Title
Augmenting EEG with Generative Adversarial Networks Enhances Brain Decoding Across Classifiers and Sample Sizes

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors
Williams, Chad C
Weinhardt, Daniel
Wirzberger, Maria
et al.

Publication Date
2023

Peer reviewed

Augmenting EEG with Generative Adversarial Networks Enhances Brain Decoding Across Classifiers and Sample Sizes

Chad C. Williams*[1], Daniel Weinhardt*[2], Maria Wirzberger[2], & Sebastian Musslick[1,3,4]

*Co-First Author
[1]Carney Institute for Brain Science, Brown University
[2]University of Stuttgart, Interchange Forum for Reflecting on Intelligent Systems
[3]Department of Cognitive, Linguistic, and Psychological Sciences, Brown University
[4]Institute of Cognitive Science, Osnabrück University
Corresponding Authors: Chad C. Williams, chad_williams@brown.edu; Daniel Weinhardt, daniel.weinhardt@ife.uni-stuttgart.de

Abstract

There is major potential for using electroencephalography (EEG) in brain decoding that has been untapped due to the need for large amounts of data. Advances in machine learning have mitigated this need through data augmentation techniques, such as Generative Adversarial Networks (GANs). Here, we gauged the extent to which GANs can augment EEG data to enhance classification performance. Our objectives were to determine which classifiers benefit from GAN-augmented EEG and to estimate the impact of sample sizes on GAN-enhancements. We investigated three classifiers—neural networks, support vector machines, and logistic regressions—across seven sample sizes ranging from 5 to 100 participants. GAN-augmented EEG enhanced classification for neural networks and support vector machines, but not logistic regressions. Further, GAN-enhancements diminished as sample sizes increased—suggesting it is most effective with small samples, which may facilitate research that is unable to collect large amounts of data.

Keywords: EEG, GAN, Data Augmentation, Neural Networks, Support Vector Machine, Logistic Regression

Introduction

Electroencephalography (EEG)—the scalp recording of electrical activity of the brain—has played a prominent role in neuroscience as a leading non-invasive method for gaining insights into human brain functioning. EEG has been used to index healthy brain functioning, for example, when defining episodes of epilepsy or diagnosing depression. However, the application of EEG in classification has been limited due to its low signal-to-noise ratio, and thus the need for large amounts of data (Lotte et al., 2018). In domains of machine learning, (e.g., image classification), the need for large amounts of data has been mitigated by data augmentation techniques such as Generative Adversarial Networks (GANs; Goodfellow et al., 2014). However, GANs have only been recently developed and have scarcely been applied to EEG data. Here, we introduce a transformer-based GAN architecture for augmenting single-trial EEG data. We investigated the efficiency of this architecture for augmenting EEG data to enhance the classification of brain states. Specifically, we examined the success of this approach across classifiers and sample sizes.

GANs are machine learning frameworks that consist of two adversarial neural network agents, namely the generator and the discriminator. The generator is trained to create novel samples that are indiscernible from real samples. In the current context, the generator produces realistic continuous EEG activity, conditioned on a set of experimental variables, which contain underlying neural features representative of the outcomes being classified. For example, depression manifests as increased alpha oscillatory activity in the EEG signal (Koo et al., 2019), and thus, an ideal generator would produce continuous EEG that includes these alpha signatures. In contrast to the generator, the discriminator determines whether a given sample is real or synthetically produced by the generator. The core insight of GANs is that the generator can effectively learn from the discriminator. Specifically, the generator will consecutively produce more realistic synthetic samples with the goal of "fooling" the discriminator into believing them as real. Once it has achieved realistic samples that the discriminator cannot discern, it can be used to generate synthetic data—or in this context, synthetic EEG data.

GANs have been shown to produce realistic EEG data, opening the door to many applications (Fahimi et al., 2019; Hartmann et al., 2018)—e.g., building clinical diagnostic tools (Brophy et al. (2021)) and enhancing brain-computer interfaces (Lotte et al. (2018)). Alongside the validation of GANs as a synthetic EEG generator, researchers explored its utility for enhancing classification performance. Indeed, research has found that GANs could be used to augment the performance of classifiers (Fahimi et al., 2020; Kan et al., 2021; Luo and Lu, 2018; Petruţiu et al., 2020; see also Dong and Ren, 2020; Hwang et al., 2019; Panwar et al., 2020 for GANs as classifiers). Augmenting EEG was shown to enhance classification performance up to 20% (Luo and Lu, 2018); however, this line of research is new and there are many open questions concerning the breadth of its applicability. For example, it is still unknown which classifiers benefit from GAN-augmented EEG and whether enhanced performance depends on the sample sizes being used.

