Neuroscience 588 (2025) 152–159

Contents lists available at ScienceDirect

Neuroscience

journal homepage: www.elsevier.com/locate/nsc

Research article

A deep learning approach to artifact removal in Transcranial Electrical Stimulation: *From shallow methods to deep neural networks and state space models*

Miguel Fernandez-de-Retana [a,*], Pablo Matanzas-de-Luis [a], Javier Peña [b], Aitor Almeida [a]

[a] Faculty of Engineering, University of Deusto, Av. de las Universidades, 24, Bilbao, Spain
[b] Faculty of Health Sciences, University of Deusto, Av. de las Universidades, 24, Bilbao, Spain

ARTICLE INFO

Keywords:
Bioinformatics
Deep learning (DL)
Electroencephalogram (EEG)
EEG denoising
Noise filtering
State space models (SSM)
Transcranial Electrical Stimulation (tES)

ABSTRACT

Transcranial Electrical Stimulation (tES) is a non-invasive neuromodulation technique that generates artifacts in simultaneous EEG recordings, hindering brain activity analysis. This study analyzes Machine Learning (ML) methods for tES noise artifact removal across three stimulation types: tDCS, tACS, and tRNS. Synthetic datasets were created by combining clean EEG data with synthetic tES artifacts. Eleven artifact removal techniques were tested and evaluated using the Root Relative Mean Squared Error (RRMSE) in the temporal and spectral domains, and the Correlation Coefficient (CC). Results indicate that method performance is highly dependent on stimulation type: for tDCS, a convolutional network (Complex CNN) performed best; while a multi-modular network (M4) based on State Space Models (SSMs) yielded the best results for tACS and tRNS. This study provides guidelines for selecting efficient artifact removal methods for different tES modalities, establishing a benchmark for future research in this area and paving the way for more robust analysis of neural dynamics in advanced clinical and neuroimaging applications.

Introduction

The introduction and development of non-invasive Transcranial Electrical Stimulation (tES) techniques have offered researchers and clinicians a valuable tool for modulating the activity of cerebral regions in humans (Yavari et al., 2018). This, in turn, has facilitated the exploration of brain-behavior relationships and the development of treatments for neurological and psychiatric disorders (Camacho-Conde et al., 2022) as well as cognitive enhancement in healthy population (Simonsmeier et al., 2018). More specifically, tES is a relatively recent technique with a wide range of applications and it works by injecting small amounts of current into the scalp via rubber electrodes that are enclosed in saline soaked sponges (Kohli and Casson, 2019). These currents create an electrical field that modulates neuronal activity based on the duration, intensity, and modality of the application and there are three main types of tES: transcranial Direct Current Stimulation (tDCS), transcranial Alternating Current Stimulation (tACS) and Random noise Stimulation (tRNS). In tDCS, a small direct current is applied between anodal (positive) and cathodal (negative) electrodes placed on the surface of the head to target specific brain areas located beneath the electrodes (Nitsche and Paulus, 2000). In tACS, an alternating current (AC) with a predetermined frequency is passed between anodal and cathodal electrodes, typically set within the EEG frequency spectrum (1–100 Hz) (Antal et al., 2008). Finally, in tRNS, the brain areas beneath both electrodes are stimulated with a current whose amplitude varies randomly in time within the frequency range of 100–640 Hz (Terney et al., 2008).

Nevertheless, our understanding of the effects induced by tES on neural activity remains limited, particularly concerning its impact on neural networks (Miniussi et al., 2012). Employing an approach like combining electroencephalography (EEG) with tES could provide deeper insights into the neural mechanisms underlying these observed changes. Electroencephalography (EEG) is a non-invasive neuroimaging method that records postsynaptic potentials resulting from the electrical activity of cortical neurons by placing electrodes on the scalp (Cohen, 2017). However, simultaneous EEG recording during stimulation via tES is currently not feasible due to the presence of stimulation artifacts in the EEG recordings. These artifacts occur during tES, are several orders of magnitude larger than the signals of interest (Kasten et al., 2018) and prevent the direct analysis of EEG data collected concurrently. Therefore, research combining tES and EEG has been mainly restricted to behavioral and after effect studies (Thut et al., 2017).

* Corresponding author.
E-mail addresses: m.fernandezderetana@deusto.es (M. Fernandez-de-Retana), aitor.almeida@deusto.es (A. Almeida).

https://doi.org/10.1016/j.neuroscience.2025.10.004
Received 25 March 2025; Accepted 5 October 2025
Available online 14 October 2025

