Augmenting EEG with Generative Adversarial Networks Enhances Brain Decoding Across Classifiers: A Replication Study

Bahadır Gezer
bahadirgezerbg@gmail.com

Berke Çalışkan
berke.caliskan@std.bogazici.edu.tr

Simar Achmet Kechagia
simar.achmet@std.bogazici.edu.tr

Department of Computer Engineering, Boğaziçi University
34342 Bebek/Istanbul, Türkiye

Information about the Original Study

Understanding our replication effort requires an overview of the original study. Conducted by Williams and titled "Augmenting EEG with Generative Adversarial Networks Enhances Brain Decoding Across Classifiers and Sample Sizes" (W. D. W. M. . M. S. Williams C. C,) the study investigates whether augmenting EEG brainwave data can improve classification performance.

Research Questions

The original study addressed two primary research questions:

1. For which classifiers is GAN-enhanced classification effective?

2. How does the impact on enhanced classifications vary with sample size?

To investigate these questions, the authors used three classifiers: neural networks, support vector machines (SVMs), and logistic regressions. They evaluated the effectiveness of GAN-augmented EEG data across seven sample sizes ranging from 5 to 100 participants.

Participants

The study involved a dataset of 500 undergraduate student participants. The dataset, provided by Williams et al. (2021) (C. C. Williams, Ferguson, Hassall, Abimbola, & Krigolson,), was split into a training set of 100 participants, a validation set of 200 participants, and a test set of 200 participants. Participants completed a two-armed bandit gambling task where they needed to discern which of two colored squares were more often rewarding through trial-and-error. Each trial presented two colored squares that the participants were to choose from and provided performance feedback as "WIN" or "LOSE," yielding two conditions of interest: winning and losing.

Design

The study design involved generating augmented EEG data using a transformer-based GAN (Goodfellow et al.,) from the original dataset of participants. For each sample size (ranging from 5 to 100 participants), synthetic data was generated to match the original participant data. The classification performance was then assessed by comparing the success rates of classifiers trained on both the original and augmented data.

The comparisons were conducted as follows: 1) for each sample size, GANs were used to generate synthetic EEG data based on the original participant data, 2) classifiers -neural network, SVM, logistic regression- were trained separately on the original data and the augmented data, and 3) the classification success rates of each classifier were evaluated and compared between the original and augmented datasets.

The primary objective was to determine whether augmenting EEG data improves classification performance, identify which classifiers benefit the most from augmentation, and assess the effectiveness of augmentation across different sample sizes.

Artifacts

The artifacts used in the study include the EEG dataset (C. C. Williams et al.,), the two-armed bandit gambling task, and the transformer-based GAN architecture. All artifacts can be found in the neatly designed website for the original paper: `https://autoresearch.github.io/EEG-GAN`.

Summary of Results

The original study found that GAN-augmented EEG data improved classification performance for neural networks and support vector machines (SVMs), but not for logistic regression. This enhancement was more pronounced with smaller sample sizes, suggesting that GANs are most effective when dealing with limited data. As sample sizes increased, the benefits of GAN augmentation diminished. These results indicate that augmenting EEG data with GANs can significantly enhance brain decoding capabilities, particularly for neural networks and SVMs, in studies with limited data.

Information about the Replication

Motivation for Conducting the Replication

We conducted this replication study to validate the original results obtained by Williams et al. (2023) (W. D. W. M. . M. S. Williams C. C,). While following their

methods closely, we utilized a different dataset (J. J. Bird,) to assess the generalizability of their findings.

Level of Interaction with the Original Experimenters

We did not have any interaction with the original researchers. However, this was not necessary due to the comprehensive documentation accompanying their research. We relied solely on the published materials, which were extensive and well-covered. The open-source dataset (C. C. Williams et al.,), source code available on GitHub [1], and a detailed Google Colab notebook provided all the necessary information for us to proceed with our replication independently.

Changes to the Original Experiment

Short Answer: We did not have different participant sizes, so we classified one size of participants with a single dataset using the three classifiers. **Long Answer:** We maintained the same training model, parameters, and classifiers as the original study. The primary change was that we conducted our experiment on a single participant size. Additionally, due to resource and time limitations, our training model had only 500 epochs instead of the 8000 epochs used in the original study. Consequently, we focused on one research question: the effect of data augmentation on different classifier techniques.