In the current study, we investigated the extent to which GANs can augment EEG data and enhance classification. We used an EEG dataset of 500 participants who performed a simple gambling task, and classified their experiences of winning and losing gambles. We focused on two primary research questions: 1) for which classifiers is GAN-enhanced classification effective, and 2) how does the impact on enhanced classifications vary with sample size? For the first research question, we assessed three classifier architectures—neural networks, support vector machines (SVM), and logistic regressions—across two data formats—the full time series, and extracted features, such as reward positivity ampli-

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

tude, delta power, and theta power. For the second research question, we assessed classification across a range of sample sizes from 5 to 100 participants. Our results indicated that GANs enhanced classification for both neural networks and SVMs, but not logistic regressions. Further, we found that this pattern persisted for both the full time series set and the extracted features set. Critically, we found that the enhancement to classification performance with GANs decreased across sample sizes, with the largest benefit occurring in smaller sample sizes. Altogether, these findings demonstrated the power of augmenting EEG with GANs for some classifiers, especially with small sample studies.

Methods and Materials

Dataset

Participants In this study, we used Williams and colleagues' (Williams et al., 2021) open-source EEG dataset of 500 undergraduate student participants (see the original manuscript for a detailed breakdown of demographics). For classification analyses, these data were split into a training set of 100 participants, a validation set of 200 participants, and a test set of 200 participants. We first explored GAN and classifier meta-parameters by observing performance with the validation set, and here report findings of performance with the test set once meta-parameters were chosen.

Task Participants completed a two-armed bandit gambling task where they needed to discern which of two coloured squares were more often rewarding through trial-and-error (see Figure 1). Each trial presented two coloured squares that the participants were to choose from, and provided performance feedback as "WIN" or "LOSE", yielding two conditions of interest, c_{win}, c_{lose}. For each pair of squares, one had a win rate of 60% while the other had a win rate of 10%. Participants saw each pair of colours twenty times consecutively. There were a total of five pairs of squares (with colours randomly determined), resulting in one hundred trials per participant. This paradigm elicits well-known frontal neural differences when contrasting the win and lose outcomes, namely in the reward positivity, delta oscillations, and theta oscillations (see Williams et al., 2021).

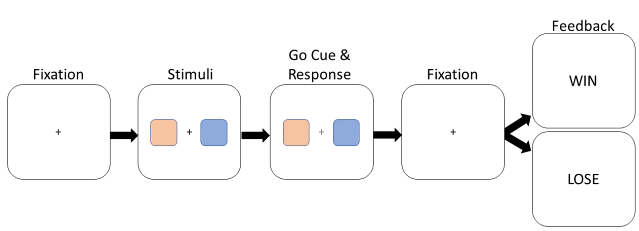

Figure 1: Two-armed bandit gambling task. Participants were tasked to determine which of two coloured stimuli were more often rewarding through trial and error. The figure was adapted from Williams et al., 2021

.

Data Acquisition and Processing EEG data were recorded on a standard EEG system (ActiCAP, Brain Products, GmbH, Munich, Germany) at a sampling rate of 500Hz with either 64 or 32 electrodes, but all data were reduced to 32 electrodes. The dataset came preprocessed with the following steps. Data were re-referenced to an averaged mastoid and filtered using Butterworth passband (0.1 to 30Hz, order 4) and notch (60Hz) filters. Eye blink artifacts were corrected using independent component analysis, and faulty electrodes were interpolated. Data were segmented to -500ms to 1,500ms around the feedback stimulus onset, baseline corrected using a -200ms to 0ms time window, and segments were removed if they surpassed artifact rejection criteria of 10μV/ms gradient and/or 100μV max-min. The data were trimmed to -200ms to $1,000$ms, resulting in trial-level data for each participant, with each trial containing 600 data points.