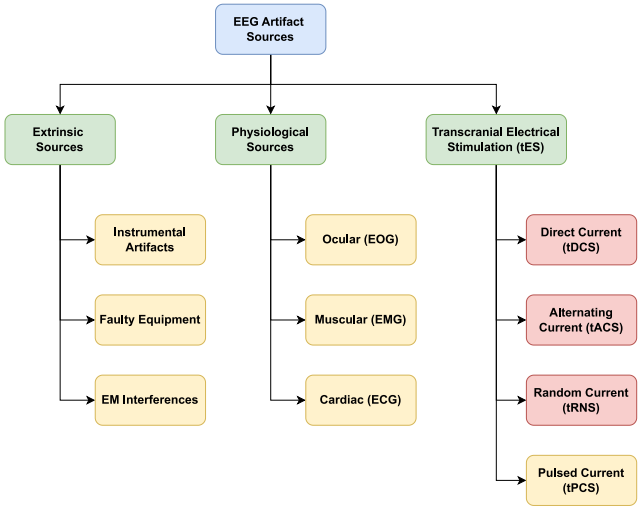

Fig. 1. Overview of Sources of Noisy Artifacts in EEG Signals. Categorization of the different sources of noisy artifacts in EEG (in red, the sources of noise studied in this work). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

While the work done addressing the filtering of EOG and EMG artifacts is extensive, in the case of the transcranial Electrical Stimulation (tES), efforts have been more limited. The generated artifacts are highly dependent on the used stimulation technique: transcranial Direct Current Stimulation (tDCS) uses a constant current; transcranial Alternating Current Stimulation (tACS) uses a sinusoidal oscillating currents, and transcranial Random Noise Stimulation (tRNS) uses randomly generated currents. Baxter et al. (2014) worked on isolating the low-frequency noise artifacts manifested during tDCS, applying Independent Component Analysis (ICA) methods . In the case of tACS, notch filter methods can be not used to filter artifacts due to the higher complexity of the problem, as discussed by Kohli and Casson (2019). The authors propose two different approaches to remove tACS artifacts from EEG recordings, based on Superposition of Moving Averages and Adaptive Filtering, which is an approach shared by Żebrowska et al. (2020). Kohli and Casson (2020) also studied machine learning (ML) methods (a Linear Discriminant Analysis classifier) to filter tACS artifacts, validating the use of ML based approaches. Haslacher et al. (2021) propose also a method to filter artifacts in tACS, Stimulation Artifact Source Separation, a signal decomposition algorithm for separating electric brain activity and stimulation signal artifacts. Barban et al. (2021) provide a comparison of four different algorithms (canonical correlation analysis, infomax, fastICA and independent vector analysis) applied also to tACS. Given the distinct temporal and spectral signatures introduced by each stimulation type, artifact removal cannot be treated as a modality-agnostic task. Each noise modality presents unique signal contamination characteristics, which in turn demand tailored processing strategies. This motivates our decision to evaluate denoising methods separately for each type of stimulation, and to design synthetic artifacts that faithfully reflect the modality-specific characteristics observed in physiological recordings. Finally, to summarize the wide field of EEG denoising under noise artifacts, Fig. 1 provides a comprehensive categorization of the different sources of noise, while Fig. 2 provides an overview of the different approaches (from traditional methods to neural network-based models).

In this paper, we present an in-depth analysis of eleven methods for tES noise artifact removal in simultaneous EEG recordings, including a novel State Space Model (SSM) approach proposed by the authors, and analyze their performance on three types of stimulation methods (tDCS, tACS and tRNS). To our knowledge, this study presents the most extensive analysis of tES filtering approach, encompassing all three main types of possible stimulation methods, while existing studies focus on one. This study accomplishes two primary objectives: it provides a guideline for researchers wanting to filter artifacts from their recordings, allowing to select the most efficient methods in each case; and it establishes a solid ground-truth for future artifact removal research.

Materials and methods

In order to carry out the evaluation of the different denoising methods, synthetic datasets were generated for each of the three types of transcranial electrical stimulation (tES) of interest: *direct* current (tDCS), *alternating* current (tACS), and *random* current (tRNS). The synthetic datasets were created by combining *clean* EEG data from the widely-adopted `EEGdenoiseNet` benchmarking dataset (Zhang et al., 2021b) with synthetic artificial noise. The `EEGdenoiseNet` dataset consists of 4514 single-channel expertly-cleaned EEG recordings — each 2 s long and sampled at 256 Hz — which were used as *ground-truth* for our models.

In this section, we first describe the process of generating the synthetic tES artifacts using a 4th-order linear system approximation of the voltage. Then, we present the mixing procedure to combine the clean EEG data with the synthetic artifacts. Finally, we review the different denoising approaches and architectures used in this study.

Synthetic tES artifact generation

In the context of transcranial stimulation, the electrical current — whether it is direct, alternating, or random — is applied to the scalp, and the resulting electric field penetrates the skull and reaches the brain. The electric field generated by the stimulation can be modeled as a voltage source, which is then distributed through the brain tissue. The voltage distribution can be approximated by a linear system (Hahn et al., 2013), which can be represented by the 4th-order system in Eq. (1):

$$V(t) = C_1 + C_2 I(t) + C_3 \dot{I}(t) + C_4 \int I(t)\,dt + C_5 \iint I(t)\,dt\,dt \tag{1}$$

where $C_1 = 3.99$, $C_2 = 3.64$, $C_3 = 1.395$, $C_4 = -9.92 \cdot 10^{-3}$, and $C_5 = 3.35 \cdot 10^{-5}$ respectively. The current $I(t)$ is the stimulation current, and the voltage $V(t)$ is the resulting electric field which is then used to generate the synthetic *noisy* EEG artifacts.

