A Privacy-Preserving Generative Adversarial Network Method for Securing EEG Brain Signals

Essam Debie, Nour Moustafa, and Monica T. Whitty
School of Engineering and Information Technology. University of New South Wales.
(Canberra)
e.debie@unsw.edu.au, nour.moustafa@unsw.edu.au, monica.whitty@unsw.edu.au

***Abstract*—Generative adversarial networks (GANs) have recently shown high success in applications such as image and time-series classification. However, those applications are vulnerable to complex hacking scenarios, for example, inference and data poisoning attacks, which would alter or infer sensitive information about systems and users. Protecting Electroencephalographic (EEG) brain signals against illegal disclosure has a great interest these days. In this paper, we propose a privacy-preserving GAN method to generate and classify EEG data effectively. Generating EEG data offers a range of capabilities, including sharing experimental data without infringing user privacy, improving machine learning models for brain-computer interface tasks and restore corrupted data. The proposed GAN model is trained under a differential privacy model to enhance the data privacy level by limiting queries of data from artificial trials that could identify the real participants from their EEG signals. The performance of the proposed method was evaluated using a motor imagery classification task, where real EEG data are augmented with artificially generated samples for training machine learning classifiers. The evaluation was performed on a benchmark EEG data set for nine subjects. The experimental outcomes revealed that the non-private version of the proposed approach could produce high-quality data that significantly improve the motor imagery classification performance. The private version showed lower but comparable performance to the standard models trained on real data only.**

I. Introduction

Brain-Computer Interfaces (BCIs) have experienced tremendous growth in various applications, including medicine, education, sports, fitness, and personalised marketing. By BCIs we mean systems that can interpret brain signal patterns and translate them into commands to control external systems such as a mouse cursor or a prosthetic limb.application. BCI applications have also been employed with the Internet of Things (IoT), especially systems of healthcare, to offer automated services to participants [1].

Large volumes of data are often required for training and testing effective machine / deep learning methods in different domains among which BCI applications [2]. Nevertheless, collecting high-quality EEG data from subjects (especially patients and elders) is difficult for various reasons including long calibration time, high subject and session variability, and corrupted data due to several experimental factors [3], [4]. In addition to the data collection challenges, sharing these data poses security risks. Providing illegal access to brain signals (i.e., EEG) or their derived patterns can significantly impair usersâ privacy. This could happen by hacking applications using cyber-attacks such as inference attacks to illegally alter or steal sensitive information about applications and their users. For example, brain spyware was used to infer users' pins, credit card information, and users' location [5]. Creating synthetic data is a potential solution to address these limitations. Generative deep learning approaches have been used to produce realistic data samples that conform to the distribution of real data sets in different domains. For these reasons, there is a need for generating synthetic EEG data that retain the structure and semantics of the real data while preserving the privacy of individuals.

Several studies have used Generative Adversarial Networks (GANs) to create synthetic data in different domains for different purposes including the performance improvement of machine/deep learning models. For example, GANs were used to generate individual participant data in biomedical and healthcare systems [6], [7]. Nonetheless, several linkages and membership inference attacks on biomedical data using machine learning models showed the capability to re-identifying users [8], [9], [10]. In the literature, a few studies have applied GANs to EEG data. In [11], GAN was applied to EEG signals, recorded while looking at images, to regenerate the shown images. The GANs outperformed Variational Auto-Encoders (VAE) models despite the low quality of the generated images. In [12], a GAN model was used to improve the spatial resolution of EEG signals. The approach showed promising results compared to baseline models . In [13], a GAN model was also utilised to explore the similarities between the original EEG and their modified version that was learned by convolutional neural networks. In [14], the authors employed several GAN architectures for generating synthetic EEG samples. While using GANs for EEG is a new research direction, existing studies indicate that it is a promising method for resolving real-world EEG challenges. Nonetheless, protecting the sensitive EEG data against illegal disclosure should be considered while building GAN architectures, as we suggest in this study.

To enhance the data privacy of EEG, a privacy-preserving technique should be applied. Privacy preservation is the process of altering, transforming, or concealing sensitive information of original data collections to prevent them from unauthorised access [15]. Differential privacy [16], as a popular type of privacy preservation, can be denoted as mathematical models used for protecting original data against common privacy attacks, for example, membership inference, homogeneity and data poisoning attacks. Differential privacy guarantees that

978-1-7281-6926-2/20/$31.00 ©2020 IEEE

generated data does not heavily depend on a single study participant. It makes the generation of new plausible individual data while disclosing almost nothing about any study participant [17]. GANs consist mainly of two components: generator; and discriminator components. The discriminator is the only component that has access to real and private data while the generator only receives feedback on the real data through the discriminator's output. Therefore, a GAN model that preserves privacy can be developed by training the discriminator component under a differential privacy scheme.

In this study, we propose a privacy-preserving GAN approach to generate synthetic EEG brain signals that look realistic enough yet does not reveal sensitive characteristics of the original data. Data privacy is achieved by limiting the maximum influence of any single user during the model training and then adding well-designed noise to the model gradients. The proposed approach is evaluated using a benchmark EEG data set for nine subjects.

The remainder of this paper is structured as follows: Section II describes the background and previous studies. The proposed method is explained in Section III. Section IV discusses the experimental design of the study, and this is followed by explaining the results in Section V. Finally, section VI summarises the study and suggests future directions to further this work.