Although the dataset used here included a minimum of 32 electrodes, we only considered electrode FCz, as it is the prime location to assess the neural correlates of feedback processing (Williams et al., 2021). Further, the EEG data were originally recorded with a frequency of 500Hz over a duration of 1,200ms (-200ms to 1,000ms surrounding the feedback stimulus onset), which resulted in a data vector of the length 600. We downsampled this vector through linear interpolation to a vector length of 100. Further, each trial was normalized to range between 0 and 1.

Data and Script Availability All data, simulations, and analysis scripts used here are openly available: `https://autoresearch.github.io/EEG-GAN`.

Generative Adversarial Network

The generative adversarial network (GAN) was introduced by Goodfellow et al., 2014 (but see Schmidhuber, 2020 for a review of similar past work) and is capable of generating realistic samples from complex data distributions after successful training. The GAN is an artificial neural network architecture that contains two opposing networks. The first is the generator G, and the second is the discriminator D. The generator G takes in a vector z sampled from a probability distribution P_z and computes a sample $\hat{y}$ according to

$$\hat{y} = G(z; \theta_G)$$

based on the random vector z and the generator parameters θ_G. The discriminator D gets either real samples y drawn from a dataset with the feature distribution P_{data} or the generated samples $\hat{y}$ from the generator's distribution P_θ. The discriminator's task is to identify the samples as either real or fake by assigning a validity score

$$v = D(\{y, \hat{y}\}; \theta_D)$$

based on the given sample y and the discriminator parameters θ_D. This score represents the probability that a sample is drawn from the real distribution covered within the dataset D. The discriminator's objective is to maximize the probability of assigning the correct validity score to a shown sample (i.e., a high validity score for real data and a low validity score for fake data), while the generator's objective is to minimize the discriminator's objective function according to

$$\min_G 1 - log(D(\hat{y})).$$

Hence, the GAN-training's objective function is

$$\min_G \max_D v(G,D) = E_{y\sim P_{data}(y)} logD(y) + E_{z\sim P_z(z)} log(1-D(y)),$$

where the training resembles a minimax game between the generator and the discriminator (Goodfellow et al., 2014).

Adapted Architecture The most significant drawbacks of the original GAN architecture are training instability and mode collapse. The training instability results from the optimization of two independent sets of network parameters θ_G and θ_D in an adversarial manner. Small changes in the set of parameters for one network can substantially alter the performance of the other network. This circumstance leads to divergence of the respective losses. Mode collapse, on the other hand, describes a lack of variance in the generated samples $\hat{y}$. To address training instability and mode collapse, we utilized the Wasserstein-GAN (WGAN) introduced by Arjovsky et al., 2017. Hereby, the original objective function is replaced by the Wasserstein distance. Specifically, the discriminator tries to maximize the Wasserstein distance

$$\max_D E_{y\sim P_{data}}[D(y)] - E_{\hat{y}\sim P_\theta}[D(\hat{y})],$$

whereas the generator tries to minimize this distance

$$\min_G E_{\hat{y}\sim P_\theta}[D(\hat{y})].$$

As such, the generator seeks to minimize the Wasserstein distance instead of fooling the discriminator. The Wasserstein distance is differentiable throughout the entire feature space. To reduce training instabilities, Arjovsky et al. (2017) suggested weight clipping for each update of the parameters θ_D. A further reduction to training instability can be achieved by replacing weight clipping with gradient penalty (GP), as suggested by Gulrajani et al. (2017), where high gradients during the parameter update are penalized. Further, the discriminator is trained for η_{iter} iterations until optimality before the generator is trained. All these modifications result in greater training stability and less collapsing at the end of training.

Time series data, as with EEG, introduces features and difficulties not covered by images or tabular data. These features contain temporal relations between subsequent points, which are difficult to capture with linear feedforward networks or convolutional neural networks. The most popular approach to processing time series data are recurrent neural networks since they capture temporal relations by computing time step by time step. The problem with this mechanism, however, is the loss of relations between distinct points in the series.

