RANDOMIZED TRIALS IN EEG CLASSIFICATION EXPERIMENTS ARE MANDATORY AS DRIFT IS PERVASIVE

Anonymous authors
Paper under double-blind review

ABSTRACT

Temporal correlation is demonstrated in three public EEG datasets. Filtering does not remove the correlation. The community is cautioned that a substantial number of recent publications are flawed due to a confound between temporal correlation and stimulus class, despite claims that filtering removes the correlation. The only known way to avoid this confound is through randomization.

1 INTRODUCTION

A recent sequence of papers (Spampinato et al., 2016; 2017; Kavasidis et al., 2017; Palazzo et al., 2017; 2018; Tirupattur et al., 2018; Palazzo et al., 2020a;b; 2021; 2024) claims to classify visual stimuli from EEG recordings. With the exception of Tirupattur et al. (2018), which uses a dataset from Kumar et al. (2018), these all use a dataset from Spampinato et al. (2017), an early version of which was reported in Spampinato et al. (2016). Li et al. (2021), Ahmed et al. (2021; 2022), and Bharadwaj et al. (2023), however, demonstrated a confound in these datasets: a spurious correlation between stimulus class and time, due to poor experiment design. This invalidates the work reported in nearly one hundred papers as discussed below. Recent responses (Palazzo et al., 2020b; 2024) argued that

I) the design of the data-collection effort in Li et al. (2021) and Ahmed et al. (2021) induced or exacerbated the temporal correlation due to long recording sessions that may introduce fatigue, and

II) the bandpass filtering performed by Spampinato et al. (2017) Kavasidis et al. (2017), and Palazzo et al. (2017; 2018; 2020a;b; 2021; 2024) could remove the temporal correlation.

Here, we demonstrate that

A) three public datasets with short recording sessions also contain a temporal correlation and
B) bandpass filtering does not remove the temporal correlation.

This supports the view that temporal correlation is pervasive in EEG data and that the only known valid way to avoid a confound is through appropriate randomization.

2 SIGNIFICANCE

Several independent lines of research have refuted a large body of flawed work (Spampinato et al., 2017; Kavasidis et al., 2017; Palazzo et al., 2017; 2018; 2020a;b; 2021; 2024) along completely different axes. Li et al. (2021) demonstrated that the dataset used (Spampinato et al., 2017), and the methods used to collect that dataset, suffer from a temporal confound, correlating stimulus class with experiment timing. Accuracy drops to chance when the confound is removed. Ahmed et al. (2021) demonstrated that this holds even with a much larger dataset. Ahmed et al. (2022) demonstrated that this holds for the additional classifiers used in Palazzo et al. (2018; 2020a;b; 2021). Bharadwaj et al. (2023) demonstrated that this holds even when using supertrials.

Here we progress beyond prior demonstrations that the particular dataset (Spampinato et al., 2017) is confounded. We offer the novel claim that the collection protocol is inherently and irreparably confounded; any dataset collected with this protocol would by its very nature suffer from the same confound. This responds specifically to claims I and II above with our counter claims A and B in ways that have not previously been addressed. This is significant because the confounded dataset

continues to be used, and confounded collection methods continue to be used to collect new confounded datasets, justified, in part, by claims I and II. This is further significant for the following reasons:

- Nearly one hundred papers (An & Cho, 2016; Spampinato et al., 2016; Ben Said et al., 2017; Bozal Chaves, 2017; Kavasidis et al., 2017; Palazzo et al., 2017; Parekh et al., 2017; Spampinato et al., 2017; Zhang et al., 2017; Du et al., 2018; Fares et al., 2018; Kumar et al., 2018; Palazzo et al., 2018; Piplani et al., 2018; Tirupattur et al., 2018; Wang et al., 2018; Zhang & Liu, 2018; Zhang et al., 2018; Zhong et al., 2018; Du et al., 2019; Hwang et al., 2019; Jiang et al., 2019; Jiao et al., 2019; Long et al., 2019; Mukherjee et al., 2019; Uys, 2019; Wang et al., 2019; Cudlenco et al., 2020; Fares et al., 2020; Li et al., 2020; Palazzo et al., 2020a;b; Wang et al., 2020; Zheng et al., 2020a;b; Palazzo et al., 2021; Zheng & Chen, 2021; Ma et al., 2021; Mo et al., 2021; Jiang et al., 2021; Lee et al., 2021; Cavazza et al., 2022; Khaleghi et al., 2022; Lee et al., 2022; Mishra et al., 2022; Mishra, 2022; Scharnagl & Groth, 2022; Shimizu & Srinivasan, 2022; Ahmadieh et al., 2023; Bai et al., 2023; Du et al., 2023; Duan et al., 2023; Hasan & A, 2023; Imani et al., 2023; Lan et al., 2023; Lee et al., 2023; Liu et al., 2023; Singh et al., 2023; Song et al., 2023; Wahengbam et al., 2023; Zeng et al., 2023b;a; Fan et al., 2024; Ferrante et al., 2024a;b; Gou et al., 2024; Lei et al., 2024; Liu et al., 2024a;b; Luvsansambuu et al., 2024; Mishra et al., 2024; Mwata-Velu et al., 2024; Ngo et al., 2024; Palazzo et al., 2024; Pan et al., 2024; Qian et al., 2024; Singh et al., 2024; Tang et al., 2024; de la Torre-Ortiz et al., 2024; Yang & Liu, 2024; Ye et al., 2024; Zheng et al., 2024b;a; Zhu et al., 2024; Deng et al., 2025; Fares, 2025; Fu et al., 2025; Lopez et al., 2025; Mehmood et al., 2025; Singh et al., 2025; Xiang et al., 2025) draw flawed conclusions based on the confounded dataset from Spampinato et al. (2017) and datasets suffering from the same confound.
- A number of new datasets have been collected with this same confounded protocol (Gou et al., 2024; Pan et al., 2024; Zhu et al., 2024; Qian et al., 2024; Uys, 2019; Shimizu & Srinivasan, 2022; Liu et al., 2024b; Wang et al., 2019; 2020; Ma et al., 2021; Cudlenco et al., 2020; Zheng et al., 2024b; Cavazza et al., 2022; Luvsansambuu et al., 2024; Liu et al., 2023; Bai et al., 2023; Parekh et al., 2017).
- A number of these have been publicly released and are used by others. For example, Singh et al. (2023), Singh et al. (2024), and Lopez et al. (2025) use the dataset reported in Kumar et al. (2018) and Duan et al. (2023), Singh et al. (2024), and Lopez et al. (2025) use the dataset reported in Ma et al. (2021).
- This is further egregious because Palazzo et al. (2020b; 2024) continue to claim that their dataset (Spampinato et al., 2017), and their results that were obtained with that dataset (Spampinato et al., 2017; Kavasidis et al., 2017; Palazzo et al., 2017; 2018; 2020a;b; 2021; 2024), are valid, despite the refutations in Li et al. (2021), Ahmed et al. (2021; 2022), and Bharadwaj et al. (2023), in part, because of claims I and II in Palazzo et al. (2020b; 2024).
- This has been used to justify continued publication of a large and growing body of flawed work based on confounded datasets (Cavazza et al., 2022; Khaleghi et al., 2022; Lee et al., 2022; Mishra et al., 2022; Mishra, 2022; Scharnagl & Groth, 2022; Shimizu & Srinivasan, 2022; Ahmadieh et al., 2023; Bai et al., 2023; Du et al., 2023; Duan et al., 2023; Hasan & A, 2023; Imani et al., 2023; Lan et al., 2023; Lee et al., 2023; Liu et al., 2023; Singh et al., 2023; Song et al., 2023; Wahengbam et al., 2023; Zeng et al., 2023b;a; Fan et al., 2024; Ferrante et al., 2024a;b; Gou et al., 2024; Lei et al., 2024; Liu et al., 2024a;b; Luvsansambuu et al., 2024; Mishra et al., 2024; Mwata-Velu et al., 2024; Ngo et al., 2024; Palazzo et al., 2024; Pan et al., 2024; Qian et al., 2024; Singh et al., 2024; Tang et al., 2024; de la Torre-Ortiz et al., 2024; Yang & Liu, 2024; Ye et al., 2024; Zheng et al., 2024b;a; Zhu et al., 2024; Deng et al., 2025; Fares, 2025; Fu et al., 2025; Lopez et al., 2025; Mehmood et al., 2025; Singh et al., 2025; Xiang et al., 2025) even after the confound became known through the work of Li et al. (2021), Ahmed et al. (2021; 2022), and Bharadwaj et al. (2023).