II. Background

A. Generative Adversarial Networks (GANs)

Several deep learning including GAN methods have been explored for feature extraction and classification of EEG data, readers are referred to [18] for a comprehensive review of those approaches. In this section, an overview of the GAN method is presented.

GAN was initially proposed by Goodfellow et al. [19] in their seminal work as an alternative generative approach using adversarial self-play for artificial data generation. Since its invention, GANs have attracted the attention of researchers and shown success as a generative model in several applications [20], [21].

GANs consist of two different neural networks: the generator (G) and the discriminator (D). The idea is based on game theory, whereby two players play to beat each other. In GANs, G is responsible for generating artificial instances, and D is responsible for determining which instance is real and which is artificial. The goal for G is to deceive the discriminator in such a way that the discriminator can no longer precisely differentiate between the actual and the generated samples. This adversarial optimisation problem can be formulated as stated in Equation 1,

$$\min_{G}\max_{D} F(D,G) = E_{X_r \text{`} p_r}\left[log\left(D\left(x_r\right)\right)\right] + E_{X_g \text{`} p_g}\left[log\left(1 - D\left(x_g\right)\right)\right] \quad (1)$$

such that g, d, x_r and x_g are G parameters, D parameters, real samples, and generated samples, respectively. $D(x)$ is the probability of x that belongs to the real or the generated data distributions. The artificial sample x_g is generated by G network from a stochastic noise input z:

$$x_g = G(z) \quad (2)$$

The two networks D and G are trained simultaneously to optimise the performance of D in assigning true labels to both the real and generated samples $log\left(D\left(x\right)\right)$, and minimise $log\left(1 - D\left(x_g\right)\right)$.

B. Differential Privacy

Dwork introduced the concept of ϵ-Differential privacy in his seminal work in 2006 [16]. The basic idea for differential privacy is that any statistical query running on a database should not be heavily dependent on the data of a particular person. Therefore the goal of differential privacy is to offer each person a level of privacy that would result from the removal of their data from a database. A mathematical definition was introduced for the privacy loss associated with any disclosure of data from the data set collected for statistical purposes under the guarantee of confidentiality.

To explain the principle of differential privacy, suppose we have a given data set D, let $\bar{D}$ refer to the data set varying from D by at most one record. Let $Range\left(\mathscr{A}\right)$ be the output range of the random algorithm $\mathscr{A}$. The formal definition of ϵ-Differential privacy is explained as follows:

a) Definition 1: A random algorithm $\mathscr{A}$ is ϵ-differentially private if for a given data set D, any $\bar{D}$, and any subset of outputs $S \subseteq Range\left(\mathscr{A}\right)$, the following holds:

$$Pr\left[\mathscr{A}\left(D\right) \in S\right] \leq e^{\epsilon} Pr\left[\mathscr{A}\left(\bar{D}\right) \in S\right] \quad (3)$$

where ϵ is a privacy budget parameter set by the user to control the privacy level of the algorithm. Smaller privacy budget ϵ implies greater privacy but less accuracy, implying a trade-off between data utility and privacy. The privacy budget is therefore inversely proportional to the accuracy (i.e., data utility) of the result of the query.

Several studies in the literature have investigated the application of differential privacy in deep learning. For example, Phan et al. [22] proposed an adaptive privacy mechanism that can be applied to different deep learning architectures. Abadi et al. [23] applied differentially private stochastic gradient descent (DP-SGD) approach to impose a privacy guarantee during the deep learning training procedure.

III. Proposed Privacy-Preserving Adversarial Network Model

In this section, we introduce the proposed privacy preserving generative adversarial network for EEG.

A. GAN Architecture

The overall architecture of the proposed privacy-preserving adversarial network method for EEG brain signals, is shown in Figure 1. The G network of the proposed GAN consists of one up-sampling and two convolution layers that involves a Tanh activation function and Adam optimiser used respectively for the activation and optimisation. Table I summarises the generator network architecture.

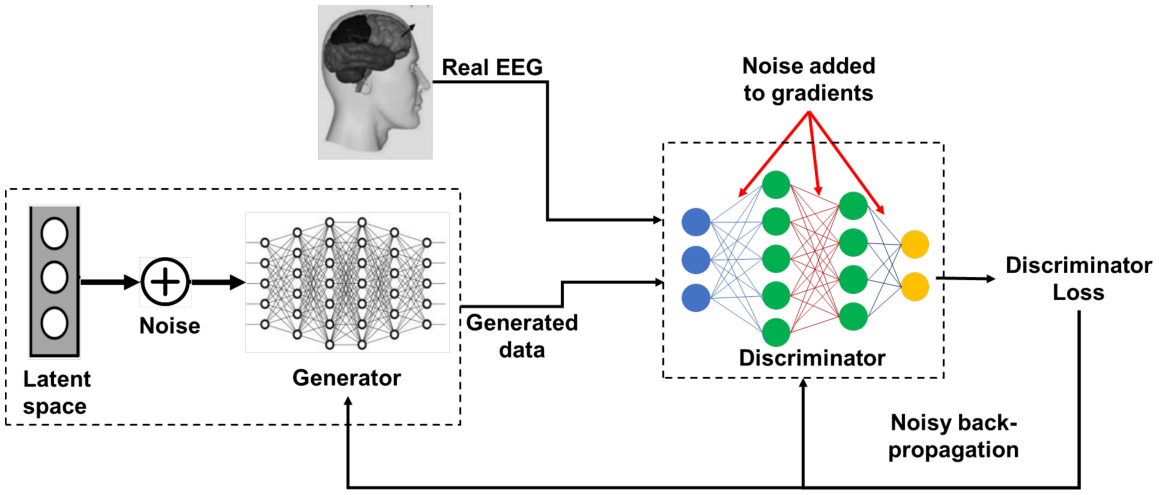


Figure 1: Overview of the proposed privacy-preserving adversarial network method .