We selected our dataset at the beginning of the replication. The dataset (J. J. Bird,) comprised EEG recordings from two participants, each classified over three-minute sessions. We considered each of these three-minute classifications as equivalent to a 'participant' in the original study. The original dataset included 'positive,' 'neutral,' and 'negative' classifications. To align with the original study's 'win' and 'lose' conditions, we filtered out the 'neutral' responses, resulting in 1416 'participant' equivalents. This dataset was split 20-80 for testing and training. We figured that using more 'participant' equivalents would be justifiable due to the limited number of actual participants. We did not alter participant sizes throughout the study because our focus was not on this variable.

Replication Methodology

The data augmentation process remained unchanged, utilizing the same model with smaller epochs but a larger training sample size. We generated the same amount of augmented data. We trained the same model with the same settings, except for reducing the epochs to 500. For each of the three classifiers—neural network, SVM, and logistic regression—we classified the test data and then the training data augmented with the synthetic data. The success rates of these classifications were then compared.

---

[1] https://github.com/AutoResearch/EEG-GAN

Comparison of Replication Results to Original Results

Our replication study yielded several consistent results with the original study by Williams et al. (2023) (W. D. W. M. . M. S. Williams C. C,). For both the neural network and SVM classifiers, the empirical and augmented classification accuracies were identical, 91% and 93% respectively, matching the original study's findings that GAN-augmented EEG data did not significantly improve performance for these classifiers. However, for the logistic regression classifier, our replication showed a slight improvement in augmented classification accuracy (86%) compared to the empirical classification accuracy (84%), contrary to the original study, which found no significant benefit of GAN-augmented data for logistic regression.

These differences can be attributed to factors such as dataset characteristics, with our smaller dataset involving only two participants and emotions classified over three-minute sessions. Each session was treated as a 'participant' equivalent, influencing classifier performance, particularly for logistic regression. Additionally, our GAN model was trained for 500 epochs compared to the 8000 epochs in the original study, potentially impacting the quality of the synthetic data. Unlike the original study, we did not vary participant sample sizes, focusing instead on a single size derived from our dataset, which may have limited the generalizability of our findings.

Despite these differences, our replication supports the original study's conclusion that augmenting EEG data with GANs can improve classification performance. The consistent results for neural networks and SVMs affirm the robustness of the original findings, while the slight improvement for logistic regression suggests further investigation with larger and more varied datasets.

Conclusions Across Studies

By combining conclusions from both the original study and our replication, we gain insights that would not have been evident from either study individually. Our findings strengthen the original study's conclusion that GAN-augmented EEG enhances classification performance for neural networks and SVMs, particularly in small-sample studies. The slight improvement observed for logistic regression classifiers in our study highlights the potential variability in GAN augmentation effectiveness across different datasets and conditions.

This replication underscores the importance of dataset characteristics and training parameters in influencing the outcomes of GAN-augmented EEG classification. Future research should explore the impact of varying participant sample sizes, different GAN architectures, and diverse datasets, including clinical populations with limited EEG data. These investigations could further validate and expand upon the findings, providing a more comprehensive understanding of the utility of GANs in EEG data

augmentation.

In conclusion, our replication study not only validates the original findings but also opens new avenues for exploring the potential of GAN-augmented EEG data in improving classification performance across various contexts.

Acknowledgments

References

Bird, J., Ekart, A., Buckingham, C., Faria, D. (2019, 04). Mental emotional sentiment classification with an eeg-based brain-machine interface..

Bird, J., Faria, D., Manso, L., Ekárt, A., Buckingham, C. (2019, 08). A deep evolutionary approach to bioinspired classifier optimisation for brain-machine interaction..

Bird, J. J. (2019). *Eeg brainwave dataset: Feeling emotions* (Tech. Rep.). Retrieved from https://www.kaggle.com/datasets/birdy654/eeg-brainwave-dataset-feeling-emotions

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014, June). Generative adversarial networks. *arXiv e-prints*.

Williams, C. C., Ferguson, T. D., Hassall, C. D., Abimbola, W., Krigolson, O. E. (2020, November). The ERP, frequency, and time-frequency correlates of feedback processing: Insights from a large sample study. *Psychophysiology*, *58*(2), e13722.

Williams, W. D. W. M. . M. S., C. C. (2023). Augmenting eeg with generative adversarial networks enhances brain decoding across classifiers and sample sizes. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*. Retrieved from https://escholarship.org/uc/item/9gz8g908