Current machine-learning conferences, and more generally, computer-science conferences and journals, are loathe to publish refutations. Observing this, Schaeffer et al. (2025) proposed that the field of machine-learning establish a "refutations and critiques" track in prominent conferences. While we applaud and support this proposal, the current lack of such a track should not be an impediment to publishing refutations. Scientific journals in other fields have long done so, often resulting in retraction of flawed work. Schaeffer et al. (2025) offer five example pieces of claimed flawed work in machine learning. Each is an individual paper. These pale in comparison to the flaws we uncover

here: a systemic flaw of the entire peer review process across an entire field of inquiry, namely classification of stimulus image class from EEG recordings, that affects seventeen datasets and ninety one papers. Moreover, none of the five examples in Schaeffer et al. (2025) are egregious; here the authors of the flawed work continue to argue for its validity despite four refereed refutations and fifty new flawed papers have been published subsequent to these four refereed refutations. This argues for the need to make the community aware of the severity of the issue.

3 METHOD

The dataset from Spampinato et al. (2017) recorded EEG data from six subjects, in four runs per subject, each run containing ten blocks, each block presenting fifty stimuli for 0.5 s each, with 10 s blanking between blocks. All fifty stimuli in each block were of the same class and the data from all subjects were pooled and divided into training and test sets, leading to the confound. We replicate this design with public EEG data as closely as possible. The twist is that two datasets (Hatlestad-Hall et al., 2021; Babayan et al., 2019) contain resting-state data and the third (Delorme, 2020) contains data where all trials have the same stimulus. We construct a simulated experiment design, repurposing and repartitioning this data in a fashion that mirrors the design of Spampinato et al. (2017) as closely as possible, associate block-level class labels with this data in the manner of Li et al. (2021, § 3.7), and train and test the LSTM classifier from Spampinato et al. (2017) and the EEGChannelNet classifier from Palazzo et al. (2018; 2020a;b; 2021; 2024) on this repurposed data. Above-chance classification accuracy, both with and without bandpass filtering, demonstrates the temporal correlation in this data and the fact that filtering fails to remove that correlation.

The design of Spampinato et al. (2017) is $6 \times 4 \times (10 \times (50 \times 0.5 \text{ s} + 10 \text{ s}) - 10 \text{ s})$, requiring 340 s of data per run and 1360 s per subject. Hatlestad-Hall et al. (2021) contains 4 min resting-state data for 111 subjects. Due to insufficient data per subject, we pooled multiple (four) subjects into a single virtual subject (six total), taking each subject's data as a run. The 4-min runs allowed simulating 50 stimulus presentations for each of 4 runs with 7 classes per run, for a total of 28 simulated classes, for each of 6 subjects: $6 \times 4 \times (7 \times (50 \times 0.5 \text{ s} + 10 \text{ s}) - 10 \text{ s})$. Babayan et al. (2019) contains 16 min of resting-state data for each of 216 subjects, organized as sixteen 60 s blocks, alternating between eyes closed and eyes open. We used only the first five eyes-closed blocks for each subject. We took each 60 s block to simulate two blocks from the design of Spampinato et al. (2017): $50 \times 0.5 \text{ s} + 10 \text{ s} + 50 \times 0.5 \text{ s}$, took five such blocks from each subject to simulate ten blocks from the design of Spampinato et al. (2017), pooled the data from four subjects to simulate each virtual subject, for a total of six virtual subjects: $6 \times 4 \times 5 \times (50 \times 0.5 \text{ s} + 10 \text{ s} + 50 \times 0.5 \text{ s})$. This allowed simulating 50 stimulus presentations for each of 20 runs with 2 classes per run, for a total of 40 simulated classes, for each of 6 subjects. Delorme (2020) contains three runs, each with 13 min of data, for each of 13 subjects, divided into 1 s trials. Each trial contained one of 3 stimuli: 70% contained the same tone A, 15% contained the same tone B, and 15% contained white noise. We extracted only the 70% of the trials with tone A, concatenated them into a simulated run of 546 s, divided this into 10 simulated blocks with 50 simulated stimulus presentations per block and 10 s blanking between blocks, pooled the data from two subjects to create each virtual subject, with each virtual subject having four virtual runs, two from each subject, for a total of six virtual subjects: $6 \times 4 \times (10 \times (50 \times 0.5 \text{ s} + 10 \text{ s}) - 10 \text{ s})$. This allowed simulating 50 stimulus presentations for each of 4 runs with 10 classes per run, for a total of 40 simulated classes, for each of 6 subjects.

Note that two of the datasets are resting state. The third is completely auditory. Thus none of the datasets required visual attention during the recording sessions, and two did not require any attention at all.