Table I: Generator architecture.

Layer	Activation	Output shape
Latent vector	-	100 x 1
Dense	LReLU	256 x 1
Dense	LReLU	512 x 1
Dense	LReLU	1024 x 1
Dense	tanh	5625 x 1
Reshape	-	1875 x 3

Table II: Discriminator architecture.

Layer	Activation	Output shape
Temporal convolution	Linear	8 x 1875 x 32 x 3
Spatial convolution	ReLU	1 x 8 x 1875 x 32 x 3
Max Pool2D	-	1 x 8 x 1875 x 8 x 3
Point-wise convolution	-	8 x 32 x 3
Fully connected	Sigmoid	2

The D network is the GAN component that has access to private data and guides the learning process of the generator through its back-propagation feedback. Therefore, the GAN performance heavily depends on the performance of the discriminator component.

In this study, we employed a convolutional neural network for the discriminator component D that learns spatio-temporal patterns in the underlying data for better classification performance. The convolutional neural network consists of four blocks of processing where each block consists of a number of neural network layers. The first three blocks are used mainly to learn spatio-temporal features from the raw EEG data while the forth block is a traditional fully-connected network used for final data classification. The detailed architecture of each block is as follows:

a) Block 1: A number P of convolutions (kernels) with kernel size $(1, S)$ is applied on raw input data in the first block to learn power spectral density (PSD) features for each EEG channel at different band-pass frequencies, where S is set to half of the sampling rate of the data and the number of convolutions is set as $P = 8$. The outcome from these convolutions is a number of power spectral features corresponding to different frequency bands for each channel used. This corresponds to a number of feature maps that depicts the brain activity at different locations. Batch normalisation [24] is then applied on the resulting data from the convolution process.

b) Block 2: In the second block, spatial patterns are learned for each spectral feature learned in the first block using depth-wise convolution layers [25] followed by batch normalisation. Rectified linear unit (RELU) activation function is then used to add non-linearity to the output of this block followed by a max-pooling layer which is used to reduce the dimensionality of the outputs.

c) Block 3: Separable convolution is applied in the third block to find the best combinations from the spectral and spatial features learned in the previous blocks. Batch normalisation is applied to the outcome of the separable convolution followed by a RELU activation function and max-pooling layer.

d) Block 4: The final block of the discriminator is a fully connected classification neural network that use the learned features in the first three blocks to classify EEG trials. Table II summarises the discriminator network architecture.

B. EEG Privacy Preservation

We employed a differentially private stochastic gradient descent (DP-SGD) approach to impose privacy guarantee during the deep learning training procedure. The basic idea of this approach is to limit the influence of each individual training example on the gradient computations so that the statistical distribution of specific individual can not be learned by the neural network model. Standard optimisation procedures in deep learning use gradients of the underlying loss function to iteratively update the model parameters. DP-SGD works by modifying the gradients used in stochastic gradient descent (SGD) which are used to update the parameters of the

Table III: Privacy parameters.

Parameter	Description
Number of micro-batches	The number of small batches used to perform the clipping. This is used to avoid slow processing when clipping gradients on a per sample basis. By splitting each mini-batch into multiple micro-batches computations can be batched and paralleled. Increasing the number of micro-batches improves the accuracy but slows down training in terms of computational time.
L2 norm clip	The upper bound on the Euclidean norm of each individual gradient computed on individual training samples. It is used to limit the sensitivity of the optimiser to individual training samples.
Noise multiplier	This controls the amount of noise added to the gradients. More noise results in better privacy but lower utility and vice versa.

discriminator network. First, the sensitivity of each gradient is bounded by clipping its computed value on each training sample at each training iteration using a predefined maximum value. Then noise is added to the clipped gradients before computing the parameter updates. The noise is drawn from a Gaussian distribution with a mean 0 and standard deviation proportional to the gradient clip size. Gradient clipping and additive noise statistically make it impossible to compare the SGD updates incorporated in the training with or without a particular data sample to determine whether that sample was included in the training data.

Three privacy specific parameters need to be tuned according to the given data set. Table III summarises these parameter.

Two measures are used to estimate the quality of the differential privacy resulting from DP-SGD process: privacy budget ϵ and privacy guarantee δ. ϵ measures the strength of the privacy guarantee achieved. It gives an upper bound on the probability that the output of a particular model can vary by including or excluding a single training sample. δ bounds the probability that a privacy guarantee does not hold. Traditionally, it is set to be less than the inverse of the size of the training data.

