for the experiments with $N = 2, 4, 8, 16,$ and 25 patients in the test set. In each category, the horizontally striped blue bar, checked red bar and dotted yellow bar indicate the results for the patient identification in Exp_{orig} , Exp_{synt} , and Exp_{rand} , respectively.

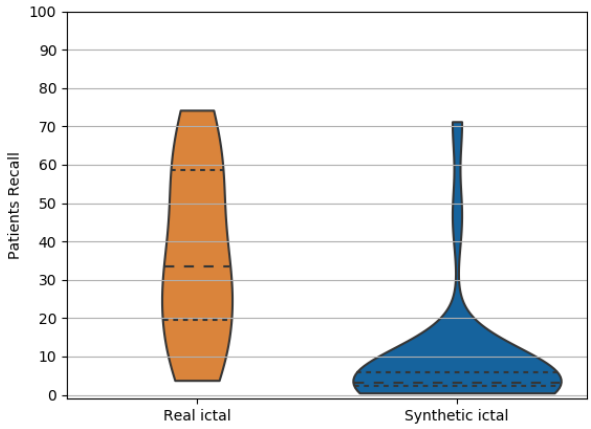


Fig. 9: Patient-specific recall for the experiments with 25 patients in the test set.

that the setup of this experiment, and in particular of Exp_{synt} is fundamentally different from the mode collapse experiment of the previous section: in the mode collpase experiment both training and test set consisted of exclusively synthetic ictal samples, while here the training set consists of real inter-ictal data and the test set of real ictal (Exp_{orig}) and synthetic ictal (Exp_{synt}) data. Therefore, this patient identification experiment investigates whether it is possible to establish an association between inter-ictal and ictal samples (real or synthetic) for a given patient, regardless of the existence or not of mode collapse.

By comparing the results for Exp_{orig} with Exp_{rand} , it is clear that, when we consider the original (real) ictal signals the patients are significantly more vulnerable to be re-identified in comparison to the random baseline. The identifiability ratio, i.e., how many times more identifiable the patients are, between Exp_{orig} (blue bars) and Exp_{rand} (yellow bars) are 1.5x, 2.8x, 4.4x, 7.1x, and 9.6x for $N = 2, 4, 8, 16,$ and 25, respectively.

On the other hand, the results of Exp_{synt} show that patient identification using synthetic data is much harder than using real data. Comparing the results of Exp_{synt} (red bars) and Exp_{rand} (yellow bars), we see that the identifiability ratios are as low as 1.1x, 1.1x, 1.4x, 1.3x, and 1.3x for $N = 2, 4, 8, 16,$ and 25, respectively. Finally, for $N = 25$ the ratio between Exp_{synt} and Exp_{orig} is 7.2x, which indicates that re-identifying a patient from its synthetic data is 7.2 times harder than with real data. These results demonstrate the privacy gains of our generative approach.

F. Patient-Specific Results

Next, we perform an in-depth study of the identification results and investigate the identifiability for every patient individually. Now, for each experiment, Exp_{orig} and Exp_{synt} , all the ictal signals from the 25 patients are fed to the trained model in order to be identified. For each patient i, the number of signals for which the i-th patient is correctly identified is denoted by TP_i. The number of signals for the i-th patient which are incorrectly associated to other patients is denoted by FN_i. Based on these two key values, we calculate the Recall (Sensitivity) metric to produce patient-specific results. The recall metric is defined as follows:

$$\text{Recall}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i} \cdot \tag{3}$$

For every test patient in every experiment we calculate the recall value. Fig. 9 shows the patient-specific recall values when using the real and synthetic ictal signals.

In Fig. 9, each violin shape represents the distribution of recall values for all the patients in the test set. The left-hand side violin, corresponds to the recall values of the experiments with real ictal data. We see that for real ictal data the recall upper quartile is 58.6%. This underlines the fact that a quarter of patients can be identified with a probability of more than 58.6% based only on their real ictal data. Note that if we randomly choose a patient for every input signal, the recall metric would be 4.0%. Furthermore, eight patients can be re-identified with more than 50.0% probability. The maximum for this recall distribution is 74.2%.