4 RESULTS

For each dataset, the simulated data for 6 subjects were pooled and randomly split into 80% training and 20% test data, with results averaged across five such splits. This was repeated twice, first without any filtering (Figs. 1 and 2 row 1 for LSTM, row 3 for EEGChannelNet) and then with 14–71 Hz bandpass filtering (row 2 for LSTM, row 4 for EEGChannelNet).[1] Note that without filtering,

---

[1]Results were obtained using the code for LSTM and EEGChannelNet from `https://github.com/perceivelab/eeg_visual_classification`, modified slightly to support z-scoring, plotting the
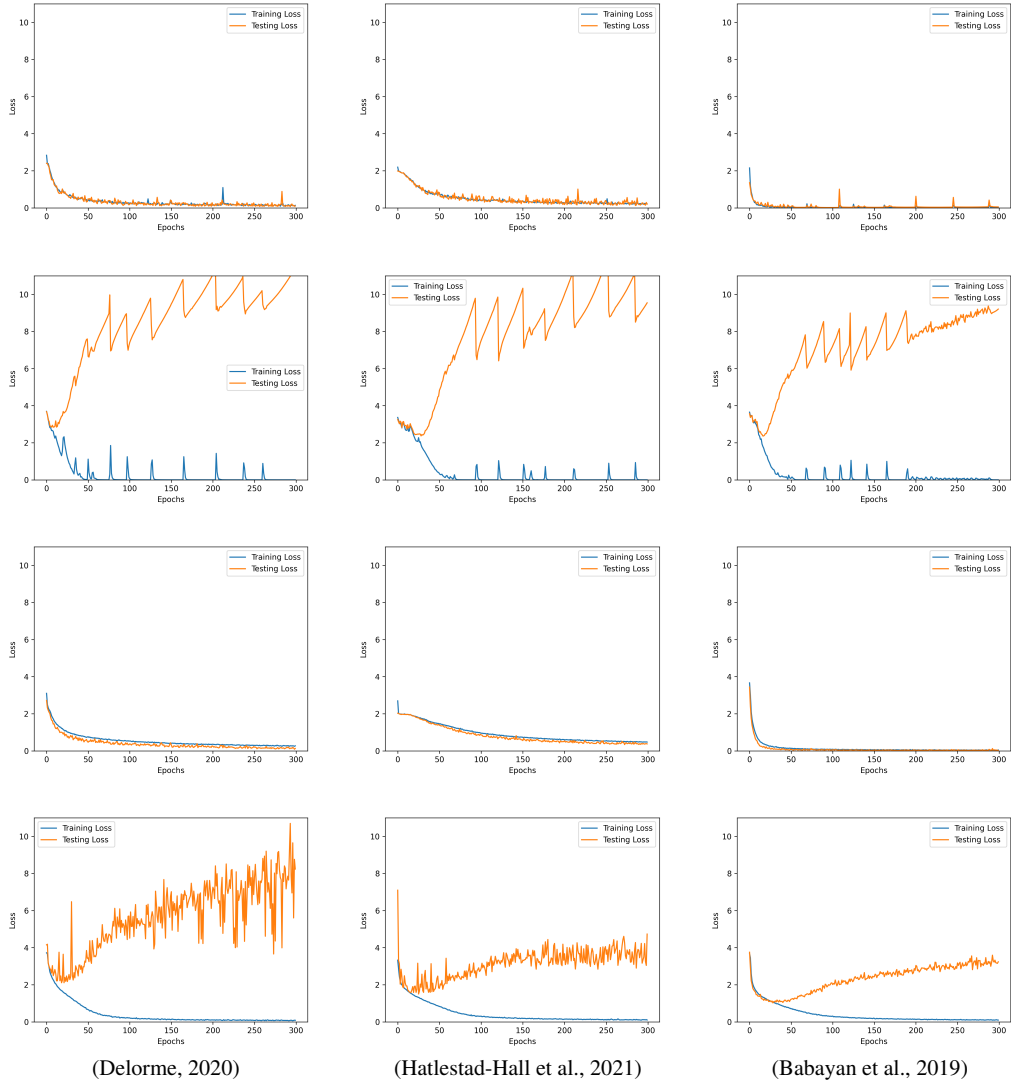
3

Figure 1: Plots of loss *vs.* epoch for the three datasets, for the LSTM classifier (rows 1 and 2) and the EEGChannelNet classifier (rows 3 and 4), averaged over splits, without filtering (rows 1 and 3) and with filtering (rows 2 and 4). Training in blue, testing in orange.
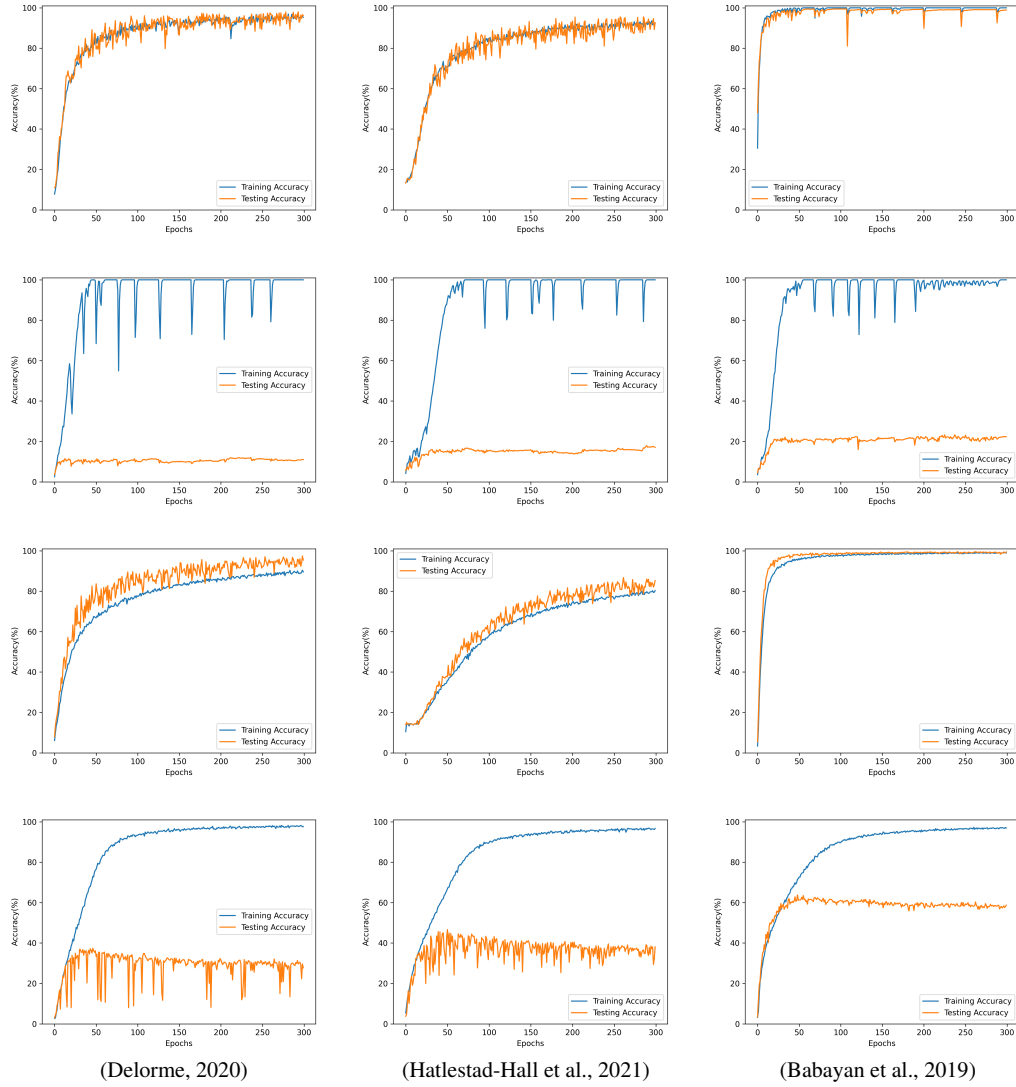
4