IV. Experimental Design

A. Data set and Preprocessing

The EEG data used in this paper is the Graz data set A (Graz A 2008) [26] for EEG Motor movement/imagery events. The EEG data were recorded from nine healthy subjects performing four different motor imagery tasks (left hand, right hand, foot, tongue). Two sessions of recording were conducted for each subject on different days (the training session and the evaluation session). Each session includes six runs separated by short breaks, where each run is composed of 12 trials per motor imagery task. Therefore, there are 48 trials per run and 288 trials in total per session. Twenty-two Ag/AgCl electrodes were used, and the signals were recorded with a sampling frequency of 250 Hz and bandpass filtered between 0.5 Hz and 100 Hz, and a notch filter at 50 Hz was applied to suppress line noise.

In our experiments, we used recording during the first session for two events only (left hand, and right hand) and three EEG channels (C3, Cz, and C4). These channels have been shown in several studies to capture brain patterns for recognising imaginary movement states [27]. Trials that were marked as bad trials by the original authors were also eliminated from the data set. Thus, the final data of each participant is of shape $136 * 3 * 1875$. The data set was normalised to zero mean and unit variance.

B. Experimental Setup

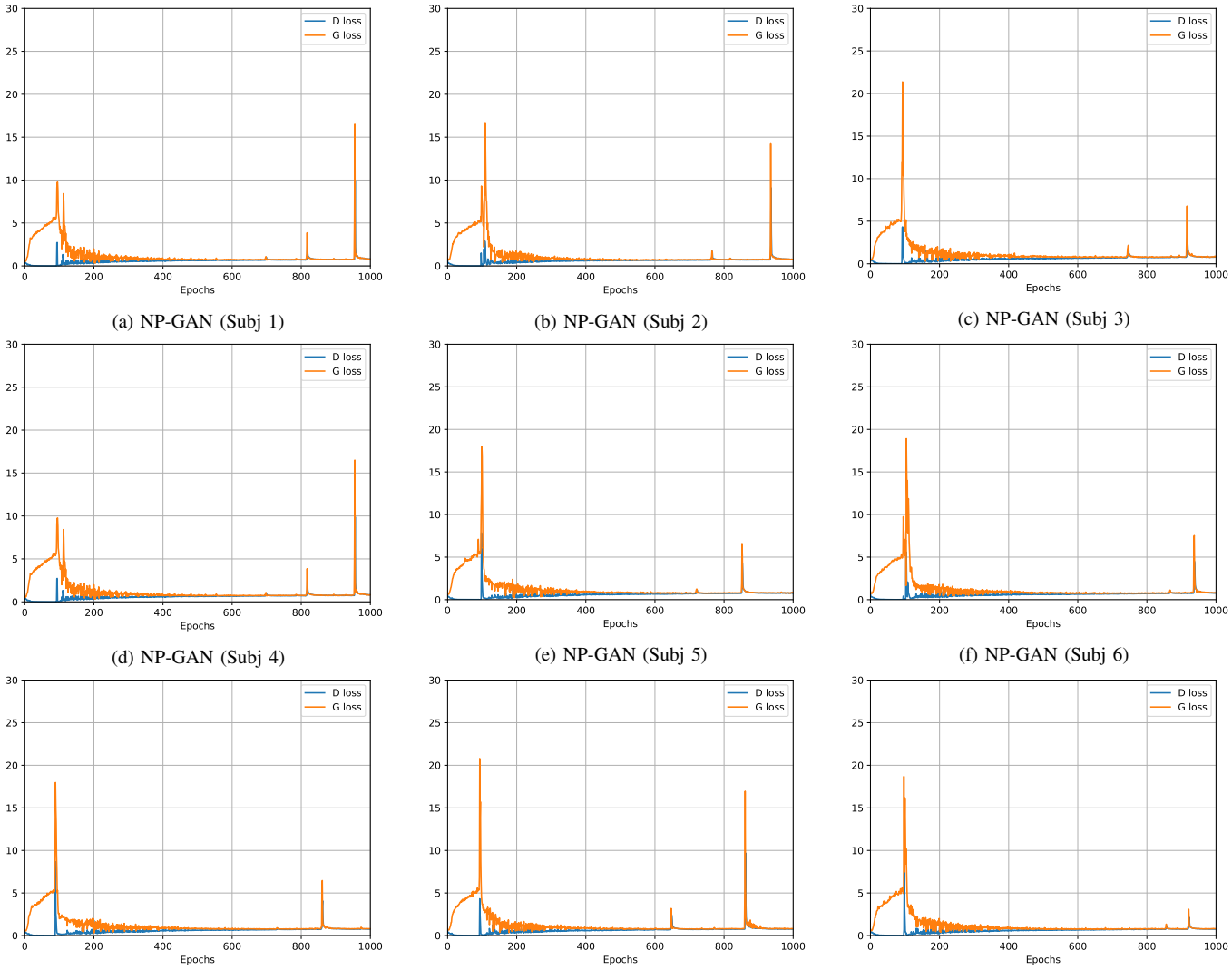
We trained two GANs using the training set to evaluate differential privacy during the generation of artificial EEG data: a non private GAN (referred to as NP-GAN throughout the remainder of this paper) and a private GAN (referred to as PP-GAN throughout the remainder of this paper) trained under differential privacy.