The stimulation currents $I(t)$ are generated as follows:

- **tDCS**: A constant current $I(t) = I_0$ in mA is applied for the duration of the EEG recording. The voltage is thus approximated as follows:

$$V(t) = C_1 + C_2 I_0 + C_4 I_0 t + \frac{1}{2} C_5 I_0 t^2 \tag{2}$$

- **tACS**: A sinusoidal current of amplitude A and frequency f is applied for the duration of the EEG recording:

$$I(t) = A \sin(2\pi f t) \tag{3}$$

- **tRNS**: A random current is applied such that the current intensities are normally distributed with 99% of the values lying between the peak-to-peak amplitude A:

$$I(t) \sim \mathcal{N}(\mu, \sigma) \tag{4}$$

where $\mu = 0$ and $\sigma = A/3$.

Signal mixing procedure

Once generated, the synthetic tES artifacts are mixed with the clean EEG data by the following procedure:

1. First, we split the reference 2-second long EEG signals as well as the synthetic tES artifacts into three sets: a training set (80%), a validation set (10%), and a test set (10%).

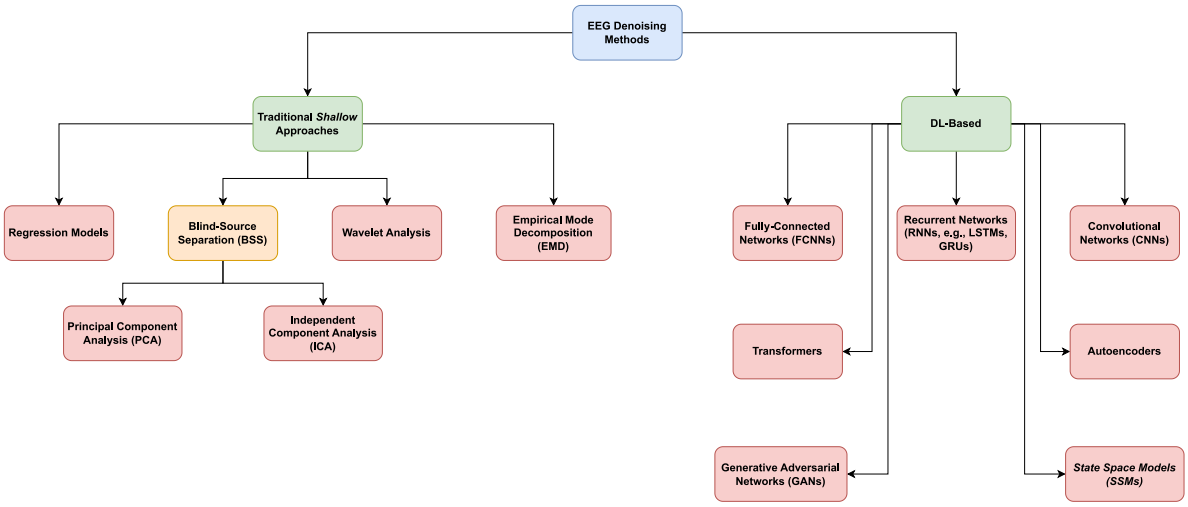

Fig. 2. Overview of Methods for Denoising Electroencephalogram (EEG) Signals. *(left)* Traditional statistical methods *(right)* Neural network architectures for removing noisy artifacts. Includes our novel State Space Model (SSM)-based approach.

2. Then, within each of the sets, pairs of clean EEG and tES artifacts are randomly combined by linearly mixing the sequences according to Eq. (5):

$$\eta(s) = s + \lambda \cdot n \tag{5}$$

where s is the clean EEG signal (*ground-truth*), $\eta(s)$ is the mixed *noisy* EEG signal, λ controls the signal-to-noise ratio (SNR), and n is the synthetic tES artifact signal.

3. The mixing procedure is repeated for each of the three tES methods (i.e., tDCS, tACS, and tRNS) at ten uniform SNR levels (ranging from −7 dB to 2 dB) according to Eq. (6):

$$\text{SNR} = 10 \log \frac{\text{RMS}(s)}{\text{RMS}(\lambda \cdot n)} \tag{6}$$

where the Root-Mean Squared (RMS) of a segment s is defined as:

$$\text{RMS}(s) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} s_i^2} \tag{7}$$

where N is the sequence length of s.

