Motor Imagery EEG Data Augmentation with cWGAN-GP for Brain-Computer Interfaces

Lucas H. dos Santos [1], **Denis G. Fantinato**[2]

[1]Center for Mathematics, Computing and Cognition (CMCC)
Federal University of ABC (UFABC)
Av. dos Estados - 5001, CEP: 09210-580 – Santo André – SP – Brazil

[2]School of Electrical and Computer Engineering (FEEC)
University of Campinas (UNICAMP)
Av. Albert Einstein - 400, CEP: 13083-852 – Campinas – SP – Brazil

heck.l@ufabc.edu.br,denisf@unicamp.br

Abstract. *Motor imagery is a paradigm in Brain-Computer Interface (BCI) systems based on EEG data. Recently, Deep Neural Networks (DNNs), such as EEGNet, have become a vital component for those systems, overcoming previous state-of-the-art techniques for classifying these data. However, most motor imagery EEG datasets are relatively small, hindering DNNs from achieving better results. In this sense, we propose using Generative Adversarial Networks to augment dataset 1 from the BCI Competition IV for classification efficiency improvement. In addition, we explore augmentation with Gaussian noise for comparison purposes. The experiments were analyzed considering the intra-subject and cross-subject perspectives.*

1. Introduction

Brain-Computer Interface (BCI) system is a device that establishes communication from neural activity to control applications. Since it may be a substitute for traditional input commands, it has several possible applications in emerging technologies for medical, assistive, or entertainment purposes [Nam et al. 2018].

Electroencephalography (EEG) allows monitoring/acquiring neural activity through electrodes positioned on the scalp [Gu et al. 2021]. Frequently, EEG is chosen instead of several other methods due to being a non-invasive, relatively safe, and comfortable technique. A BCI system should be able to process noisy signals in EEG data and generate an output according to a paradigm, such as P300, evoked potentials, or motor imagery. Among them, motor imagery has gained increased attention among researchers for being a very challenging task, which requires the classification of an imagined motor movement (i.e., without any motor execution) within a predefined set of possibilities.

One of the remarkable techniques used in BCI systems for motor imagery EEG data is the Filter Bank Common Spatial Pattern (FBCSP) [Ang et al. 2008], which combines the use of spatial filters and the Common Spatial Pattern algorithm to obtain relevant temporal-spatial features to be classified. The reliable performance achieved by FBCSP in several EEG datasets inspired the design of Deep Learning (DL) structures based on it. In this perspective, we highlight the Shallow Convolutional Neural Network (SCNN) and

the Deep Convolutional Neural Network (DCNN) [Schirrmeister et al. 2017], which allied the temporal-spatial features from FBCSP to the learning flexibility of Convolutional Neural Networks (CNNs). Based on SCNN, a Deep Neural Network called EEGNet has been proposed [Lawhern et al. 2018], including two new layers, the depthwise and pointwise convolution, which allowed an efficient transformation of the feature, using a network with a reduced number of parameters.

The outstanding performance of these DNNs in several motor imagery EEG data, mainly the EEGNet, consolidated its use in BCI systems [Gu et al. 2021]. However, DNNs requires a considerably large amount of data for proper learning during training, which might not be available. In that sense, data augmentation techniques might be used to generate synthetic data. Particularly in the DL context, Generative Adversarial Networks (GANs) have shown promising results in several problems [Aggarwal et al. 2021], generating artificial data as a result of a competition between a generator and a discriminator model. Regarding EEG data, some initial efforts with GANs were made [Lashgari et al. 2020], but there remains a lack considering small (motor imagery) datasets and more sophisticated models of GANs, such as the conditional Wasserstein GAN with gradient penalty (cWGAN-GP) [Mirza and Osindero 2014, Gulrajani et al. 2017].

Based on this, in this work, we propose using cWGAN-GP to perform data augmentation for dataset 1 of BCI Competition IV [Blankertz et al. 2007], which presents considerably small EEG data, with only four subjects and 200 window samples each. SCNN, DCNN, and the EEGNet are used to evaluate the proposed approach. In addition, the results are compared to data augmentation using noise, considering both the intra-subject and cross-subject perspectives.