V. Results and Discussions

Figure 2 and Figure 3 show the generator and the discriminator losses over 1000 epochs for the nine participants in NP-GAN and PP-GAN respectively. Results for all subjects have shown similar patterns where the generator loss gradually decreases while the discriminator loss increases until they reach equilibrium (both losses are very close to each other) after approximately 300 epochs in non-private case and 600 epochs in the private case.

A. Sensitivity Analysis of the Noise Multiplier

In this section, we analyse the impact of varying levels of noise multiplier on the GAN performance in terms of quality of the generated data and preserving privacy through the estimation of the privacy budget. To achieve this aim, we trained four different classifiers to distinguish left from right hand movement (2 classes). These classifiers were namely support vector machine (SVM), random forest (RF), linear discriminant analysis (LDA), and logistic regression (LR). Two different sources of data were used for each classifier: real EEG data, and real data augmented by artificial samples generated by the private model (PP-GAN). We use the classification accuracy for each classifier to estimate the quality of the generated data in response to varying noise multiplier. Figure 4 shows the estimate of the privacy budget ϵ in blue calculated at $\delta = 10^{-3}$ which is set to be less than the inverse of the number of training data. Figure 4 also shows the classification accuracy in red using augmented training with $ar = 100$ artificial samples for varying noise multiplier in the range $(0, 2)$. The number of artificial samples ar was set manually after experimenting with varying number of samples in the range of $[50, 200]$. 100 samples was shown to provide the best performance. Therefore, we kept the number of artificial samples in all experiments to $ar = 100$. The choice of the



(a) NP-GAN (Subj 1)

(b) NP-GAN (Subj 2)

(c) NP-GAN (Subj 3)

(d) NP-GAN (Subj 4)

(e) NP-GAN (Subj 5)

(f) NP-GAN (Subj 6)

(g) NP-GAN (Subj 7)

(h) NP-GAN (Subj 8)

(i) NP-GAN (Subj 9)

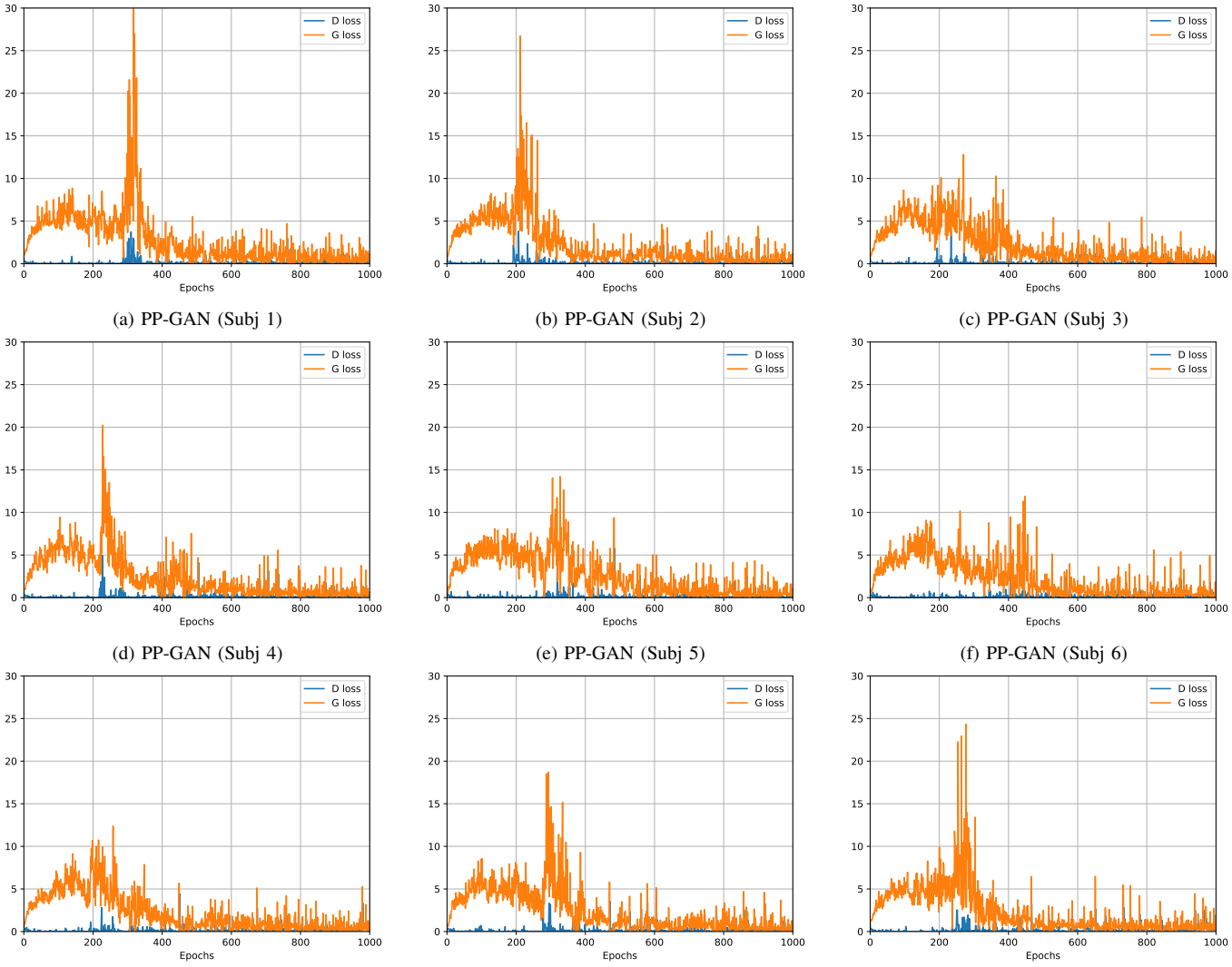
Figure 2: Non-private network loss.