On the other hand, the distribution of recall values for the synthetic data is shown in the right-hand side violin in Fig. 9. The median is only 3.2%, which means that with synthetic data the probability of identifying more than half of the patients is the same as random. This demonstrates that the synthetic data generated by the EpilepsyGAN are less prone to privacy vulnerability and re-identification, when compared to the real data.

G. Discussion

The experiments presented in this section show that:

1) Synthetic seizures can be used to train an algorithm for seizure detection, i.e., for discriminating real inter-ictal and real ictal signals.

2) The generative model does not collapse to a single mode, i.e., generates different signals for different patients.
3) Given real inter-ictal signals, it is easier to identify the patient from real ictal samples than from synthetic ictal samples.

These experiments are complementary since 1) determines the discriminability between real ictal and inter-ictal samples given synthetic ictal samples; 2) discriminates synthetic ictal samples across patients; and 3) studies the relationship between inter-ictal and ictal samples for a given patient. The results imply that while synthetic ictal signals are different across patients (which is desired) it is harder to identify the patient from synthetic ictal signals than from real ones, without degradation in the seizure detection task.

VII. Conclusion

In this work we have presented EpilepsyGAN, a generative model that successfully generates synthetic EEG signals of epileptic seizures. This constitutes a direct application of deep generative models in healthcare and a step forward in the treatment of epilepsy. Furthermore, we have proposed an evaluation procedure that allows us to systematically and quantitatively evaluate the quality and utility of the generated synthetic ictal samples. Our results underline that training using exclusively synthetic seizures can achieve comparable performance to training with generic real samples.

To the best of our knowledge, for the first time in the medical domain, we have generated synthetic data sets that can train detection algorithms and reach (and even improve) state-of-the-art results based on real data, which demonstrates the quality of our synthetic data. This solution circumvents both the costs related to seizure recording and labeling, as well as the privacy concerns derived from sharing personal and sensitive data.Although we have shown that our GAN model can generate realistic seizure signals, further improvement on the quality of the signals may require modification of the model and the training strategy. In particular, a variational version of our model working with unpaired data is a promising candidate since it should be able to learn the matching between distributions (inter-ictal and ictal) rather than between samples. We believe that further research into unpaired and conditional deep generative models may advance the field towards personalized medicine based on synthetic data.

Finally, we have shown that it is possible to re-identify patients from their ictal data and demonstrated that using the synthetic signals produced by our proposed GAN model alleviates the privacy concerns associated with sharing data for epileptic seizure detection without a loss in detection performance.

Acknowledgements

This work has been partially supported by the ML-edge Swiss National Science Foundation (NSF) Research project (GA No. 200020182009/1), the MyPreHealth research project (Hasler Foundation project No. 16073), the PEDESITE Swiss NSF Sinergia project (GA No. SCRSII5_193813/1), and the WASP Program funded by the Knut and Alice Wallenberg Foundation.

References

[1] D. Hirtz, D. Thurman, K. Gwinn-Hardy, M. Mohamed, A. Chaudhuri, and R. Zalutsky, "How common are the "common" neurologic disorders?" *Neurology*, vol. 68, no. 5, pp. 326–337, 2007.

[2] H. Blumenfeld, "Impaired consciousness in epilepsy," *The Lancet Neurology*, vol. 11, no. 9, pp. 814–826, 2012.

[3] M. R. Sperling, H. Feldman, J. Kinman, J. D. Liporace, and M. J. O'Connor, "Seizure control and mortality in epilepsy," *Annals of neurology*, vol. 46, no. 1, pp. 45–50, 1999.

[4] T. Tomson, L. Nashef, and P. Ryvlin, "Sudden unexpected death in epilepsy: current knowledge and future directions," *The Lancet Neurology*, vol. 7, no. 11, pp. 1021–1031, 2008.

[5] S. Debener, R. Emkes, M. De Vos, and M. Bleichner, "Unobtrusive ambulatory eeg using a smartphone and flexible printed electrodes around the ear," *Scientific reports*, vol. 5, p. 16743, 2015.