Table 1: GAN Parameters

Description	Variable	Value
Number of Classes	$\eta_{D,cl}$	1
Number of Transformer Encoder in Discriminator	$\eta_{D,d}$	3
Attention Drop Rate	$\eta_{drop,a}$	0.5
Forward Drop Rate	$\eta_{drop,f}$	0.5
Embedding Dimension	η_{emb}	10
Number of Generated Channels	$\eta_{G,ch}$	1
Number of Transformer Encoder in Generator	$\eta_{G,d}$	3
Latent Dimension	η_{lat}	16
Patch Size	η_{patch}	20
Sequence Length	η_{seq}	100
Number of Discriminator Training Iterations Before Generator Training	η_{iter}	5

Note that descriptions of each parameter are documented at `https://autoresearch.github.io/EEG-GAN`.

An alternative to recurrent neural networks are tansformers, which do not suffer from loss of dependencies regardless of the time between two consecutive points since the underlying attention mechanism learns to pay attention to different parts of the sequence regardless of their distance (Vaswani et al., 2017). Further, transformers have significantly fewer parameters than comparable convolutional or recurrent neural networks (Vaswani et al., 2017). Therefore, we implemented a transformer-encoder-based generator and discriminator for time series (TTS) as introduced by Li et al. (2022). The TTS-GAN parameters are listed in Table 1.

Transformers originate from natural language processing, where each word is tokenized as part of the whole sequence. Further, each token is computed with a positional embedding. Likewise, our time series needed tokenizable parts, which were introduced by Li et al. (2022) as patches. Hereby, each patch was a subvector of the original vector with a length $n_{sub} \leq n_{seq}$. Since each patch is of the same length, the vector length n_{seq} needed to be divisible by n_{sub}. An investigation of the patch length showed that good results were achieved with $n_{sub} = 20$. Hence, the original vector was divided into 5 parts with equal lengths.

Since the data consists of EEG samples under two experimental conditions $c = \{c_{win}, c_{lose}\}$, the GAN's input includes this information as well. That is, the generator receives as input information about the experimental condition c, in addition to the noise vector z, while the discriminator gets this information in addition to the sample y or $\hat{y}$. This architecture is called *conditional* GANs (Mirza and Osindero, 2014).

In summary, we implemented a conditional TTS-WGAN-GP architecture for more stable and less collapsing training along with the advantages of transformers regarding time series data. Note that this GAN architecture is unique in comparison to the aforementioned literature investigating GAN-enhanced classification of EEG (Fahimi et al., 2020; Kan et al., 2021; Luo and Lu, 2018; Petruţiu et al., 2020).

Training Procedure The GAN was trained on trial-by-trial EEG data (and, thus, neither on participant-averaged nor grand-averaged waveforms). Training was performed with an ADAM optimizer with the learning rate $l_r = 0.0001$ and decay rates $\beta_1 = 0$, $\beta_2 = 0.9$ (Gulrajani et al., 2017). Furthermore, to mimic empirical processing steps, the same Butterworth band-pass filter (0.1Hz to 30Hz, order 4) was applied to the samples prior to feeding them to the discriminator. The discriminator was trained for $\eta_{iter} = 5$ iterations each time before the generator was trained once, as suggested by Arjovsky et al. (2017).

Evaluation

Methods

Before determining whether GAN-augmented EEG enhances classification performance, we examined whether it produced realistic EEG samples with underlying neural features in the first place (Fahimi et al., 2019; Hartmann et al., 2018). So far, there is no gold standard for evaluating the quality of synthetic EEG data produced by GANs (Brophy et al., 2021). However, there are various criteria EEG data should satisfy.

Qualitatively, we evaluated our GANs through visual inspection 1) of trial-level EEG data, 2) grand-averaged ERPs, and 3) variance along the first two principal components. For these evaluations, we trained a GAN on the trial-level data of all 500 participants. Following this, we generated 50,000 samples for the win condition and 50,000 samples for the lose condition. We then randomly selected and visually compared five empirical and five synthetic trials. We also averaged the synthetic data across all trials for each condition to produce grand-averaged ERPs. These were then compared to empirical ERPs. Finally, we projected the synthetic and real EEG data onto the first two principal components of the pooled dataset and inspected the overlap between the projected synthetic and real data.