random noise multiplier range $(0, 2)$ was recommended in recent GAN studies [19], [22]. The standard models trained on only real data classifier provide the best baseline classification accuracy. Classifier trained on augmented data with different noise multiplier values achieved worse but comparable accuracy than the baseline due to the added noise as expected. It is observable that setting the noise multiplier to 1.4 or higher causes significant deterioration in classification performance for three classifiers (SVM, RF, and LDA) while for LR the noise multiplier was 1.2.

B. Classification Performance Using Augmented Data

We conducted another experiment to analyse the impact of increasing the ratio of artificial to real data on classification performance. To do this, we trained the four classifiers (SVM, RF, LDA, and LR) to distinguish hand movement on three different sources of data: real EEG data, real data augmented by artificial data generated by the non-private model (NP-GAN), and real data augmented by artificial data generated by the private model (PP-GAN). In this case, results of non-private model provide an approximate upper bound for expected performance when using augmented training. Classification performance using these classifiers is reported over 10 runs where in each run only real data is split into train/test subsets. For augmented models, training data is augmented by artificial samples. To measure the effect of the ratio of artificial data to real data on classification performance we created four augmented training data sets, with 50, 100, 150, and 200 artificial samples respectively. We trained the four classifiers on each of the four augmented data sets in addition to original data set and recorded each model's classification accuracy on the test set. Privacy parameters were set as follows: noise multiplier was set to 1.2, and $\delta = 10^{-3}$.

Table IV summarises the classification performance (average accuracy and standard deviations) for the standard models trained on real data only, models trained on a combination of real data and NP-GAN generated trials. As the results show, augmenting training data with up to 150 artificially generated



(a) PP-GAN (Subj 1)

(b) PP-GAN (Subj 2)

(c) PP-GAN (Subj 3)

(d) PP-GAN (Subj 4)

(e) PP-GAN (Subj 5)

(f) PP-GAN (Subj 6)

(g) PP-GAN (Subj 7)

(h) PP-GAN (Subj 8)

(i) PP-GAN (Subj 9)

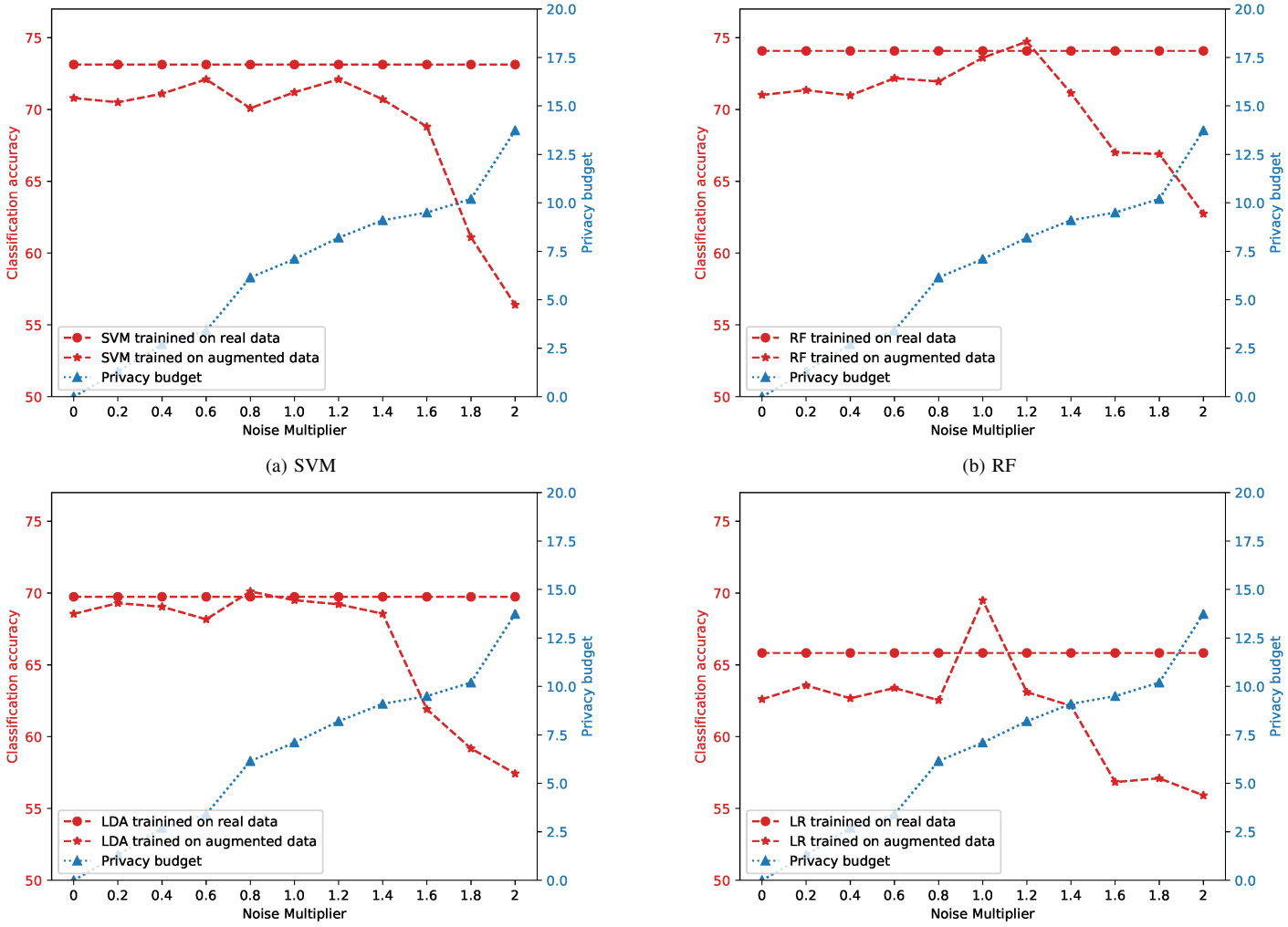
Figure 3: Privacy-preserving network loss.