The organization of the works is as follows. In Section 2, the related works are presented, followed by a description of the methodology used for data processing, classification, and data augmentation using GANs in Section 3. Section 4 present the comparative results and the discussion. Finally, Section 5 concludes the work.

2. Related Works

As shown by Lashgari et al. [Lashgari et al. 2020], many researchers have contributed with approaches to applying GANs to EEG data. These amounted to up to 21% of the papers linking data augmentation to EEG, and only a fraction of those dealt with raw EEG signal, with many proposing mapping to other domains. In [Ko et al. 2019], the authors proposed a semi-supervised GAN with three distinct loss functions. Before training the GAN, the authors used a sliding window to increase the amount of data and, after training, obtained an average increase of 2.2% in accuracy when compared to a Recurrent Neural Network (RNN) classifier. In [Fahimi et al. 2020], the authors proposed a conditional DCGAN, which uses the traditional minimax loss, in a motor imagery dataset and t-SNE as visualization to verify that the GAN converged. They obtained an average increase of 3.6% in accuracy compared to a CNN classifier. In the recent work of [Zhang et al. 2022], the authors proposed an approach very similar to that presented in this work, in which they use a conditional WGAN but applied to P300 detection. They ran comparisons against other data augmentation methods, such as SMOTE and ADASYN, but GAN had better results with an average increase of 9% over the P300 dataset's original baseline.

3. Methodology

In this work, we propose using Generative Adversarial Networks to a Motor Imagery dataset for classification efficiency improvement. We first introduced state-of-the-art CNNs that classify EEG data, aiming to evaluate their performance in the selected dataset without data augmentation. However, these CNNs are known to demand a considerably large dataset for efficient training. Since we consider a relatively small dataset, we propose using a conditional Wasserstein Generative Adversarial Network with gradient penalty (cWGAN-GP) to generate synthetic samples. To evaluate it, we extract the Power Spectral Density of the synthetic data and compare it with that of the original data. Additionally, for comparison purposes, we also consider the noise augmentation method and the baseline results. In the following subsections, we describe the dataset and the proposed approach.

3.1. Dataset and Preprocessing

We use the motor imagery dataset 1 from BCI Competition IV [Blankertz et al. 2007] for all subsequent analyses. This dataset is relevant because it considers some representative imagined actions in a real scenario but with limited complexity. Three behaviors are possible: moving the left hand (LH), right hand (RH), and foot/feet (F). Each subject must only execute two of those (left hand movement is compulsory). One objective of the BCI competition for this dataset was to find out which subjects were artificially created. For our analysis, we remove these and keep only real ones (labeled as A, B, F, and G). The data acquisition occurred over four sessions that generated 50 trials each, summing up to 200 trials per subject. The recording setup used 59 EEG electrodes (most over sensory-motor areas) distributed according to the international 10-20 system. The acquired signals were digitized at 1000 Hz, then a band-pass filter between 0.05 and 200 Hz was applied.

At the competition, the EEG signals were also available down-sampled to 100 Hz, which we chose to use. Our experiments consider an intra-subject (using data from a single subject) and a cross-subject perspective. To validate the results, we used repeated cross-validation, splitting the data into four equal parts, two assigned to compose the training set, one to the validation set, and one to the test set. We performed experiments using 1, 2, 10, and 30 independent simulations but chose two as there was no relevant accuracy variation. We also proposed experiments with different electrodes to find a better (more suitable for GAN training) set, as the 59 electrodes may present exceeding complexity. Based on Park *et al* [Park and Chung 2020], for the 23 electrodes selection, the considered sets and locations are shown in Figure 1.

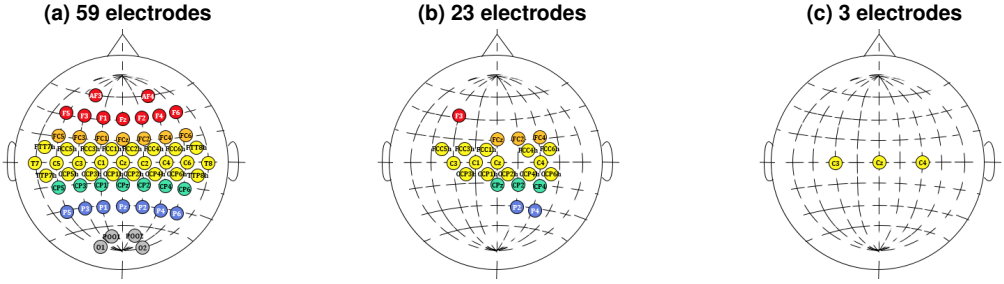

Figure 1. Distinct sets of electrodes chosen to be investigated.