Learning process

In terms of the learning process of our denoising models, given a pair of noisy/clean EEG signals (x, y), where $x = \eta(y)$ denotes the linearly-mixed *noisy* EEG, and y denotes the original *ground-truth* signal. In order to improve the denoising capabilities of the models, we normalize the signals according to Eq. (8):

$$\tilde{x} = \frac{x}{\sigma_x} \qquad \tilde{y} = \frac{y}{\sigma_x} \tag{8}$$

where σ_x is the standard deviation of the noisy EEG signal x.

Our goal is to train the denoising models to learn a non-linear mapping $\mathcal{F}(\cdot)$ between the contaminated EEG signal $\tilde{x}$ and the clean EEG signal $\tilde{y}$:

$$\hat{y} = \mathcal{F}(\tilde{x}; \boldsymbol{\theta}) \tag{9}$$

where $\hat{y}$ denotes the estimated *denoised* EEG signal, and $\boldsymbol{\theta}$ denotes the learnable parameters of the model.

The learning process is carried out by minimizing the Mean Squared Error (MSE) loss between the estimated denoised EEG signal $\hat{y}$ and the original clean EEG signal $\tilde{y}$ as in Eq. (10):

$$\mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} \left\| \hat{y}_i - \tilde{y}_i \right\|_2^2 = \frac{1}{N}\sum_{i=1}^{N} \left\| \mathcal{F}(\tilde{x}_i; \boldsymbol{\theta}) - \tilde{y}_i \right\|_2^2 \tag{10}$$

Our objective is thus to find the optimal set of parameters $\boldsymbol{\theta}^*$ that minimizes the loss function $\mathcal{L}_{\text{MSE}}(\boldsymbol{\theta})$ in Eq. (10).

$$\boldsymbol{\theta}^* = \underset{\boldsymbol{\theta}}{\text{argmin}}\, \mathcal{L}_{\text{MSE}}(\boldsymbol{\theta}) \tag{11}$$

For this purpose, we used the Adam optimizer (Kingma and Ba, 2014) with a learning rate of $\alpha = 1 \cdot 10^{-4}$, and trained the models with a batch size of 128. To increase the statistical power of our benchmarks, we trained, validated and tested independently for 10 times with randomly generated signal combinations.

All our models were implemented in Python 3.10 with PyTorch 2.1, running on a computer with two NVIDIA RTX A6000 GPUs. For pre-published models for which open-source code has been released publicly, we used their original implementation. Finally, the complete workflow followed in this study is summarized in Fig. 3.

Denoising models

In this section, we present the different denoising models used in this study. We divide the models into four categories: traditional *shallow* denoisers (i.e., models that are not based on neural networks), the original `EEGdenoiseNet` networks from Zhang et al. (2021b) (i.e., the most widely used benchmark for EEG denoising), alternative denoising models proposed throughout the literature, and finally our prosed novel SSM-based networks.

Traditional shallow models: Empirical Mode Decomposition (EMD)

As a baseline for comparing our *deeper* models, we include in this study a traditional *shallow* method for EEG denoising such as the Empirical Mode Decomposition (EMD) with Mutual Information from Mert and Akan (2014). The EMD method is a simple signal processing technique that decomposes a signal into a finite number of intrinsic mode functions (IMFs) (Huang et al., 1998). The IMFs are obtained by iteratively extracting the local oscillatory modes of the signal as in Algorithm 1. Note that we can then reconstruct the original signal by summing the IMFs and the residual as in Eq. (12):

$$x(t) = \sum_{i=1}^{n} \text{IMF}_i(t) + r(t) \tag{12}$$

where $x(t)$ is the original signal, $\text{IMF}_i(t)$ are the IMFs, n is the total number of extracted IMFs, and $r(t)$ is the monotonic residual.

Once the intrinsic components have been extracted, the Mutual Information (MI) criterion based on Shannon's entropy in Eq. (13) is

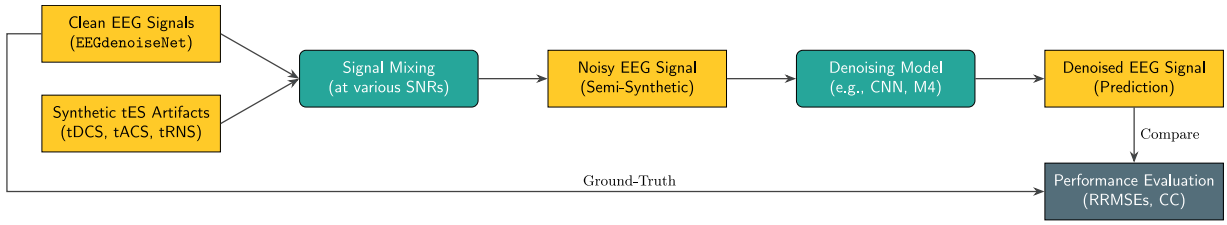

Fig. 3. Experimental Workflow. The diagram illustrates the complete denoising pipeline used in this study, from data generation to model evaluation.