For quantitative evaluations, we followed the Train on Synthetic, Test on Real (TSTR) method (Esteban et al., 2017). This method trains a classifier exclusively on synthetically generated data and then tests its performance on a held-out sample of real data. For comparison, we also conducted a Train on Real, Test on Real (TRTR) analysis. For the TSTR analysis, we trained a GAN on the 100 participant training set and then generated 5,000 samples for each of the win and lose conditions (in line with empirical trial counts; Williams et al., 2021). For the TRTR analysis, we trained a classifier on the 100 participant training set (i.e., the same set used to train the GAN in TSTR). For both analyses, we determined classifier performance using the test set. The classifier used here was a neural network, and we followed the same procedures as described in the Classification section.

Results: Synthetic Data Reflected Realistic EEG

We found that empirical and synthetically generated trial-level EEG data were qualitatively indistinguishable, see Figure 2. We also found that synthetic ERPs qualitatively matched empirical ERPs and contained the same underlying features, see Figure 3. Further, we found that both the generated and empirical data shared the same principal component space, see Figure 4.

Both the TSTR and TRTR classifications achieved a performance of 70% (SD_{TSTR}: 0.89%, SD_{TRTR}: 3.33%), indicating equal predictability between synthetic and empirical classifications. Altogether, these results suggest that the GANs do accurately produce realistic EEG data with underlying cognitive features, such as the neural reflections of the reward positivity ERP.

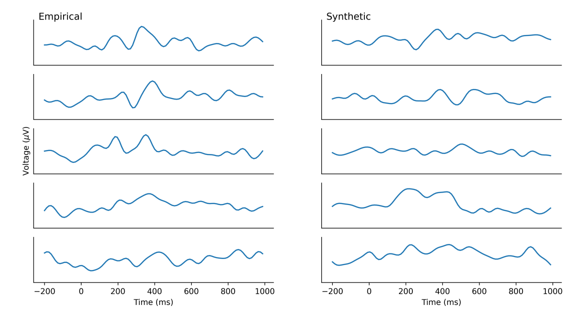

Figure 2: Trial-level empirical and synthetically derived EEG data.

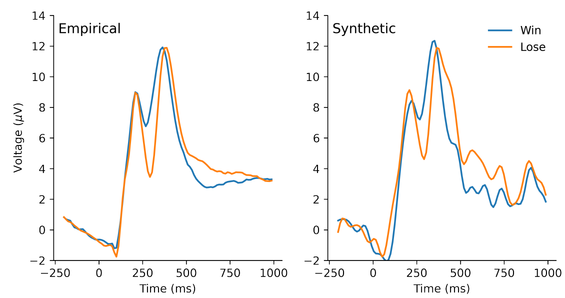

Figure 3: Grand-averaged empirical and synthetically derived ERPs. Synthetic data scaled to match empirical units.

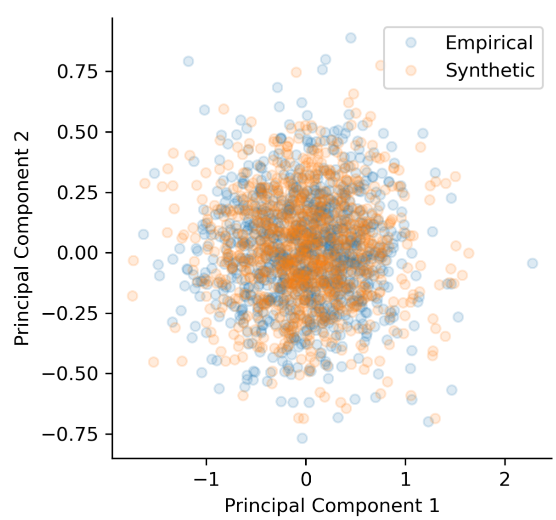

Figure 4: Principal component space of the first and second principal components for empirical and synthetic data.

Classification

Methods

Experimental Factors Participants completed a two-armed bandit gambling task in which they received win and lose feedback. As such, we classified whether the participants experienced a win or a lose. We assessed the difference between empirical and augmented classification performance across three factors: classifier (neural networks, SVMs, and logistic regressions), data format (full time series, and extracted features), and sample size (five to 100 participants).