216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269

(Delorme, 2020)   (Hatlestad-Hall et al., 2021)   (Babayan et al., 2019)

Figure 2: Plots of accuracy *vs.* epoch for the three datasets, for the LSTM classifier (rows 1 and 2) and the EEGChannelNet classifier (rows 3 and 4), averaged over splits, without filtering (rows 1 and 3) and with filtering (rows 2 and 4). Training in blue, testing in orange.

5

classification accuracy is near perfect for both LSTM and EEGChannelNet. While classification accuracy is much lower with filtering, it is still far above chance (2.5% for Delorme 2020 and Babayan et al. 2019, 3.6% for Hatlestad-Hall et al. 2021). Also note that with filtering, EEGChannelNet yields considerably higher classification accuracy than LSTM. This suggests that EEGChannelNet particularly focuses on the confound, even more so than LSTM. This, together with prior results demonstrating chance performance of LSTM and EEGChannelNet on nonconfounded data (Li et al., 2021; Ahmed et al., 2021; 2022; Bharadwaj et al., 2023), suggests that LSTM and EEGChannelNet focus particularly on the confound, and not the stimulus-related brain activity, calling all results in Spampinato et al. (2017), Kavasidis et al. (2017), and Palazzo et al. (2017; 2018; 2020a;b; 2021; 2024) into question.

5 DISCUSSION

¡The results clearly demonstrate that these three public datasets exhibit a temporal correlation. Note that we make no claim that the studies underlying these datasets exhibit a confound; we are clearly using these datasets for a different purpose than they were intended. As pointed out by Li et al. (2021), one must be careful when repurposing data not to introduce a confound. We do so here only to demonstrate that temporal correlation appears to be pervasive in EEG data, not just the data collected in Spampinato et al. (2017). By design, our repurposing demonstrates that it exists even in short runs. The strong correlation between training and test, for both loss and accuracy, as a function of epoch throughout the entire training regimen clearly indicates label leakage.

The results also clearly demonstrate that the temporal correlation is not removed by bandpass filtering. High-pass filtering removes low-frequency components, not slow spectral change. One way to understand this is the fact that broadcast radio works by modulating an HF carrier with an audio-frequency signal, tuning a receiver with a bandpass filter to a specific channel, and demodulating the filtered broadcast to recover the audio-frequency signal. This, perhaps, is the most foundational principle of Electrical Engineering.

An extensive body of work is flawed because of this confound. Not only is all work using the dataset from Spampinato et al. (2017) flawed, new datasets (Gou et al., 2024; Pan et al., 2024; Zhu et al., 2024; Qian et al., 2024; Uys, 2019; Shimizu & Srinivasan, 2022; Liu et al., 2024b; Wang et al., 2019; 2020; Ma et al., 2021; Cudlenco et al., 2020; Zheng et al., 2024b; Cavazza et al., 2022; Luvsansambuu et al., 2024; Liu et al., 2023; Bai et al., 2023; Parekh et al., 2017) have been collected with the same confounded design. Ma et al. (2021) states:

> to improve the problem of low signal-to-noise ratio of EEG signals, the same category of images were presented successively to bring continuous stimulation to the brain

Bharadwaj et al. (2023) demonstrates how it is possible to increase the signal-to-noise ratio of EEG signals and achieve better classification accuracy without introducing a confound due to temporal correlation.

6 CONCLUSION

It appears that temporal correlation is pervasive in EEG data, even for short runs, where the subject is not fatigued. When conducting experiments with EEG data, one must be careful not to let temporal correlation introduce a confound. Temporal correlation cannot be removed *post hoc* by filtering. The only known way of preventing temporal correlation from introducing a confound is through appropriate randomization that breaks the temporal correlation.

AUTHOR CONTRIBUTIONS

Removed for blind review.

---

loss and accuracy as a function of epoch, and these datasets which have 62 and 64 channels instead of the original 128.

ETHICS STATEMENT

This work debunks nearly one hundred published papers whose results are based on the same confound: a correlation between stimulus class and temporal drift. This confound has been found in eighteen available EEG datasets. Just as with an inconsistent set of axioms one can prove anything, a confounded dataset can be used to support any claim, even ones that are false or absurd. That is what many recent publications based on this confound do: things like generating high fidelity renderings of images, or even 3D CAD models of objects, from EEG recordings.

A research community, knowingly or unknowingly, has discovered that one can use confounded datasets to churn out a plethora of flawed results without reviewers noticing. They have also discovered that one can collect new confounded datasets to churn out even more flawed results without reviewers noticing. The temptation to do this is so strong that the community continues to do so four years after details of the confound were published.

It is conceivable that the flaws in these datasets may be a driving factor behind their frequent reuse. When a dataset is severely confounded, it becomes relatively easy to achieve an extremely high accuracy, which can in turn be used to support sensational claims, and ultimately directs further attention to the dataset. In business, this phenomenon is referred to as "the bad money drives out the good money."