Algorithm 1 Signal Decomposition into Intrinsic Mode Functions (IMFs) (*Sifting* Process)

Input: Signal $x(t)$
Output: Set of IMFs $\{\text{IMF}_1, \text{IMF}_2, \ldots, \text{IMF}_n\}$
Initialize the residual $r(t) = x(t)$
Set IMF counter $i = 1$
while residual $r(t)$ is not a monotonic function **do**
    $h(t) = r(t)$
    repeat
        Identify all local maxima and minima of $h(t)$
        Interpolate the local maxima and minima to form the upper and lower envelopes $e_u(t)$ and $e_l(t)$ respectively
        Compute the mean envelope $m(t) = \frac{e_u(t)+e_l(t)}{2}$
        Update the signal $h(t) = h(t) - m(t)$
    until $h(t)$ is an IMF according to stopping criteria
    Set $\text{IMF}_i = h(t)$
    Update the residual $r(t) = r(t) - \text{IMF}_i$
    Increment IMF counter $i = i + 1$
end while
Return $\{\text{IMF}_1, \text{IMF}_2, \ldots, \text{IMF}_n\}$

Algorithm 2 Empirical Mode Decomposition (EMD) with Mutual Information for EEG Denoising

Input: Noisy EEG signal $x(t)$
Output: Denoised EEG signal $\hat{y}(t)$
Decompose the noisy EEG signal $x(t)$ into n IMFs $\{\text{IMF}_1, \text{IMF}_2, \ldots, \text{IMF}_n\}$ using Algorithm 1
Compute the Autocorrelation Function (ACF) for the original signal $\rho_k(x)$ and the IMFs $\rho_k(\text{IMF}_i)$
For every IMF, compute the Mutual Information (MI) between its ACF and the ACF of the original signal $I_i = I(\rho_k(\text{IMF}_i), \rho_k(x))$ as in Eq. (13)
Normalize the MI values $I_i = \frac{I_i}{\max(I_i)}$
if $I_{max} - I_{min} > 0.8$ **then**
    The signal does not contain additive noise
end if
Determine the threshold $\tau = \frac{1}{2}(I_3 - I_1) + I_1$
Reconstruct the denoised EEG signal $\hat{y}(t)$ as $\hat{y}(t) = \sum_j \text{IMF}_j, j = \{i \mid I_i > \tau\}$
Return $\hat{y}(t)$

used to select the IMFs that are most informative for the denoising process (i.e., which IMFs are noise-free and can be used to reconstruct the clean EEG). The whole method is summarized in Algorithm 2.

$$H(X) = -\sum_x p(x)\log p(x) \tag{13}$$

$$H(X \mid Y) = -\sum_y p(y)\sum_x p(x \mid y)\log\left[p(x \mid y)\right] \tag{14}$$

$$\begin{aligned} I(X,Y) &= H(X) - H(X \mid Y) \\ &= \sum_x p(x)\sum_y p(x \mid y)\log\left[\frac{p(x \mid y)}{p(x)}\right] \end{aligned} \tag{15}$$

EeGdenoiseNet models

Next, we present the original `EEGdenoiseNet` models proposed by Zhang et al. (2021b). These models are a set of deep learning architectures that have shown *state-of-the-art* performance in EEG denoising under ocular and muscular, and have been widely adopted as a reference benchmark for EEG denoising tasks. The original `EEGdenoiseNet` models consist of four different architectures: **(1)** a Fully-Connected Neural Network (FCNN) with 4 hidden layers, dropout regularization, and ReLU activations; **(2)** a *Simple* Convolutional Neural Network (Simple CNN) with 4 layers of 1D convolutional filters (with 64 filters each); **(3)** a Complex Convolutional Network (Complex CNN) with 4 layers of 1D convolutional filters (with 64 filters each) and residual connections; and **(4)** a recurrent network using LSTM units followed by 3 fully-connected layers with dropout regularization. The models were trained using the Adam optimizer with a learning rate of $1 \cdot 10^{-4}$, and a batch size of 128, as in the original work.

Alternative denoising models

In addition to the aforementioned architectures, we also include in our study a set of alternative denoising models proposed throughout the literature. These models include a Novel CNN originally proposed in Zhang et al. (2021a) for removing muscular artifacts by using an exponentially increasing number of 1D filters in each layer, a Transformer-based network (EEGDnet) proposed by Pu et al. (2022), and a multi-modular convolutional network (MMNN-4) proposed by Zhang et al. (2022) which not only estimates the clean EEG but also outputs the estimated noise. Finally, we also include the Wavelet-CNN network which provides a computationally efficient alternative as it relies only on two single Conv1D layers and a pair of fully-connected layers. Instead of passing the clean EEG signal through the network, the Wavelet-CNN network first extracts the Wavelet decomposition (Sharma et al., 2017; Borse, 2015; Yu, 2009) of the signal, and then feeds the transformed signal to the network for denoising. The architecture of the Wavelet-CNN network is shown in Fig. 4.