We used grid search to determine classification performance for each of the three **classifiers** (conducted using Scikit-Learn; Pedregosa et al., 2011; all parameters not discussed were left at default). We permitted a search across five neural network structures. Each network consisted of one to three layers. We varied the number of units N_i in hidden layer i, resulting in the following architectures: $(N_1 = 25)$, $(N_1 = 50)$, $(N_1 = 25, N_2 = 25)$, $(N_1 = 50, N_2 = 50)$, $(N_1 = 50, N_2 = 25, N_3 = 50)$. Furthermore, we searched across the following training parameters: 0.0001 and 0.05 L2 regularization strengths; logistic, tanh, and ReLU activation functions; constant, inverse scaling, and adaptive learning rate schedules; and 10,000 and 20,000 maximum iterations. For the SVMs, we searched across 0.1, 1, 10, and 100 inverse regularization strengths; radial basis function, polynomial, and sigmoid kernel types; and 1.0, 0.1, 0.01, 0.001 kernel coefficients. For the logistic regressions, we fixed the solver to be liblinear, but searched across L1 and L2 penalties; 20 equal log steps from 0.0001 to 1,000 inverse regularization strengths; and 5,000, 10,000, 20,000, and 50,000 maximum iterations. For classification, we used participant averaged data, rather than trial-by-trial segments. For example, the training set with 100 participants and two conditions would result in 200 samples.

Classification performance was assessed separately on two **data formats**: the full time series and extracted features. For the full time series analyses, we used the entire 100 data points as predictors in our classification models. For the extracted features analyses, we had three predictors corresponding to the aforementioned neural signals of this task—namely, reward positivity amplitude, delta power, and theta power. Our analyses focused on electrode FCz, where these neural differences are strongest (Williams et al., 2021). We extracted the reward positive amplitude for each participant and condition by averaging EEG data within the same time window as in Williams et al. (2021), i.e., 264ms to 356ms post-feedback onset. We extracted delta and theta power for each participant and condition by applying a Fast Fourier Transform (FFT) to the data and averaging within the 0-3Hz and 4-8Hz frequency ranges for delta and theta power, respectively.

Our study also sought to investigate the magnitude of this potential increase across different **sample sizes**. Thus, we randomly extracted (without replacement) sample sizes of 5, 10, 15, 20, 30, 60, and 100[1] from the training set of 100 participants. We repeated this process five times for each sample size, resulting in 35 separate datasets (note that the five sets of 100 participants would be identical).

Altogether, for each combination of classifier (neural network, SVM, logistic regression), and feature set (full time series, extracted features), we had a total of seven sample sizes, each with five iterations of randomly selected data. We conducted 10 classifications for each of these.

Data Augmentation For each of the 35 datasets, we trained a GAN on trial-level data and generated 5,000 (half win, half lose) synthetic samples. Although the empirical data had been filtered using a Butterworth pass-band filter (0.1 to 30Hz, order 4), there were no constraints on the frequencies that the GAN could produce. Thus, we applied the same filter to the synthetic samples. We also baseline-corrected the data using the -200 to 0ms time window. Since classification for the empirical data was based on participant-averaged data, we averaged the generated samples in bins of 50 within each condition. This process resulted in 50 synthetic "participants", each with a participant-averaged win and lose time series. To create a series of 35 augmented training sets, this data was randomly inserted among the corresponding empirical training sets.

Results: Augmented EEG Improved Classification Performance

We found that augmenting EEG data enhanced classification performance for both neural networks and SVMs, but not for logistic regressions (see Figure 5). Further, for sample sizes under 60, the best classification performance across all classifiers was achieved with augmented data. For example, augmented classification using SVM at sample size 30 outperformed empirical classifications for all classifiers.

The augmented classification for the neural network and SVM classifiers also persisted across both of the data formats (full time series and extracted features). Augmented EEG had the largest effect on neural network classifiers, reaching up to 12%, while its effect on SVMs reached up to 8%. Furthermore, GAN-enhanced classification was larger when using the extracted features as predictors than when using the full time series as predictors.

We also investigated whether the empirical sample size impacted GAN-augmented classification enhancements. We found for both the neural network and SVM classifiers that EEG-augmented enhancements diminished with increasing sample sizes, indicating that enhancements are most beneficial for smaller sample size studies. In some cases, augmentation with smaller sample sizes could even enhance classification to the point where it was on par with the largest unenhanced sample sizes here investigated. For example, when considering the extracted features set, augmented SVM performance at an empirical sample size of 15 matched the SVM performance at an empirical sample size of 100.