More prominent exposure of these flawed methods and consequent false results will allow resources wasted on continued use of these confounded datasets and flawed methods to be reallocated. The debunked work also causes direct ongoing harm:

- grant proposals can be rejected due to preliminary results not being competitive with results demonstrating falsely-inflated performance based on confounded data or faulty methods;
- manuscripts can be rejected for the same reason;
- grants can be awarded based on false pretenses
- manuscripts can be accepted for the same reason;
- degrees can be awarded for the same reason;
- resources can be wasted attempting to replicate the debunked results;
- resources can be wasted having people read and review flawed papers, and learn flawed methods; and
- because the debunked work relates to brain-computer interfaces—whose primary application is helping people with disabilities (*e.g.*, paralysis) interact with the world—the harm caused is not merely scientific but also medical, with disproportionate impact on people with disabilities.

REPRODUCIBILITY STATEMENT

All code needed to replicate the results presented here will be released upon publication. All data needed to replicate the results presented here are publicly available. The aforementioned code downloads the data.

REFERENCES

Hajar Ahmadieh, Farnaz Gassemi, and Mohammad Hasan Moradi. A hybrid deep learning framework for automated visual image classification using EEG signals. *Neural Computing and Applications*, pp. 1–17, 2023.

Hamad Ahmed, Ronnie B. Wilbur, Hari M. Bharadwaj, and Jeffrey Mark Siskind. Object classification from randomized EEG trials. In *Computer Vision and Pattern Recognition*, pp. 3845–3854, 2021.

Hamad Ahmed, Ronnie B. Wilbur, Hari M. Bharadwaj, and Jeffrey Mark Siskind. Confounds in the data—Comments on "Decoding brain representations by multimodal learning of neural activity

and visual features". *Transactions on Pattern Analysis and Machine Intelligence*, 44(12):9217–9220, 2022.

Jinwon An and Sungzoon Cho. Hand motion identification of grasp-and-lift task from electroencephalography recordings using recurrent neural networks. In *International Conference on Big Data and Smart Computing*, pp. 427–429, 2016.

Anahit Babayan, Miray Erbey, Deniz Kumral, Janis D. Reinelt, Andrea M. F. Reiter, Josefin Röbbig, H. Lina Schaare, Marie Uhlig, Alfred Anwander, Pierre-Louis Bazin, Annette Horstmann, Leonie Lampe, Vadim V. Nikulin, Hadas Okon-Singer, Sven Preusser, André Pampel, Christiane S. Rohr, Julia Sacher, Angelika Thöne-Otto, Sabrina Trapp, Till Nierhaus, Denise Altmann, Katrin Arelin, Maria Blöchl, Edith Bongartz, Patric Breig, Elena Cesnaite, Sufang Chen, Roberto Cozatl, Saskia Czerwonatis, Gabriele Dambrauskaite, Maria Dreyer, Jessica Enders, Melina Engelhardt, Marie Michele Fischer, Norman Forschack, Johannes Golchert, Laura Golz, C. Alexandrina Guran, Susanna Hedrich, Nicole Hentschel, Daria I. Hoffmann, Julia M. Huntenburg, Rebecca Jost, Anna Kosatschek, Stella Kunzendorf, Hannah Lammers, Mark E. Lauckner, Keyvan Mahjoory, Ahmad S. Kanaan, Natacha Mendes, Ramona Menger, Enzo Morino, Karina Näthe, Jennifer Neubauer, Handan Noyan, Sabine Oligschläger, Patricia Panczyszyn-Trzewik, Dorothee Poehlchen, Nadine Putzke, Sabrina Roski, Marie-Catherine Schaller, Anja Schieferbein, Benito Schlaak, Robert Schmidt, Krzysztof J. Gorgolewski, Hanna Maria Schmidt, Anne Schrimpf, Sylvia Stasch, Maria Voss, Annett Wiedemann, Daniel S. Margulies, Michael Gaebler, and Arno Villringer. A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults, 2019. URL https://www.nature.com/articles/sdata2018308.

Yunpeng Bai, Xintao Wang, Yan-pei Cao, Yixiao Ge, Chun Yuan, and Ying Shan. Dreamdiffusion: Generating high-quality images from brain EEG signals. *arXiv*, 2306.16934, 2023.

Ahmed Ben Said, Amr Mohamed, Tarek Elfouly, Khaled Harras, and Z. Jane Wang. Multimodal deep learning approach for joint EEG-EMG data compression and classification. In *Wireless Communications and Networking Conference*, 2017.

Hari M. Bharadwaj, Ronnie B. Wilbur, and Jeffrey Mark Siskind. Still an ineffective method with supertrials/ERPs—Comments on "Decoding brain representations by multimodal learning of neural activity and visual features". *Transactions on Pattern Analysis and Machine Intelligence*, 45 (11):14052–14054, 2023.

Alberto Bozal Chaves. Personalized image classification from EEG signals using deep learning. B.S. thesis, Universitat Politècnica de Catalunya, 2017.

Jacopo Cavazza, Waqar Ahmed, Riccardo Volpi, Pietro Morerio, Francesco Bossi, Cesco Willemse, Agnieszka Wykowska, and Vittorio Murino. Understanding action concepts from videos and brain activity through subjects' consensus. *Scientific Reports*, 12(1):19073, 2022.

Nicolae Cudlenco, Nirvana Popescu, and Marius Leordeanu. Reading into the mind's eye: Boosting automatic visual recognition with EEG signals. *Neurocomputing*, 386:281–292, 2020.

Carlos de la Torre-Ortiz, Michiel M. Spapé, Niklas Ravaja, and Tuukka Ruotsalo. Cross-subject EEG feedback for implicit image generation. *Transactions on Cybernetics*, 54(10):6105–6117, 2024.

Arnaud Delorme. EEG data from an auditory oddball task, 2020. URL https://openneuro.org/datasets/ds003061/versions/1.0.1.

Xia Deng, Shen Chen, Jiale Zhou, and Lei Li. Mind2matter: Creating 3D models from EEG signals. *arXiv*, 2504.11936, 2025.

Changde Du, Changying Du, and Huiguang He. Doubly semi-supervised multimodal adversarial learning for classification, generation and retrieval. In *International Conference on Multimedia*, pp. 13–18, 2019.

Changde Du, Kaicheng Fu, Jinpeng Li, and Huiguang He. Decoding visual neural representations by multimodal learning of brain-visual-linguistic features. *Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10760–10777, 2023.

Changying Du, Changde Du, Xingyu Xie, Chen Zhang, and Hao Wang. Multi-view adversarially learned inference for cross-domain joint distribution matching. In *International Conference on Knowledge Discovery & Data Mining*, pp. 1348–1357, 2018.

Yiping Duan, Shuzhan Hu, Xin Ma, and Xiaoming Tao. Multi-class image generation from EEG features with conditional generative adversarial networks. In *International Conference on Wireless Communications and Signal Processing*, pp. 534–539, 2023.