[6] V. Goverdovsky, W. von Rosenberg, T. Nakamura, D. Looney, D. J. Sharp, C. Papavassiliou, M. J. Morrell, and D. P. Mandic, "Hearables: Multimodal physiological in-ear sensing," *Scientific reports*, vol. 7, no. 1, p. 6948, 2017.

[7] D. Sopic, A. Aminifar, and D. Atienza, "e-Glass: A wearable system for real-time detection of epileptic seizures," in *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2018, pp. 1–5.

[8] D. Pascual, A. Aminifar, and D. Atienza, "A self-learning methodology for epileptic seizure detection with minimally-supervised edge labeling," in *2019 Design, Automation & Test in Europe Conference & Exhibition (DATE)*. IEEE, 2019, pp. 764–769.

[9] F. Forooghifar, A. Aminifar, L. Cammoun, I. Wisniewski, C. Ciumas, P. Ryvlin, and D. Atienza, "A self-aware epilepsy monitoring system for real-time epileptic seizure detection," *Mobile Networks and Applications*, pp. 1–14, 2019.

[10] F. Forooghifar, A. Aminifar, and D. Atienza, "Resource-aware distributed epilepsy monitoring using self-awareness from edge to cloud," *IEEE transactions on biomedical circuits and systems*, 2019.

[11] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. A. El-Samie, "Eeg seizure detection and prediction algorithms: a survey," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 183, 2014.

[12] U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, and H. Adeli, "Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals," *Computers in biology and medicine*, vol. 100, pp. 270–278, 2018.

[13] S. Beniczky and P. Ryvlin, "Standards for testing and clinical validation of seizure detection devices," *Epilepsia*, vol. 59, pp. 9–13, 2018.

[14] S. O. Dyke, E. S. Dove, and B. M. Knoppers, "Sharing health-related data: a privacy test?" *NPJ genomic medicine*, vol. 1, p. 16024, 2016.

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680. [Online]. Available: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf

[16] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "Stargan: Unified generative adversarial networks for multi-domain image-to-image translation," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[17] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of GANs for improved quality, stability, and variation," in *International Conference on Learning Representations*, 2018.

[18] L.-C. Yang, S.-Y. Chou, and Y.-H. Yang, "Midinet: A convolutional generative adversarial network for symbolic-domain music generation," in *ISMIR*, 2017.

[19] C.-C. Hsu, H.-T. Hwang, Y.-C. Wu, Y. Tsao, and H.-M. Wang, "Voice conversion from unaligned corpora using variational autoencoding wasserstein generative adversarial networks," in *INTERSPEECH*, 2017.

[20] Z. Yang, W. Chen, F. Wang, and B. Xu, "Improving neural machine translation with conditional sequence generative adversarial nets," *arXiv preprint arXiv:1703.04887*, 2017.

[21] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient," in *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

[22] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, 2018.

[23] H.-C. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. P. Andriole, and M. Michalski, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," in *International Workshop on Simulation and Synthesis in Medical Imaging*. Springer, 2018, pp. 1–11.

[24] P. Costa, A. Galdran, M. I. Meyer, M. Niemeijer, M. Abràmoff, A. M. Mendonça, and A. Campilho, "End-to-end adversarial retinal image synthesis," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 781–791, 2018.

[25] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (medical) time series generation with recurrent conditional gans," *arXiv preprint arXiv:1706.02633*, 2017.

[26] Z. Che, Y. Cheng, S. Zhai, Z. Sun, and Y. Liu, "Boosting deep learning risk prediction with generative adversarial networks for electronic health records," in *2017 IEEE International Conference on Data Mining (ICDM)*. IEEE, 2017, pp. 787–792.

[27] Z. C. Lipton, D. C. Kale, and R. Wetzel, "Modeling missing data in clinical time series with rnns," *Machine Learning for Healthcare*, 2016.

[28] W. N. Price and I. G. Cohen, "Privacy in the age of medical big data," *Nature medicine*, vol. 25, no. 1, p. 37, 2019.