[1]These sample sizes were chosen to both include a high density of small sample sizes (i.e., 5-20), as well as, typical sample sizes seen in the EEG literature (i.e., 30-100).

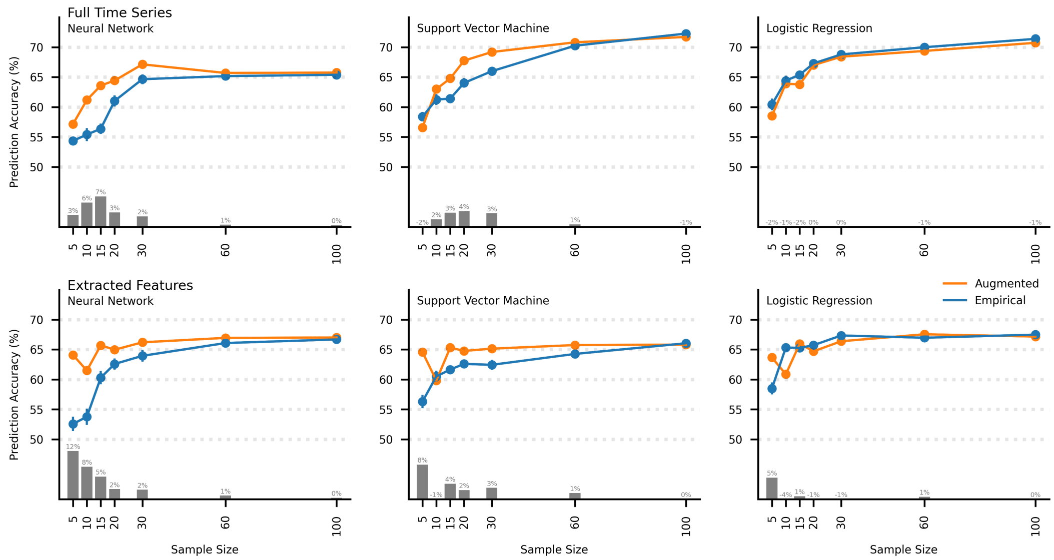

Figure 5: Classification performance across classifiers with two data formats, and across seven sample sizes for empirical (blue line) and augmented (orange line) data. Error bars reflect the standard error of the mean. The left, middle, and right panels show classification performance using a neural network, SVM, and logistic regression, respectively. The top panels show classification performance based on full EEG time series, whereas the bottom panels show classification performance for extracted EEG features (see main text for more details). Grey bars in each panel indicate the difference between augmented and empirical performance, with positive values reflecting enhanced performance for the augmented classifications.

Discussion

We investigated the extent to which augmenting EEG data with GANs can improve classification performance. Our analyses focused on GAN-enhancements across three classifiers and seven sample sizes. In line with previous efforts (Fahimi et al., 2019; Hartmann et al., 2018), we confirmed that GANs can generate realistic EEG samples that contain cognitive features. Next, we found that augmenting EEG enhanced classification performance for neural networks and SVMs, but not logistic regressions. Indeed, the best classifications for sample sizes under 60 were achieved with augmented data. These findings persisted across two data formats where classifiers were provided inputs of the full time series or of three extracted features—the reward positivity, delta power, and theta power. Finally, we found that classification enhancements diminished as sample sizes increased.

These findings suggest that GAN-derived EEG augmentation can enable better classification performance with less data—saving researchers time and money (Lotte et al., 2018). Increased classification performance with smaller sample sizes could also permit research with studies where it is difficult to collect large amounts of data. One exciting avenue would be to investigate whether our findings generalize to classify clinical populations with limited EEG data (e.g., Parkinson's disease) (Brophy et al., 2021). Although useful for small sample sizes, our technique did not influence classification with large sample sizes. This is because the data generated from GAN models provide classifiers with unique samples, effectively interpolating between empirical samples and increasing the signal-to-noise ratio. However, large sample studies contain a dense distribution of samples, which is already sufficient for classifier training.