Xiaoya Fan, Yuntao Liu, and Zhong Wang. Electroencephalogram helps few-shot learning. In *International Conference on Acoustics, Speech and Signal Processing*, pp. 8015–8019, 2024.

Ahmed Fares. A novel spatiotemporal framework for EEG-based visual image classification through signal disambiguation. *Applied System Innovation*, 8(5):121, 2025.

Ahmed Fares, Shenghua Zhong, and Jianmin Jiang. Region level bi-directional deep learning framework for EEG-based image classification. In *International Conference on Bioinformatics and Biomedicine*, pp. 368–373, 2018.

Ahmed Fares, Sheng-hua Zhong, and Jianmin Jiang. Brain-media: A dual conditioned and lateralization supported GAN (DCLS-GAN) towards visualization of image-evoked brain activities. In *International Conference on Multimedia*, pp. 1764–1772, 2020.

Matteo Ferrante, Tommaso Boccato, Stefano Bargione, and Nicola Toschi. Decoding EEG signals of visual brain representations with a CLIP based knowledge distillation. In *Learning from Time Series For Health*, 2024a.

Matteo Ferrante, Tommaso Boccato, Stefano Bargione, and Nicola Toschi. Decoding visual brain representations from electroencephalography through knowledge distillation and latent diffusion models. *Computers in Biology and Medicine*, 178:108701, 2024b.

Honghao Fu, Hao Wang, Jing Jih Chin, and Zhiqi Shen. BrainVis: Exploring the bridge between brain and visual signals via image reconstruction. In *International Conference on Acoustics, Speech and Signal Processing*, pp. 1–5, 2025.

Mingyu Gou, Ying-Jie Zhang, Ren-Jie Dai, Hao-Long Yin, Tianzhen Chen, Fei Cheng, Bao-Liang Lu, Jiang Du, and Wei-Long Zheng. Addressing temporal and auditory factors in meditative EEG with self-supervised learning. In *International Conference on Bioinformatics and Biomedicine*, pp. 1954–1959, 2024.

Haitham S Hasan and Al-Sharqi Mais A. EEG-based image classification using an efficient geometric deep network based on functional connectivity. *Periodicals of Engineering and Natural Sciences*, 11(1):208–215, 2023.

Christoffer Hatlestad-Hall, Trine Waage Rygvold, and Stein Andersson. SRM resting-state EEG, 2021. URL https://openneuro.org/datasets/ds003775/versions/1.0.0.

Sunhee Hwang, Kibeom Hong, Guiyoung Son, and Hyeran Byun. EZSL-GAN: EEG-based zero-shot learning approach using a generative adversarial network. In *International Winter Conference on Brain-Computer Interface*, pp. 1–4, 2019.

Zahra Imani, Mehdi Ezoji, and Timothée Masquelier. Brain-guided manifold transferring to improve the performance of spiking neural networks in image classification. *Journal of Computational Neuroscience*, 51(4):475–490, 2023.

Jianmin Jiang, Ahmed Fares, and Sheng-Hua Zhong. A context-supported deep learning framework for multimodal brain imaging classification. *Transactions on Human-Machine Systems*, 49(6): 611–622, 2019.

Jianmin Jiang, Ahmed Fares, and Sheng-Hua Zhong. A brain-media deep framework towards seeing imaginations inside brains. *Transactions on Multimedia*, 23:1454–1465, 2021.

Zhicheng Jiao, Haoxuan You, Fan Yang, Xin Li, Han Zhang, and Dinggang Shen. Decoding EEG by visual-guided deep neural networks. In *International Joint Conference on Artificial Intelligence*, 2019.

Isaak Kavasidis, Simone Palazzo, Concetto Spampinato, Daniela Giordano, and Mubarak Shah. Brain2Image: Converting brain signals into images. In *International Conference on Multimedia*, pp. 1809–1817, 2017.

Nastaran Khaleghi, Tohid Yousefi Rezaii, Soosan Beheshti, Saeed Meshgini, Sobhan Sheykhivand, and Sebelan Danishvar. Visual saliency and image reconstruction from EEG signals via an effective geometric deep network-based generative adversarial network. *Electronics*, 11(21):3637, 2022.

Pradeep Kumar, Rajkumar Saini, Partha Pratim Roy, Pawan Kumar Sahu, and Debi Prosad Dogra. Envisioned speech recognition using EEG sensors. *Personal and Ubiquitous Computing*, 22(1): 185–199, 2018.

Yu-Ting Lan, Kan Ren, Yansen Wang, Wei-Long Zheng, Dongsheng Li, Bao-Liang Lu, and Lili Qiu. Seeing through the brain: image reconstruction of visual perception from human brain signals. *arXiv*, 2308.02510, 2023.

Pilhyeon Lee, Sunhee Hwang, Seogkyu Jeon, and Hyeran Byun. Subject adaptive EEG-based visual recognition. In *Asian Conference on Pattern Recognition*, pp. 322–334, 2021.

Pilhyeon Lee, Sunhee Hwang, Jewook Lee, Minjung Shin, Seogkyu Jeon, and Hyeran Byun. Inter-subject contrastive learning for subject adaptive EEG-based visual recognition. In *International Winter Conference on Brain-Computer Interface*, pp. 1–6, 2022.

Pilhyeon Lee, Seogkyu Jeon, Sunhee Hwang, Minjung Shin, and Hyeran Byun. Source-free subject adaptation for EEG-based visual recognition. In *International Winter Conference on Brain-Computer Interface*, pp. 1–6, 2023.

Weixian Lei, Yixiao Ge, Kun Yi, Jianfeng Zhang, Difei Gao, Dylan Sun, Yuying Ge, Ying Shan, and Mike Zheng Shou. VIT-LENS: Towards omni-modal representations. In *Computer Vision and Pattern Recognition*, pp. 26637–26647, 2024.

Dan Li, Changde Du, and Huiguang He. Semi-supervised cross-modal image generation with generative adversarial networks. *Pattern Recognition*, 100, 2020.

Ren Li, Jared S. Johansen, Hamad Ahmed, Thomas V. Ilyevsky, Ronnie B. Wilbur, Hari M. Bharadwaj, and Jeffrey Mark Siskind. The perils and pitfalls of block design for EEG classification experiments. *Transactions on Pattern Analysis and Machine Intelligence*, 43(1):316–333, 2021.

Dongjun Liu, Weichen Dai, Hangkui Zhang, Xuanyu Jin, Jianting Cao, and Wanzeng Kong. Brain-machine coupled learning method for facial emotion recognition. *Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10703–10717, 2023.