SSM-based networks

Finally, we test the performance of our proposed novel SSM-based networks: that is, a *vanilla* implementation of the Mamba architecture (Gu and Dao, 2023), and the more complex Multi-Modular Macro Mamba (M4) architecture in Fig. 5. Theoretically, SSM-based networks leverage the limitations of *traditional* deep sequential models by essentially providing parallelizable training (in the style of CNNs and Transformers) while also delivering constant inference complexity (as in RNNs) (Gu et al., 2021; Gu, 2023). That is, they can be written as a convolution (16):

$$\begin{aligned} \bar{\boldsymbol{K}} &= \left(\boldsymbol{C}\bar{\boldsymbol{B}}, \boldsymbol{C}\bar{\boldsymbol{A}}\bar{\boldsymbol{B}}, \ldots, \boldsymbol{C}\bar{\boldsymbol{A}}^k\bar{\boldsymbol{B}}\right) \\ y &= x * \bar{\boldsymbol{K}} \end{aligned} \tag{16}$$

or as a recurrence (17):

$$\begin{aligned} h_t &= \bar{\boldsymbol{A}}h_{t-1} + \bar{\boldsymbol{B}}x_t \\ y_t &= \boldsymbol{C}h_t \end{aligned} \tag{17}$$

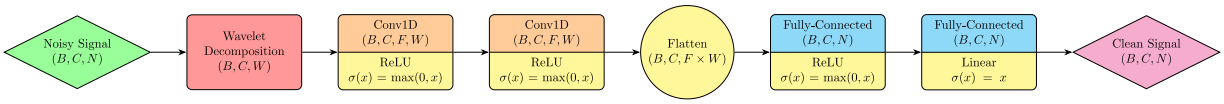

Fig. 4. Architecture of the Wavelet-CNN Network. The model uses the wavelet decomposition of artifact-contaminated EEG as input features. Here, W represents the dimension of the concatenated wavelet coefficients of x. Following two Conv1D layers, the signal is mapped back to its original sequence length N through the fully-connected layers. The output of the network is the denoised EEG signal.

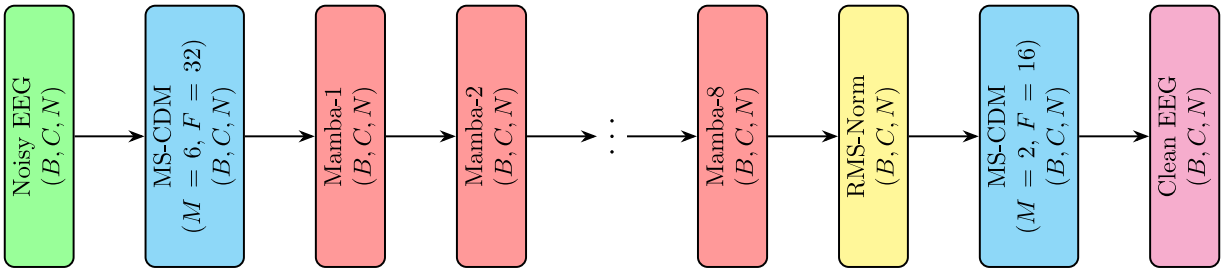

Fig. 5. Architecture of Multi-Modular Macro Mamba (M4). The artifact-contaminated EEG is first processed by an MS-CDM block with 6 CDMs, each containing 32 filters in their Conv1D layers. Next, 8 Mamba blocks are stacked sequentially. The output from the last Mamba block is then normalized and fed into another MS-CDM block with 2 CDMs, each having 16 filters in their Conv1D layers.

Likewise, *selective* SSMs, such as the Mamba blocks used in our models, allow the network to focus on the most relevant information for the denoising task, and to ignore the irrelevant information by employing input-dependent parameters (Gu and Dao, 2023). As a consequence of this, the convolutional model of computation is no longer valid, and thus *scanning* — a hardware efficient recurrence operation — is used.

However, more recently, some authors have argued that in practice SSMs pose a number of limitations. For instance, Dao and Gu (2024) acknowledges that although SSMs are theoretically appealing for their computational complexities, they are not as efficient in practice. In particular, the authors argue that the *hyper-optimization* of modern accelerators such as GPUs and TPUs can be exploited and thus propose a new matrix form of computing SSMs based on structured matrices: i.e., matrices that can be represented in a compact (subquadratic) representation, and that have fast algorithms for matrix–vector operations by operating directly on these compressed representations.

$$y = \boldsymbol{M} x \tag{18}$$

$$\boldsymbol{M} = \boldsymbol{L} \circ \boldsymbol{C} \bar{\boldsymbol{B}}^T \tag{19}$$

$$\boldsymbol{M}_{ij} = \boldsymbol{C}_i^T \bar{\boldsymbol{A}}_i \dots \bar{\boldsymbol{A}}_{j+1} \bar{\boldsymbol{B}}_j \tag{20}$$

where $\boldsymbol{M}$ is the sequence transformation matrix, $\boldsymbol{L}$ is the structured mask matrix, and $\circ$ denotes the Hadamard product (i.e., element-wise multiplication).