3.2. Classification

Signals acquired with EEG are, generally, represented as a multivariate time series, with $\mathbf{X}_k \in \mathcal{R}^{d \times T}$ describing the trial k, composed of d electrodes (or channels) that lasts for a time period T. Although an EEG signal is similar to images (two-dimensional data), using regular CNNs (e.g., VGG16) is not efficient due to inherent characteristics dissimilarities in EEG data. Classical networks, such as filter-bank CSP, on the other hand, take advantage of those characteristics and, even with lower complexity, can obtain better results. Based on this, Schirrmeister *et al.* [Schirrmeister et al. 2017] proposed two networks: Shallow Convolutional Neural Network (SCNN) and Deep Convolutional Neural Network (DCNN). For their construction, they noted that EEG signals have local and global modulations that evolve with time, requiring two unique convolutions: temporal and spatial. As shown in Figure 2a and Figure 2b, these convolutions are placed in the first two layers and have become a typical strategy in most CNNs related to EEG.

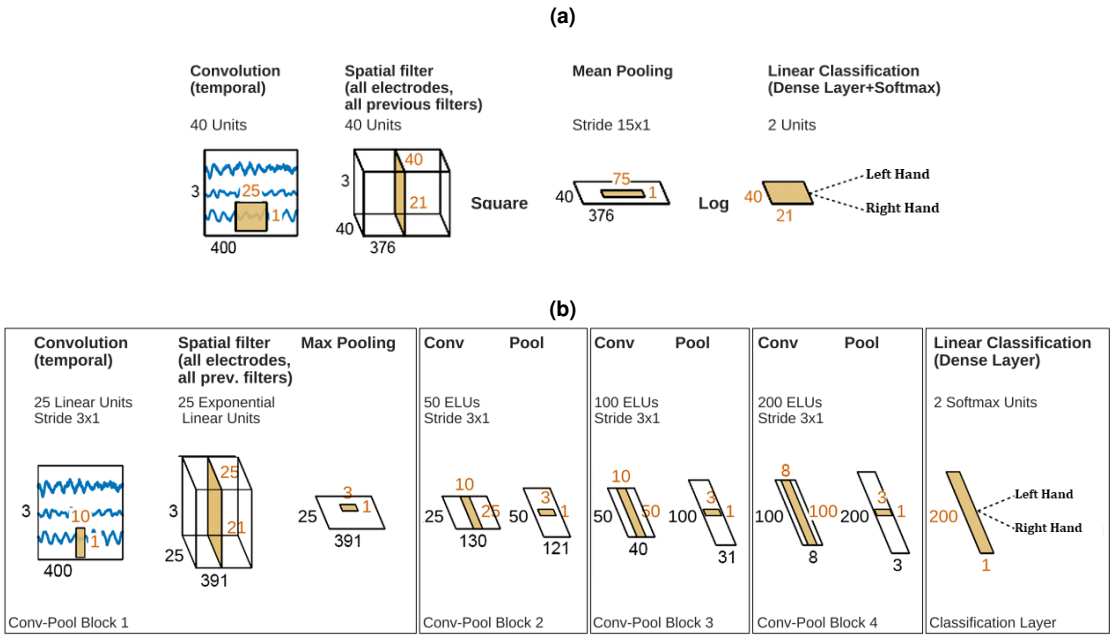

Figure 2. Architecture of the (a) Shallow Convolutional Neural Network and (b) Deep Convolutional Neural Network. Adapted from [Schirrmeister et al. 2017]

The objective of the SCNN is to decode band-power features, a direct influence from FBCSP, as seen from the power and log operations applied after the first two convolutions. While this network is efficient in many datasets, such operations come at the cost that the model is meant to work more efficiently for oscillatory signals. For DCNN, these operations were removed and replaced by regular convolutions at the detriment of additional parameters, which means that a larger dataset is necessary. Lawhern *et al.* [Lawhern et al. 2018] proposed the EEGNet (shown in Figure 3) based on the SCNN, but removing the power and log operation, keeping the temporal and spatial convolutions and also introducing another pair: a depthwise and a pointwise convolution. They allow the model to learn frequency-specific spatial filters while combining them with significant