Dongjun Liu, Jin Cui, Zeyu Pan, Hangkui Zhang, Jianting Cao, and Wanzeng Kong. Machine to brain: facial expression recognition using brain machine generative adversarial networks. *Cognitive Neurodynamics*, 18(3):863–875, 2024a.

Xuan-Hao Liu, Yan-Kai Liu, Yansen Wang, Kan Ren, Hanwen Shi, Zilong Wang, Dongsheng Li, Bao-Liang Lu, and Wei-Long Zheng. EEG2video: Towards decoding dynamic visual perception from EEG signals. In *Advances in Neural Information Processing Systems*, pp. 72245–72273, 2024b.

Yanfang Long, Wanzeng Kong, Xuanyu Jin, Jili Shang, and Can Yang. Visualizing emotional states: A method based on human brain activity. In *Human Brain and Artificial Intelligence*, pp. 248–258, 2019.

Eleonora Lopez, Luigi Sigillo, Federica Colonnese, Massimo Panella, and Danilo Comminiello. Guess what I think: Streamlined EEG-to-image generation with latent diffusion models. In *International Conference on Acoustics, Speech and Signal Processing*, pp. 1–5, 2025.

Uurtsaikh Luvsansambuu, Tengis Tserendondog, Munkbayar Bat-Erdene, and Batmunkh Amar. A deep learning model for classifying the thoughts of multiple individuals based on visual event. In *International Conference on Electrical, Computer and Energy Technologies*, pp. 1–6, 2024.

Xin Ma, Yiping Duan, Shuzhan Hu, Xiaoming Tao, and Ning Ge. EEG based visual classification with multi-feature joint learning. In *International Conference on Image Processing*, pp. 264–268, 2021.

Tariq Mehmood, Hamza Ahmad, Muhammad Haroon Shakeel, and Murtaza Taj. CATVis: Context-aware thought visualization. *arXiv*, 2507.11522, 2025.

Abhijit Mishra, Shreya Shukla, Jose Torres, Jacek Gwizdka, and Shounak Roychowdhury. Thought2text: Text generation from EEG signal using large language models (llms). *arXiv*, 2410.07507, 2024.

Alankrit Mishra. Enhancing machine vision using human cognition from EEG analysis. Master's thesis, Lakehead University, 2022.

Alankrit Mishra, Nikhil Raj, and Garima Bajwa. EEG-based image feature extraction for visual classification using deep learning. In *International Conference on Intelligent Data Science Technologies and Applications*, pp. 181–188, 2022.

Liangyan Mo, Yuhan Wang, Wenhui Zhou, Xingfa Shen, and Wanzeng Kong. A bi-LSTM based network with attention mechanism for EEG visual classification. In *International Conference on Unmanned Systems*, pp. 858–863, 2021.

Pranay Mukherjee, Abhirup Das, Ayan Kumar Bhunia, and Partha Pratim Roy. Cogni-Net: Cognitive feature learning through deep visual perception. In *International Conference on Image Processing*, pp. 4539–4543, 2019.

Tat'y Mwata-Velu, Erik Zamora, Juan Irving Vasquez-Gomez, Jose Ruiz-Pinales, and Humberto Sossa. Multiclass classification of visual electroencephalogram based on channel selection, minimum norm estimation algorithm, and deep network architectures. *Sensors*, 24(12):3968, 2024.

Huyen Ngo, Khoi Do, Duong Nguyen, Viet Dung Nguyen, and Lan Dang. How homogenizing the channel-wise magnitude can enhance EEG classification model? *arXiv*, 2407.20247, 2024.

Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, and Mubarak Shah. Generative adversarial networks conditioned by brain signals. In *International Conference on Computer Vision*, pp. 3410–3418, 2017.

Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, and Mubarak Shah. Decoding brain representations by multimodal learning of neural activity and visual features. *arXiv*, 1820.10974v1, 2018.

Simone Palazzo, Francesco Rundo, Sebastiano Battiato, Daniela Giordano, and Concetto Spampinato. Visual saliency detection guided by neural signals. In *International Conference on Automatic Face and Gesture Recognition*, pp. 434–440, 2020a.

Simone Palazzo, Concetto Spampinato, Joseph Schmidt, Isaak Kavasidis, Daniela Giordano, and Mubarak Shah. Correct block-design experiments mitigate temporal correlation bias in EEG classification. *arXiv*, 2012.03849, 2020b.

Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, Joseph Schmidt, and Mubarak Shah. Decoding brain representations by multimodal learning of neural activity and visual features. *Transactions on Pattern Analysis and Machine Intelligence*, 43(11):3833–3849, 2021.

Simone Palazzo, Concetto Spampinato, Isaak Kavasidis, Daniela Giordano, Joseph Schmidt, and Mubarak Shah. Rebuttal to "Comments on 'Decoding brain representations by multimodal learning of neural activity and visual features' ". *Transactions on Pattern Analysis and Machine Intelligence*, 46(12):11540–11542, 2024.

Hongguang Pan, Zhuoyi Li, Yunpeng Fu, Xuebin Qin, and Jianchen Hu. Reconstructing visual stimulus representation from EEG signals based on deep visual representation model. *Transactions on Human-Machine Systems*, 54(6):711–722, 2024.

Viral Parekh, Ramanathan Subramanian, Dipanjan Roy, and CV Jawahar. An EEG-based image annotation system. In *National Conference on Computer Vision, Pattern Recognition, Image Processing, and Graphics*, pp. 303–313, 2017.

Tanya Piplani, Nick Merill, and John Chuang. Faking it, making it: Fooling and improving brain-based authentication with generative adversarial networks. In *International Conference on Biometrics Theory, Applications and Systems*, 2018.

Dongguan Qian, Hong Zeng, Wenjie Cheng, Yu Liu, Taha Bikki, and Jianjiang Pan. NeuroDM: Decoding and visualizing human brain activity with EEG-guided diffusion model. *Computer Methods and Programs in Biomedicine*, 251:108213, 2024.

Rylan Schaeffer, Joshua Kazdan, Yegor Denisov-Blanch, Brando Miranda, Matthias Gerstgrasser, Susan Zhang, Andreas Haupt, Isha Gupta, Elyas Obbad, Jesse Dodge, Jessica Zosa Forde, Francesco Orabona, Sanmi Koyejo, and David Donoho. Position: Machine learning conferences should establish a "refutations and critiques" track. *arXiv*, 2506.19882, 2025.

Bastian Scharnagl and Christian Groth. Evaluation of different deep learning approaches for EEG classification. In *International Conference on Artificial Intelligence for Industries*, pp. 42–47, 2022.

Hirokatsu Shimizu and Ramesh Srinivasan. Improving classification and reconstruction of imagined images from EEG signals. *Plos one*, 17(9):e0274847, 2022.