Furthermore, Merrill et al. (2024) also explores the relationship between the expressive power of SSMs and Transformers, and Dao and Gu (2024) shows how the structured matrix $\boldsymbol{L}$ can be used to define different types of attention.

Results

In this section, we first present the metrics used to assess the quality of the denoising models on the `EEGdenoiseNet` dataset. Then, we provide a short discussion on the choice of hyperparameters used in the training process. Finally, we present the results of the denoising models on the synthetic datasets generated for each of the three tES methods (tDCS, tACS, and tRNS).

Evaluation metrics

In order to evaluate the performance of the different models we use the same regression metrics as in the original benchmarking dataset (Zhang et al., 2021b). These metrics have become a staple of the EEG denoising literature, and are widely used to compare the performance of different models. The metrics used are the Root Relative Mean Squared Error (RRMSE) for both the temporal and spectral domains (Eqs. (21) and (22) respectively), and the Correlation Coefficient (CC) in Eq. (23) between the estimated and the ground-truth EEG signals:

$$\text{RRMSE}_{temporal} = \frac{\text{RMS}(\hat{y} - y)}{\text{RMS}(y)} \tag{21}$$

$$\text{RRMSE}_{spectral} = \frac{\text{RMS}(\text{PSD}(\hat{y}) - \text{PSD}(y))}{\text{RMS}(\text{PSD}(y))} \tag{22}$$

$$\text{CC} = \frac{\text{Cov}(\hat{y}, y)}{\sqrt{\text{Var}(\hat{y})\,\text{Var}(y)}} \tag{23}$$

where $\hat{y}$ is the estimated EEG signal, y is the *ground-truth* EEG signal, $\text{RMS}(\cdot)$ denotes the Root-Mean Squared error as in Eq. (7), $\text{PSD}(\cdot)$ denotes the Power Spectral Density, $\text{Cov}(\cdot)$ denotes the covariance, and $\text{Var}(\cdot)$ denotes the variance.

In practice, smaller values of the RRMSEs, and values of the CC closed to 1, are preferred as indicating a better match between the estimated and the ground-truth EEG signals.

Hyperparameter configuration

For the hyperparameter configurations of the evaluated models, we adhered to the optimizer, batch-size, number of epochs, and other hyperparameters (e.g., the number of convolutional filters and their kernel size) as specified in their respective publications.

In training the Mamba-based networks, both *vanilla* and Multi-Modular variant, we employed early stopping regularization with a patience of 5 epochs, a batch-size of 128, and the Adam optimizer. The learning rate, $\alpha = 10^{-4}$, was chosen based on a sweep over learning rates ranging from 10^{-1} to 10^{-6}. The $\boldsymbol{\beta}$ coefficients were set to the default values of $\beta_1 = 0.9$ and $\beta_2 = 0.999$. Likewise, for the Mamba blocks, we set the inner dimension (the latent internal space dimension) to match the sequence length. Practically, this means $N = E = 512$, given that our signals are 2 s long with a sampling rate of 256 Hz. Finally, to enhance the statistical robustness of our benchmarks, we performed training, validation, and testing independently 10 times with randomly generated signal combinations. Random weight initialization followed PyTorch's default seed mechanism for each model class. No fixed seed was enforced; instead, 10 distinct random seeds were implicitly used for each run, to support robust estimates across trials.

Experimental results

Table 2 shows the performance of the proposed denoising models in the context of artifact removal from EEG signals under transcranial stimulation. Specifically, the Relative Root Mean Square Error (RRMSEs) — in both temporal and spectral domains —, and the correlation coefficient (CC) between the *ground-truths* and the denoised signals are presented for alternating, direct, and random current stimulations.

To determine the statistical significance of these results, we performed a series of pairwise comparisons. For each stimulation modality and metric, every model was compared against the designated best-performing model (in **bold** in Table 2), serving as our control group. The comparisons were conducted using Welch's t-test on the summary statistics (mean, standard deviation) gathered from the 10 independent runs. To account for multiple comparisons, the Bonferroni correction was applied. A result was deemed not significantly different if the corrected p-value exceeded our significance threshold of $\alpha = 0.05$.

In the case of tDCS, the Empirical Mode Decomposition (EMD) method, a traditional shallow denoiser based on Mutual Information, yielded an RRMSE of 0.313 in the temporal domain and 0.138 in the spectral domain, with a CC of 0.952. In comparison, the Original `EEGdenoiseNet` Networks demonstrated superior performance, with the Fully Connected Neural Network (FCNN) achieving an RRMSE of 0.131 (temporal) and 0.075 (spectral), and a CC of 0.993. Notably, the Simple CNN and Complex CNN models further reduced these errors, with the Complex CNN achieving the lowest RRMSE of 0.021 (temporal) and 0.010 (spectral), and a perfect CC score of 1. Alternative models such as the Multi-Modal Neural Network-4 (MMNN-4) also performed well, with an RRMSE of 0.066 (temporal) and 0.029 (spectral), and a CC of 0.999. The Novel SSM-Based Networks, particularly the *vanilla* Mamba implementation, achieved an RRMSE of 0.042 (temporal) and 0.023 (spectral), with a CC of 1.