Table IV: Classification performance for models using NP-GAN .

	Real data	**50 trial augmentation**	**100 trial augmentation**	**150 trial augmentation**	**200 trial augmentation**
SVM	73.13 ± 2.3	74.17 ± 7.1	77.9 ± 1.8	78.87 ± 3.7	71.03 ± 3.3
RF	74.08 ± 1.9	74.10 ± 4.3	76.1 ± 0.8	77.51 ± 1.4	69.51 ± 4.1
LDA	69.74 ± 2.7	71.46 ± 5.2	74.14 ± 1.4	76.13 ± 2.1	70.25 ± 3.9
LR	65.83 ± 2.1	67.14 ± 4.5	74.94 ± 2.5	75.37 ± 3.5	66.19 ± 3.7

Table V: Classification performance for models using PP-GAN with $\epsilon = 0.01$.

	Real data	**50 trial augmentation**	**100 trial augmentation**	**150 trial augmentation**	**200 trial augmentation**
SVM	73.13 ± 2.3	72.08 ± 5.3	72.10 ± 3.5	70.71 ± 4.1	69.14 ± 4.9
RF	74.08 ± 1.9	74.44 ± 2.6	74.73 ± 4.9	71.14 ± 3.1	70.87 ± 5.0
LDA	69.74 ± 2.7	69.55 ± 3.1	69.22 ± 2.9	68.56 ± 4.7	62.33 ± 4.5
LR	65.83 ± 2.1	66.01 ± 5.2	63.17 ± 3.6	62.15 ± 4.4	59.19 ± 7.1



(a) SVM (b) RF

(c) LDA (d) LR

Figure 4: Classification performance and privacy budgets ϵ with varying noise multiplier between 0 to 2. ϵ is calculated at $\delta = 10^{-3}$.

data increased classification accuracy for all three classifiers tested in this experiment. It is observable that augmenting training data with 200 artificial samples performs the worst among the augmentation models. We hypothesise that the addition of too many artificial samples produce noise to the models since the quality of the samples generated by GAN is not optimal. To determine whether the difference between each augmented model (50, 100 and 150 samples) and the baseline model is statistically significant, we applied a pairwise t-test with 95% confidence. The null hypothesis is that there is no significant difference between an augmented model and the baseline model. The statistical results indicated that the null hypothesis was rejected for models with 100, and 150 artificial samples while it was accepted for the augmented models with 50 samples. These results suggest that each augmented model with 100, and 150 artificial samples significantly outperformed their baseline models (trained with real data only) respectively. The results suggest that the ratio of artificial to real data can be tuned to improve the classification performance of different classifiers.

Table V summarises the classification performance (average accuracy and standard deviations) for the standard models trained on real data only, models trained on a combination of real data and PP-GAN generated trials. Classifier trained on augmented data set achieves worse but comparable accuracy than the baseline due to the additive noise. Adding more artificial samples deteriorates the performance of all four classifiers used in the experiment.

VI. Conclusions and Future Work

This paper has introduced a privacy preserving generative adversarial neural network for producing artificial EEG brain signal in BCI applications. The proposed approach addresses data leakage by adding noise sampled from a Gaussian distribution to the gradients of the discriminator network during the parameter update process. Based on the results obtained from the experiment on nine subjects, the proposed approach was able to generate artificial EEG data that possess similar patterns to the real EEG samples. We evaluated the proposed approach in a data augmentation scenario where

machine learning models were trained on real EEG data augmented with artificial samples generated by the proposed approach in both non-private and private modes. The results showed that the non-private mode produced realistic samples that significantly improved the classification perform acne of four classifiers. Data generated by the private version of the approach produces lower but comparable performance due to the added noise. Sensitivity analysis of the noise multiplier parameter showed that a noise multiplier between 1.2 and 1.4 can achieve a good trade-off between data quality and data privacy.