Prajwal Singh, Pankaj Pandey, Krishna Miyapuram, and Shanmuganathan Raman. EEG2IMAGE: Image reconstruction from EEG brain signals. In *International Conference on Acoustics, Speech and Signal Processing*, pp. 1–5, 2023.

Prajwal Singh, Dwip Dalal, Gautam Vashishtha, Krishna Miyapuram, and Shanmuganathan Raman. Learning robust deep visual representations from EEG brain recordings. In *Winter Conference on Applications of Computer Vision*, pp. 7538–7547, 2024.

Pushapdeep Singh, Jyoti Nigam, Medicherla Vamsi Krishna, Arnav Bhavsar, and Aditya Nigam. EAD: An EEG adapter for automated classification. *arXiv*, 2505.23107, 2025.

Yonghao Song, Bingchuan Liu, Xiang Li, Nanlin Shi, Yijun Wang, and Xiaorong Gao. Decoding natural images from EEG for object recognition. *arXiv*, 2308.13234, 2023.

Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Mubarak Shah, and Nasim Souly. Deep learning human mind for automated visual classification. *arXiv*, 1609.00344, 2016.

Concetto Spampinato, Simone Palazzo, Isaak Kavasidis, Daniela Giordano, Nasim Souly, and Mubarak Shah. Deep learning human mind for automated visual classification. In *Computer Vision and Pattern Recognition*, pp. 6809–6817, 2017.

Jiajia Tang, Yutao Yang, Qibin Zhao, Yu Ding, Jianhai Zhang, Yang Song, and Wanzeng Kong. Visual-guided dual-spatial interaction network for fine-grained brain semantic decoding. *Transactions on Instrumentation and Measurement*, 73:1–14, 2024.

Praveen Tirupattur, Yogesh Singh Rawat, Concetto Spampinato, and Mubarak Shah. ThoughtViz: Visualizing human thoughts using generative adversarial network. In *International Conference on Multimedia*, pp. 950–958, 2018.

Pieter Johannes Uys. Image classification from EEG brain signals using machine learning and deep learning techniques. Master's thesis, Stellenbosch University, 2019.

Kanan Wahengbam, Kshetrimayum Linthoinganbi Devi, and Aheibam Dinamani Singh. Fortifying brain signals for robust interpretation. *Transactions on Network Science and Engineering*, 10(2): 742–753, 2023.

Fang Wang, Sheng Hua Zhong, Jianfeng Peng, Jianmin Jiang, and Yan Liu. Data augmentation for EEG-based emotion recognition with deep convolutional neural networks. *Lecture Notes in Computer Science*, 10705:82–93, 2018.

Pan Wang, Danlin Peng, Ling Li, Liuqing Chen, Chao Wu, Xiaoyi Wang, Peter Childs, and Yike Guo. Human-in-the-loop design with machine learning. In *International Conference on Engineering Design*, pp. 2577–2586, 2019.

Pan Wang, Shuo Wang, Danlin Peng, Liuqing Chen, Chao Wu, Zhen Wei, Peter Childs, Yike Guo, and Ling Li. Neurocognition-inspired design with machine learning. *Design science*, 6:e33, 2020.

Xin Xiang, Wenhui Zhou, and Guojun Dai. Electroencephalography-driven three-dimensional object decoding with multi-view perception diffusion. *Engineering Applications of Artificial Intelligence*, 156:111180, 2025.

Guangyu Yang and Jinguo Liu. A new framework combining diffusion models and the convolution classifier for generating images from EEG signals. *Brain Sciences*, 14(5):478, 2024.

Zesheng Ye, Lina Yao, Yu Zhang, and Sylvia Gustin. Self-supervised cross-modal visual retrieval from brain activities. *Pattern Recognition*, 145:109915, 2024.

Hong Zeng, Nianzhang Xia, Dongguan Qian, Motonobu Hattori, Chu Wang, and Wanzeng Kong. DM-RE2I: A framework based on diffusion model for the reconstruction from EEG to image. *Biomedical Signal Processing and Control*, 86:105125, 2023a.

Hong Zeng, Nianzhang Xia, Ming Tao, Deng Pan, Haohao Zheng, Chu Wang, Feifan Xu, Wael Zakaria, and Guojun Dai. DCAE: A dual conditional autoencoder framework for the reconstruction from EEG into image. *Biomedical Signal Processing and Control*, 81:104440, 2023b.

Wenxiang Zhang and Qingshan Liu. Using the center loss function to improve deep learning performance for EEG signal classification. In *International Conference on Advanced Computational Intelligence*, pp. 578–582, 2018.

X. Zhang, L. Yao, Q. Z. Sheng, S. S. Kanhere, T. Gu, and D. Zhang. Converting your thoughts to texts: Enabling brain typing via deep feature learning of EEG signals. In *International Conference on Pervasive Computing and Communications*, 2018.

Xiang Zhang, Lina Yao, Dalin Zhang, Xianzhi Wang, Quan Z. Sheng, and Tao Gu. Multi-person brain activity recognition via comprehensive EEG signal analysis. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services*, 2017.

Linfeng Zheng, Peilin Chen, and Shiqi Wang. EidetiCom: A cross-modal brain-computer semantic communication paradigm for decoding visual perception. *arXiv*, 2407.14936, 2024a.

Xiao Zheng and Wanzhong Chen. An attention-based bi-LSTM method for visual object classification via EEG. *Biomedical Signal Processing and Control*, 63, 2021.

Xiao Zheng, Wanzhong Chen, Mingyang Li, Tao Zhang, Yang You, and Yun Jiang. Decoding human brain activity with deep learning. *Biomedical Signal Processing and Control*, 56, 2020a.

Xiao Zheng, Wanzhong Chen, Yang You, Yun Jiang, Mingyang Li, and Tao Zhang. Ensemble deep learning for automated visual classification using EEG signals. *Pattern Recognition*, 102, 2020b.

Xu Zheng, Ling Wang, Kanghao Chen, Yuanhuiyi Lyu, Jiazhou Zhou, and Lin Wang. EIT-1M: One million EEG-image-text pairs for human visual-textual recognition and more. *arXiv*, 2407.01884, 2024b.

Saisai Zhong, Yadong Liu, Zongtan Zhou, and Dewen Hu. ELSTM-based visual decoding from singal [sic]-trial EEG recording. In *International Conference on Software Engineering and Service Science*, pp. 1139–1142, 2018.

Shuqi Zhu, Ziyi Ye, Qingyao Ai, and Yiqun Liu. EEG-ImageNet: An electroencephalogram dataset and benchmarks with image visual stimuli of multi-granularity labels. *arXiv*, 2406.07151, 2024.