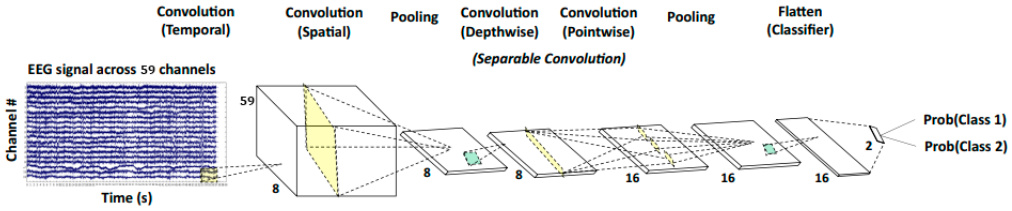

Figure 3. Architecture of the EEGNet. Adapted from [Nahmias et al. 2020]

synergy, achieving better results and increasing computational efficiency, as they require much fewer parameters than regular convolutions.

The classification training, rooted in the work of Lawhern *et al*. [Lawhern et al. 2018], is done in two steps. The first training step, executed with 1500 epochs, has early stopping based on validation accuracy. The second step combines the training and validation sets. The network continues training with this set until the validation loss reaches a value equal to the best training loss from the first step. Optimization happens using Adam with default parameters.

3.3. Data Augmentation with GANs

Generative Adversarial Networks [Goodfellow et al. 2014] are composed of two models, trained alternately, that compete against one another. One is the generator G, whose objective is to match an artificial distribution ($p_{\mathcal{Z}}$) to a real one ($p_{\mathcal{X}}$). The second is the discriminator, D, that receives samples from both distributions and has to predict whether it belongs to the original data, X, or not. The generator, G, does not require prior knowledge on $p_{\mathcal{X}}$ to produce new samples, unlike most other generative models, instead, it is induced indirectly to it based on the output of the discriminator.

Goodfellow *et al*. [Goodfellow et al. 2014] defined, for what is called the classical GAN, an objective function based on the binary cross-entropy. There are two conflicting objects to be achieved: identifying the real sample as 1 and the synthetic as 0. To improve results, many authors have proposed other functions. For instance, the Wasserstein GAN (WGAN), proposed by Arjovsky *et al*. [Arjovsky et al. 2017], has a very distinct objective and has shown exceptional results with relatively low fine-tuning, with fast convergence in many datasets. The authors hypothesize that the use of f-divergences (which includes the classical GAN) as a metric to approximate $p_{\mathcal{Z}}$ to $p_{\mathcal{X}}$ is not a promising strategy. Instead, they propose using Earth-Moving (or Wasserstein-1) distance, given that it allows the use of a gradient optimization method, independent of the distance between the distributions. Gulrajani *et al*. [Gulrajani et al. 2017] further improves the WGAN, proposing a penalty over the norm of the gradient, that better satisfies an internal condition of Wasserstein-1 distance. The object function for the WGAN-GP is defined as:

$$\min_{G} \max_{D \in \mathcal{D}} \mathbb{E}_{\mathbf{x} \sim p_{\mathcal{X}}} [D(x)] - \mathbb{E}_{\mathbf{z} \sim p_{\mathcal{Z}}} [D(G(\mathbf{z}))] - \lambda \mathbb{E}_{\hat{x} \sim p_{\mathcal{G}}} \left[(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2 \right], \quad (1)$$

where $\hat{x}$ is a sample from distribution $p_{\mathcal{G}}$, corresponding to a linear distribution interpolated between $p_{\mathcal{X}}$ and $p_{\mathcal{Z}}$, and λ is the penalty coefficient.

Another topic of importance in GANs is the architecture of the models. If models of high complexity are chosen, then it is relatively easy to guarantee that the discriminator will converge, on the other hand, this will probably mean the opposite for the

generator. Once the generator fails to converge, it tends to search for easier paths, leading to issues such as mode collapse, in which it will generate samples from few or single class, and may even generate a single sample. The conditional GAN, proposed by Mirza *et al.* [Mirza and Osindero 2014], aims to solve this issue by adding a class conditional perspective to the objective function in the form of a new variable as input into both models. In fact, Zheng *et al.* [Zheng et al. 2020] demonstrate that, for theirs datasets, using conditional on the WGAN-GP resulted in increased quality of synthetic data over the non-conditional one. For the architecture, it is suggested to start building on from DCGAN [Radford et al. 2015] and then try adding or reducing the number of layers – although a very empirical approach, it is the most common one, as there is not a fixed rule that works for all cases.