Given the novelty of this approach, there are many future directions to consider. For example, as different GAN architectures have not been extensively contrasted within an EEG context, it will be important to compare the effectiveness of our method—a conditional TTS-WGAN-GP architecture—with alternative GAN architectures (WGAN, TTS-GAN, etc.). Further, it will be important to test our method with different datasets, such as the classification of clinical disorders (Brophy et al., 2021) and neural markers for brain-computer interfacing (Lotte et al., 2018). Separately, one could investigate other applications of GANs in EEG research—for example, as an approach to recover faulty electrodes or increase the electrode density of mobile EEG.

In conclusion, we investigated the degree to which GANs can augment EEG to enhance classification performance. We assessed which classifiers, and which empirical sample sizes, benefit from this augmentation. We found that GAN-augmented EEG enhanced classification with neural networks and SVMs, but not logistic regressions, suggesting that researchers must carefully select which classifiers to use. Further, we found the largest enhancements with smaller sample sizes, opening the potential for small sample studies to reach reliable classification outcomes.

Acknowledgements

This research was supported by Schmidt Science Fellows, in partnership with the Rhodes Trust, as well as the Carney BRAINSTORM program at Brown University.

References

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

Brophy, E., Wang, Z., She, Q., & Ward, T. (2021). Generative adversarial networks in time series: A survey and taxonomy. *arXiv preprint arXiv:2107.11098*.

Dong, Y., & Ren, F. (2020). Multi-reservoirs eeg signal feature sensing and recognition method based on generative adversarial networks. *Computer Communications*, *164*, 177–184.

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

Fahimi, F., Dosen, S., Ang, K. K., Mrachacz-Kersting, N., & 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.

Fahimi, F., Zhang, Z., Goh, W. B., Ang, K. K., & Guan, C. (2019). Towards eeg generation using gans for bci applications. *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, 1–4.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *NIPS*.

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein gans. *Advances in neural information processing systems*, *30*.

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

Hwang, S., Hong, K., Son, G., & Byun, H. (2019). Ezsl-gan: Eeg-based zero-shot learning approach using a generative adversarial network. *2019 7th International Winter Conference on Brain-Computer Interface (BCI)*, 1–4.

Kan, C. N. E., Povinelli, R. J., & Ye, D. H. (2021). Enhancing multi-channel eeg classification with gramian temporal generative adversarial networks. *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1260–1264.

Koo, P. C., Berger, C., Kronenberg, G., Bartz, J., Wybitul, P., Reis, O., & Hoeppner, J. (2019). Combined cognitive, psychomotor and electrophysiological biomarkers in major depressive disorder. *European archives of psychiatry and clinical neuroscience*, *269*(7), 823–832.

Li, X., Metsis, V., Wang, H., & Ngu, A. H. H. (2022). Tts-gan: A transformer-based time-series generative adversarial network. *arXiv preprint arXiv:2202.02691*.

Lotte, F., Bougrain, L., Cichocki, A., Clerc, M., Congedo, M., Rakotomamonjy, A., & Yger, F. (2018). A review of classification algorithms for eeg-based brain–computer interfaces: A 10 year update. *Journal of neural engineering*, *15*(3), 031005.

Luo, Y., & Lu, B.-L. (2018). Eeg data augmentation for emotion recognition using a conditional wasserstein gan. *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2535–2538.

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

Panwar, S., Rad, P., Jung, T.-P., & Huang, Y. (2020). Modeling eeg data distribution with a wasserstein generative adversarial network to predict rsvp events. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *28*(8), 1720–1730.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Petruţiu, V. M., Palcu, L. D., Lemnaru, C., Dınşoreanu, M., Potolea, R., Murseşan, R., & Moca, V. V. (2020). Enhancing the classification of eeg signals using wasserstein generative adversarial networks. *2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)*, 29–34.

Schmidhuber, J. (2020). Generative adversarial networks are special cases of artificial curiosity (1990) and also closely related to predictability minimization (1991). *Neural Networks*, *127*, 58–66.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Williams, C. C., Ferguson, T. D., Hassall, C. D., Abimbola, W., & Krigolson, O. E. (2021). The erp, frequency, and time–frequency correlates of feedback processing: Insights from a large sample study. *Psychophysiology*, *58*(2), e13722.