The results reported in this study were restricted to subject-wise and within session contexts. In the future, it is worth investigating an inter-subject analysis which has potential applications for transfer learning. In addition, investigating privacy preserving data generation that depicts session variability is of great importance, particularly in security applications.

VII. Acknowledgement

We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro P6000 GPU used for this research.

References

[1] Seonghun Park, Ho-Seung Cha, and Chang-Hwan Im. Development of an online home appliance control system using augmented reality and an ssvep-based brain–computer interface. *IEEE Access*, 7:163604–163614, 2019.

[2] Essam Debie and Kamran Shafi. Implications of the curse of dimensionality for supervised learning classifier systems: theoretical and empirical analyses. *Pattern Analysis and Applications*, 22(2):519–536, 2019.

[3] Qiqi Zhang and Ying Liu. Improving brain computer interface performance by data augmentation with conditional deep convolutional generative adversarial networks. *arXiv preprint arXiv:1806.07108*, 2018.

[4] Yun Luo and Bao-Liang Lu. Eeg data augmentation for emotion recognition using a conditional wasserstein gan. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 2535–2538. IEEE, 2018.

[5] Ivan Martinovic, Doug Davies, Mario Frank, Daniele Perito, Tomas Ros, and Dawn Song. On the feasibility of side-channel attacks with brain-computer interfaces. In *Presented as part of the 21st ${$USENIX$}$ security symposium (${$USENIX$}$ Security 12)*, pages 143–158, 2012. 00103.

[6] Edward Choi, Siddharth Biswal, Bradley Malin, Jon Duke, Walter F Stewart, and Jimeng Sun. Generating multi-label discrete patient records using generative adversarial networks. *arXiv preprint arXiv:1703.06490*, 2017. 00000.

[7] Cristóbal Esteban, Stephanie L Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional gans. *arXiv preprint arXiv:1706.02633*, 2017. 00000.

[8] Simson L Garfinkel. De-identification of personal information. *National institute of standards and technology*, 2015. 00000.

[9] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart. Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pages 1322–1333. ACM, 2015. 00000.

[10] Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 3–18. IEEE, 2017. 00000.

[11] Isaak Kavasidis, Simone Palazzo, Concetto Spampinato, Daniela Giordano, and Mubarak Shah. Brain2image: Converting brain signals into images. In *Proceedings of the 25th ACM international conference on Multimedia*, pages 1809–1817. ACM, 2017. 00021.

[12] Isaac A. Corley and Yufei Huang. Deep EEG super-resolution: Upsampling EEG spatial resolution with Generative Adversarial Networks. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 100–103. IEEE, 2018. 00009.

[13] Fatemeh Fahimi, Zhuo Zhang, Wooi Boon Goh, Tih-Shih Lee, Kai Keng Ang, and Cuntai Guan. Inter-subject transfer learning with end-to-end deep convolutional neural network for EEG-based BCI. *Journal of neural engineering*, 2018. 00006.

[14] Kay Gregor Hartmann, Robin Tibor Schirrmeister, and Tonio Ball. EEG-GAN: Generative adversarial networks for electroencephalograhic (EEG) brain signals. *arXiv:1806.01875 [cs, eess, q-bio, stat]*, June 2018. 00016 arXiv: 1806.01875.

[15] Marwa Keshk, Benjamin Turnbull, Nour Moustafa, Dinusha Vatsalan, and Kim-Kwang Raymond Choo. A privacy-preserving framework based blockchain and deep learning for protecting smart power networks. *IEEE Transactions on Industrial Informatics*, 2019.

[16] Cynthia Dwork. Differential privacy [G]?? Automata, Languages and Programming. *ser. Lecture Notes in Computer Scienc*, 4052:112, 2006. 00013.

[17] Martin Abadi, Andy Chu, Ian Goodfellow, H Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318. ACM, 2016.

[18] Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. Deep learning for electroencephalogram (eeg) classification tasks: a review. *Journal of neural engineering*, 16(3):031001, 2019.

[19] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 14468.

[20] Emily L. Denton, Soumith Chintala, and Rob Fergus. Deep generative image models using a? laplacian pyramid of adversarial networks. In *Advances in neural information processing systems*, pages 1486–1494, 2015. 01281.

[21] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, and Zhen Wang. Multi-class generative adversarial networks with the L2 loss function. *arXiv preprint arXiv:1611.04076*, 5, 2016. 00102.

[22] NhatHai Phan, Xintao Wu, Han Hu, and Dejing Dou. Adaptive laplace mechanism: Differential privacy preservation in deep learning. In *2017 IEEE International Conference on Data Mining (ICDM)*, pages 385–394. IEEE, 2017.

[23] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 308–318. ACM, 2016.

[24] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.

[25] François Chollet. Xception: Deep learning with depthwise separable convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1251–1258, 2017. 01618.

[26] C Brunner, R Leeb, G Müller-Putz, A Schlögl, and G Pfurtscheller. BCI Competition 2008–Graz data set A. *Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology*, 16, 2008.

[27] Sanqing Hu, Hui Wang, Jianhai Zhang, Wanzeng Kong, and Yu Cao. Causality from cz to c3/c4 or between c3 and c4 revealed by granger causality and new causality during motor imagery. In *2014 International Joint Conference on Neural Networks (IJCNN)*, pages 3178–3185. IEEE, 2014.