From this idea, we reached the architecture shown in Table 1. We chose EEGNet as the discriminator due to its efficiency in classifying the data, since it obtained high accuracy in the dataset of interest, and it should mean that it was able to extract relevant features (which should also be in the synthetic samples). GAN training is similar to classifier training, with 250 epochs for the training set, after which the training and validation sets were combined and trained for 50 additional epochs. The power spectral density of the synthetic signals was calculated every 50 epochs for visual inspection of the quality of samples during training. Synthetically augmented datasets were constructed once the GAN training was concluded, with proportions of 1:1 (equally amounts of real and synthetic samples), 1:2 (one real to two synthetic samples), 1:3, and 1:5. These sets were used to train a classifier and compared against each other and the baselines.

Table 1. GAN architecture.

Generator		**Discriminator (EEGNet)**	
Layer	**Output**	**Layer**	**Output**
Input	(16)	Input	(3, 400, 1)
Linear	(512)	Conv and BNorm	(16, 3, 401)
Linear and BNorm	(1024)	Conv and BNorm	(64, 1, 400
Upsample and BNorm.	(64, 1, 50)	Mean Pooling	(64, 1, 100)
Conv and BNorm	(64, 3, 50)	Conv	(64, 1, 101)
TConv and BNorm	(64, 3, 100)	Conv and BNorm	(64, 1, 101)
Upsample and BNorm	(64, 3, 200)	Mean Pooling	(64, 1, 12)
Conv and BNorm	(128, 3, 200)	Flatten	(768)
TConv and BNorm	(64, 3, 400)	Linear	(2)
Conv	(1, 3, 400)	LogSoftmax	(2)

BNorm: Batch Norm
Conv: Convolution
TConv: Transposed Convolution

3.4. Data Augmentation with Gaussian Noise

Data augmentation with generative models falls into the same issue as when classifying, i.e., if the dataset is small, then the many parameters tend to overfit the data, causing poor generalization. In that sense, there are many other techniques not based on neural

networks for data augmentation, with augmentation by Gaussian noise being one of them. Although very simple, it has been proven effective in many works centered around EEG data [Wang et al. 2018, Pei et al. 2021]. Its application can be done sample by sample with

$$\tilde{\mathbf{x}} = \mathbf{x} + \mathcal{N}\left(\mu, \sigma^2\right), \tag{2}$$

where μ and σ are, respectively, the mean and standard deviation.

Training with Gaussian noise meant constructing synthetically augmented distributions, as done in the GAN training. Hence, for comparison purposes, we shall include this method in the analysis. In the first experiments, we chose $\mu = 0$ and $\sigma = 0.2$. In the end we show additional experiments for a wider range of signal-to-noise ratios by varying σ from 0.02 to 0.5.

3.5. Performance Evaluation of Data Augmentation

The Fréchet Inception Distance (FID) [Heusel et al. 2017] is used to evaluate how similar the synthetic distribution generated by the GAN is to the real one. It became the norm when comparing different GANs. However, this metric tends to be uncertain for data that differs from natural images. A promising approach is to use a specific network trained with EEG data. However, this approach can not be used straightforwardly due to the high variability in EEG signals. Visual inspection of raw EEG data is also possible, but it requires specialized personnel to evaluate the quality of the synthetic data. Hence, as an alternative, we propose using Power Spectral Density (PSD) to assess the quality of the GAN training. Since motor imagery activation occurs predominantly in the μ and α bands (8–13 Hz), both distributions' PSD should be similar or, at least, predominantly occurring in these bands. Note that PSD may determine how similar the original and synthetic data are (which can be related to the quality of the samples) but do not provide any insight into diversity.

4. Experimental Results and Discussion

Experimental results are divided into the selection of the classifier, defining a performance baseline, and evaluation of the data augmentation approach, as presented below.