[29] K. G. Hartmann, R. T. Schirrmeister, and T. Ball, "Eeg-gan: Generative adversarial networks for electroencephalograhic (eeg) brain signals," *arXiv preprint arXiv:1806.01875*, 2018.

[30] N. K. N. Aznan, A. Atapour-Abarghouei, S. Bonner, J. Connolly, N. A. Moubayed, and T. Breckon, "Simulating brain signals: Creating synthetic eeg data via neural-based generative models for improved ssvep classification," *arXiv preprint arXiv:1901.07429*, 2019.

[31] I. A. Corley and Y. Huang, "Deep eeg super-resolution: Upsampling eeg spatial resolution with generative adversarial networks," in *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, 2018, pp. 100–103.

[32] L. Xie, K. Lin, S. Wang, F. Wang, and J. Zhou, "Differentially private generative adversarial network," *arXiv preprint arXiv:1802.06739*, 2018.

[33] J. Yoon, J. Jordon, and M. van der Schaar, "PATE-GAN: Generating synthetic data with differential privacy guarantees," in *International Conference on Learning Representations*, 2019.

[34] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.

[35] T. Li and N. Li, "On the tradeoff between privacy and utility in data publishing," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2009, pp. 517–526.

[36] E. Choi, S. Biswal, B. Malin, J. Duke, W. F. Stewart, and J. Sun, "Generating multi-label discrete patient records using generative adversarial networks," *arXiv preprint arXiv:1703.06490*, 2017.

[37] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[38] S. Pascual, A. Bonafonte, and J. Serrà, "Segan: Speech enhancement generative adversarial network," in *INTERSPEECH*, 2017.

[39] S. Pascual, A. Bonafonte, J. Serrà, and J. A. Gonzalez, "Whispered-to-voiced alaryngeal speech conversion with generative adversarial networks," *arXiv preprint arXiv:1808.10687*, 2018.

[40] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[41] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *The IEEE International Conference on Computer Vision (ICCV)*, Oct 2017.

[42] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier nonlinearities improve neural network acoustic models," in *Proc. icml*, vol. 30, no. 1, 2013, p. 3.

[43] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in *6th International Conference on Learning Representations, ICLR 2018, Vancouver, BC, Canada, April 30 - May 3, 2018, Conference Track Proceedings*, 2018.

[44] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training gans," in *Advances in neural information processing systems*, 2016, pp. 2234–2242.

[45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[46] M. Ihle, H. Feldwisch-Drentrup, C. A. Teixeira, A. Witon, B. Schelter, J. Timmer, and A. Schulze-Bonhage, "Epilepsiae–a european epilepsy database," *Computer methods and programs in biomedicine*, vol. 106, no. 3, pp. 127–138, 2012.

[47] C. Hoppe, M. Feldmann, B. Blachut, R. Surges, C. E. Elger, and C. Helmstaedter, "Novel techniques for automated seizure registration: patients' wants and needs," *Epilepsy & Behavior*, vol. 52, pp. 1–7, 2015.

[48] G. H. Klem, H. O. Lüders, H. Jasper, C. Elger *et al.*, "The ten-twenty electrode system of the international federation," *Electroencephalogr Clin Neurophysiol*, vol. 52, no. 3, pp. 3–6, 1999.

[49] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[50] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[51] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.

[53] R. Díaz-Uriarte and S. A. De Andres, "Gene selection and classification of microarray data using random forest," *BMC bioinformatics*, vol. 7, no. 1, p. 3, 2006.

[54] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[55] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series," *Physical review letters*, vol. 88, no. 17, p. 174102, 2002.

[56] P. J. Fleming and J. J. Wallace, "How not to lie with statistics: the correct way to summarize benchmark results," *Communications of the ACM*, vol. 29, no. 3, pp. 218–221, 1986.

[57] I. Osorio, H. Zaveri, M. Frei, and S. Arthurs, *Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics*. CRC Press, 2016.

[58] S. A. Barnett, "Convergence problems with generative adversarial networks (gans)," *arXiv preprint arXiv:1806.11382*, 2018.