4.1. Baseline Classification

Table 2 1 shows the results obtained with the optimized hyper-parameters of SCNN, DCNN, and EEGNet classifiers. The DCNN could not provide results as good as the other networks due to the dataset's small size. EEGNet and SCNN got very similar results, with the difference coming from subject B data, which, from our analysis, has shown the lowest accuracy in all scenarios. The performance from subjects F and G data were very close on both nets, and subject A being classified with the same accuracy. When moving to cross-subject, SCNN has a marginal improvement over EEGNet. Even with this, we still chose EEGNet for the rest of the experiment, as the mean intra-subject accuracy was slightly better.

Table 3 shows the accuracy obtained with the three proposed sets of electrodes. The obtained results for 23 electrodes led to the best results. It also indicates that many electrodes may not necessarily mean better performance, even more since some electrodes are in regions on the scalp where motor imagery has little influence. With three electrodes,

Table 2. Accuracy (%) for the test set, with mean and standard deviation.

Network	A	B	F	G	Mean	Cross-Subject
EEGNet	79.33 ± 5.56	62.08 ± 8.50	73.08 ± 7.45	87.92 ± 3.74	75.60	68.18 ± 3.55
SCNN	79.17 ± 4.97	56.58 ± 6.68	75.83 ± 7.00	85.75 ± 4.98	74.33	68.74 ± 2.94
DCNN	63.83 ± 6.15	51.67 ± 4.85	56.75 ± 5.95	84.75 ± 4.48	64.25	58.83 ± 4.77

both the mean and cross-subject accuracy degraded. Yet, the results are still sharp, i.e., for subjects A, B, and G, the performance loss is subtle, especially for G, whose results are almost equal to that with 59 electrodes. The computational complexity with only three electrodes is also lower since this amount is equivalent to less than 15% of the 23 and only 5% of the 59. Thus, this may aid the GAN training. For this reason, we chose to keep the three electrodes set for the rest of the experiments.

Table 3. Accuracy (%) with EEGNet for distinct sets of electrodes.

# Electrodes	A	B	F	G	Mean	Cross-Subject
59	83.00	56.67	**86.25**	88.17	78.52	79.18
23	**83.52**	**68.83**	84.67	**90.58**	**81.90**	**83.32**
3	79.33	62.08	73.08	87.92	75.60	68.18

4.2. Data Augmentation

Figure 4 shows the PSD obtained for subject G data using the proposed GAN architecture (Table 1) after training. These results are important evidence that GANs present the capability of generating samples similar to the real ones. For the left-hand data PSD (Figure 4a), there is a difference with synthetic samples, with half the expected power (0.2 against 0.4 of the expected), but not for the right hand (Figure 4b). Nonetheless, the main characteristic, the frequency bands, are similar in both cases. These figures are also relevant because they allow the classification of the imagined movement to be clearly recognizable, which explains why subject G had such good results in either of the three CNNs.

Once we had some evidence that the GAN had been trained successfully (i.e., verified that the PSD of the real and synthetic data – conditionally to the class – are close, we stopped the training of the GAN), we moved on to the classification step. Table 4 displays the results obtained with both GAN and noise augmentation under an intra-subject perspective. For subject A data, neither of the approaches could improve the results, with GAN severely degrading the accuracy. For subject B data, the results are opposite with both improving, but noise augmentation showed to be the most relevant, raising the accuracy by almost 4%. Subject F data had a similar result to subject A data for GAN, and noise had no significant role. A marginal improvement was observed for subject G data, with noise augmentation improving by almost 1%.

For the cross-subject experiments, there are three classes, turning it into a more complex problem, but with more input samples since data from all subjects are gathered together, so GAN is expected to have better efficiency. The obtained results are shown in

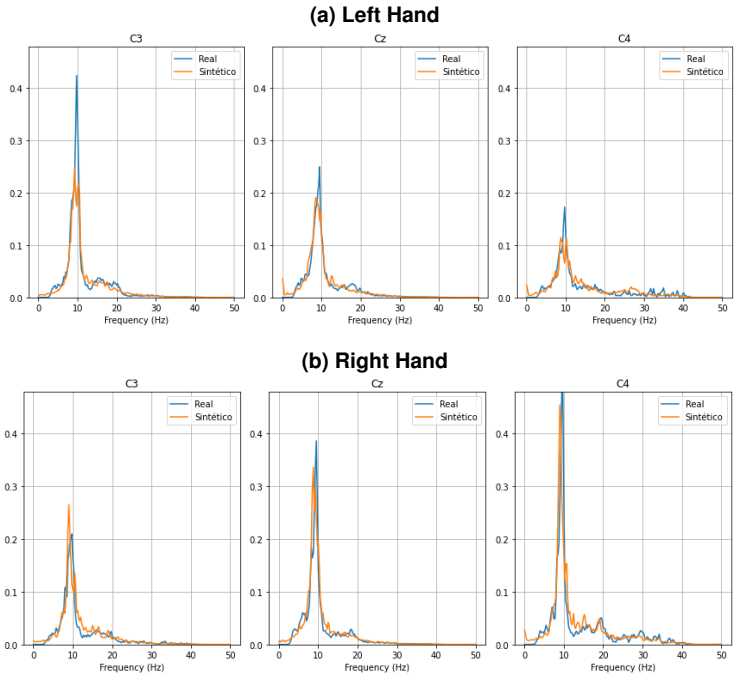

Figure 4. Power spectral density from synthetic samples with 3 electrodes for subject G.

Table 4. Accuracy (%) with data augmentation in intrasubject classification.

Subject	Augmentation	Baseline	1:1	1:2	1:3	1:5
A	GAN	**79.33**	75.25	73.58	75.58	72.33
	Noise($\sigma = 0.2$)	**79.33**	78.67	78.58	78.33	76.92
B	GAN	62.08	63.42	62.50	63.83	**63.92**
	Noise($\sigma = 0.2$)	62.08	62.75	65.42	65.65	**65.75**
F	GAN	**73.08**	67.45	66.54	67.09	66.63
	Noise($\sigma = 0.2$)	73.08	73.17	72.67	72.83	**73.20**
G	GAN	87.92	**88.50**	87.75	88.42	88.33
	Noise($\sigma = 0.2$)	87.92	88.00	**88.83**	88.33	88.75

Table 5. As expected, both techniques show considerable efficiency. The GAN improves the accuracy by 2% and noise improves by 4%. With the increase in synthetic samples proportion, there is a negative impact using GAN and a positive using noise.

Table 5. Accuracy (%) with data augmentation in cross-subject classification.

Augmentation	Baseline	1:1	1:2	1:3	1:5
GAN	68.18	**70.42**	70.25	69.02	69.21
Noise($\sigma = 0.2$)	68.18	71.03	71.86	71.97	**72.71**

As the previous experiments demonstrated, noise augmentation is an efficient method. To gain a better insight into it, we conducted additional experiments with increased signal-to-noise ratios, which resulted in Table 6. Subject B data was chosen, as it had a higher positive change in accuracy. These results prove, once again, that this method is capable of increasing the accuracy, even for a large level of noise ($\sigma = 0.5$). Nonetheless, the most prominent increase was obtained with the lowest standard deviation and largest synthetic-real ratio.

Table 6. Accuracy (%) achieved for subject B (62.08%) with noise augmentation.

σ (SNR [db])	1:1	1:2	1:3	1:5	Mean Gain
0.02 (34.0)	65.50	66.42	66.08	**67.25**	**+4.23**
0.1 (20.0)	62.92	66.67	**66.92**	65.50	+3.42
0.2 (14.0)	62.83	65.33	**66.08**	65.75	+2.92
0.3 (10.4)	64.33	63.67	**66.75**	66.50	+3.23
0.5 (6.0)	60.50	64.33	64.00	**65.42**	+1.48

We applied the noise augmentation for all subjects and cross-subject with the best σ and proportion observed for subject B data. The results in Table 7 show that $\sigma = 0.2$ and $\sigma = 0.02$ contribute similarly since we had subject B data as a reference. Both increase the intra-subject and cross-subject mean accuracy by a similar percentage.

Table 7. Accuracy (%) achieved with 0.2 and 0.02 standard deviation and 1:5 proportion.

σ	A	B	F	G	Mean	Cross-Subject
Baseline	**79.33**	62.08	73.08	87.92	75.60	68.18
0.02	77.42	**67.25**	72.08	**89.50**	**76.56**	**72.94**
0.2	76.92	65.75	**73.20**	88.75	76.15	72.71

5. Conclusion

In this paper, we proposed a data augmentation approach with GANs, focusing on a small dataset of motor imagery EEG signals. Extracting the power spectral density from samples generated by the cWGAN-GP proved clear evidence of convergence, with changes

of activity in similar frequencies to those of the original distribution. The use of GAN to augment the number of samples from the BCI competition IV dataset 1 has shown positive results for the intra-subject perspective for two out of four subjects, increasing the classification accuracy in a higher amount in the cross-subject scenario. The small number of training samples has possibly limited the potential of the GAN, surpassed by the noise augmentation technique. Further research should explore data augmentation, with techniques such as noise and sliding window, to aid the GAN training, which still depends on a relatively large number of samples.

6. Acknowledgements

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, and grant #2020/10014-2, São Paulo Research Foundation (FAPESP).

References

Aggarwal, A., Mittal, M., and Battineni, G. (2021). Generative adversarial network: An overview of theory and applications. *International Journal of Information Management Data Insights*, 1(1):100004.

Ang, K. K., Chin, Z. Y., Zhang, H., and Guan, C. (2008). Filter bank common spatial pattern (fbcsp) in brain-computer interface. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pages 2390–2397. IEEE.

Arjovsky, M., Chintala, S., and Bottou, L. (2017). Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR.

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., and Curio, G. (2007). The non-invasive berlin brain–computer interface: fast acquisition of effective performance in untrained subjects. *NeuroImage*, 37(2):539–550.

Fahimi, F., Dosen, S., Ang, K. K., Mrachacz-Kersting, N., and Guan, C. (2020). Generative adversarial networks-based data augmentation for brain–computer interface. *IEEE transactions on neural networks and learning systems*, 32(9):4039–4051.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

Gu, X., Cao, Z., Jolfaei, A., Xu, P., Wu, D., Jung, T.-P., and Lin, C.-T. (2021). Eeg-based brain-computer interfaces (bcis): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(5):1645–1666.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., and Courville, A. (2017). Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028*.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30.

Ko, W., Jeon, E., Lee, J., and Suk, H.-I. (2019). Semi-supervised deep adversarial learning for brain-computer interface. In *2019 7th international winter conference on brain-computer interface (BCI)*, pages 1–4. IEEE.

Lashgari, E., Liang, D., and Maoz, U. (2020). Data augmentation for deep-learning-based electroencephalography. *Journal of Neuroscience Methods*, page 108885.

Lawhern, V. J., Solon, A. J., Waytowich, N. R., Gordon, S. M., Hung, C. P., and Lance, B. J. (2018). Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013.

Mirza, M. and Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Nahmias, D. O., Civillico, E. F., and Kontson, K. L. (2020). Deep learning and feature based medication classifications from eeg in a large clinical data set. *Scientific Reports*, 10(1):1–11.

Nam, C. S., Nijholt, A., and Lotte, F. (2018). *Brain–Computer Interfaces Handbook: Technological and Theoretical Advances*. CRC Press.

Park, Y. and Chung, W. (2020). Optimal channel selection using correlation coefficient for csp based eeg classification. *IEEE Access*, 8:111514–111521.

Pei, Y., Luo, Z., Yan, Y., Yan, H., Jiang, J., Li, W., Xie, L., and Yin, E. (2021). Data augmentation: Using channel-level recombination to improve classification performance for motor imagery eeg. *Frontiers in Human Neuroscience*, 15:645952.

Radford, A., Metz, L., and Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., and Ball, T. (2017). Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420.

Wang, F., Zhong, S.-h., Peng, J., Jiang, J., and Liu, Y. (2018). Data augmentation for eeg-based emotion recognition with deep convolutional neural networks. In *International conference on multimedia modeling*, pages 82–93. Springer.

Zhang, R., Zeng, Y., Tong, L., Shu, J., Lu, R., Yang, K., Li, Z., and Yan, B. (2022). Erp-wgan: A data augmentation method for eeg single-trial detection. *Journal of Neuroscience Methods*, 376:109621.

Zheng, M., Li, T., Zhu, R., Tang, Y., Tang, M., Lin, L., and Ma, Z. (2020). Conditional wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification. *Information Sciences*, 512:1009–1023.