Regarding tRNS, the EMD method recorded the highest RRMSE values of 0.648 (temporal) and 0.485 (spectral), and a CC of 0.831. Among the EEGdenoiseNet models, the FCNN achieved an RRMSE of 0.405 (temporal) and 0.317 (spectral), and a CC of 0.904; while the Complex CNN model showed improvements with an RRMSE of 0.514 (temporal) and 0.420 (spectral), and a CC of 0.864. The best performance in this category was observed with the LSTM, which scored a 0.405 (temporal) and 0.306 (spectral) for the RRMSEs, and a CC of 0.903. Additionally, we highlight the performance of the Wavelet-CNN which outperforms both CNN implementations from the original EEGdenoiseNet framework and outperform the NovelCNN in the spectral domain at a fractional computational cost. At last, the M4 model achieved the lowest RRMSE values of 0.340 (temporal) and 0.289 (spectral), and the highest CC of 0.927.

Finally, when removing tACS artifacts, the EMD method exhibited an RRMSE of 0.635 (temporal) and 0.545 (spectral), and a CC of 0.816. The `EEGdenoiseNet` networks once again demonstrated significant enhancements, with the FCNN achieving an RRMSE of 0.211 (temporal) and 0.124 (spectral), and a CC of 0.975. The Simple CNN and Complex CNN models further optimized these metrics, with the Complex CNN achieving an RRMSE of 0.179 (temporal) and 0.106 (spectral), and a CC of 0.979. The M4 model consistently delivered the best results, with an RRMSE of 0.163 (temporal) and 0.127 (spectral), and a CC of 0.989.

Overall, the Complex CNN model demonstrated the best performance for tDCS, yielding the lowest RRMSEs and a *perfect* CC of 1.0. Likewise, for tACS and tRNS artifacts, the Multi-Modular Macro Mamba model consistently outperformed other methods, providing the best combination of low RRMSEs and high CC. Finally, we highlight the remarkable robustness of the Wavelet-CNN model in both the spectral and temporal domains due to its use of wavelet components of the signal as input to the convolutional layers. Likewise, its parallelizability and computational efficiency (i.e., similar results to the other CNN models with a fraction of the weights) might prove of utmost importance in situations such as in-device denoising or computational resource scarcity. A succinct summary of the results is provided in Table 1.

Table 1
Summary of Top-Performing Denoising Models for Each tES Modality. Arrows indicate whether lower (↓) or higher (↑) values are better.

Stimulation	Best Model	RRMSE (T|S) ↓	CC ↑
tDCS	Complex CNN	0.021 | 0.010	1.000
tACS	M4	0.163 | 0.127	0.989
tRNS	M4	0.340 | 0.289	0.927

Discussion

In this study, we have presented a comprehensive evaluation of the performance of different deep learning-based denoising models in the context of artifact removal from EEG signals under transcranial stimulation. We have compared the performance of traditional *shallow* denoisers — such as the Empirical Mode Decomposition (EMD) method using entropy as an information criterion for selecting the most informative Intrinsic Mode Functions (IMFs) —, the original `EEGdenoiseNet` networks, alternative denoising models proposed throughout the literature — including an alternative CNN, a Transformer-based network, a multi-modular convolutional network, and an efficient CNN using wavelets as input features —, and novel SSM-based networks (both a *vanilla* Mamba block and the more complex Multi-Modular Macro Mamba (M4) architecture).

Our results show that, in the presence of constant current stimulation (tDCS), the Complex CNN model from the original `EEGdenoiseNet` framework delivered the best performance metrics, providing the best combination of low RRMSEs and highest CC. However, our statistical analysis demonstrates that it is not a standalone winner; the Simple CNN and the original Mamba model performed at a statistically indistinguishable level ($p > 0.05$) suggesting that researchers can select from a tier of top-performing models for tDCS based on secondary criteria, such as computational complexity or implementation ease.

Conversely, for the more complex, non-stationary artifacts generated by alternating (tACS) and random current stimulation (tRNS), our proposed Multi-Modular Macro Mamba (M4) architecture was the definitive top performer, with the lowest RRMSE values (in both the temporal and spectral domains) and the highest CC. The statistical analysis confirmed the M4 model's superiority, particularly for tRNS. It is particularly noteworthy that the simpler, non-modular *vanilla* Mamba implementation was also highly effective. For tACS artifact removal, its performance was statistically indistinguishable from the more complex M4 architecture, emphasizing the denoising strength of the underlying SSM approach. Finally, we highlight the remarkable robustness of the Wavelet-CNN model in both the spectral and temporal domains due to its use of wavelet components of the signal as input to the convolutional layers.

In more practical terms, the high performance of the M4 architecture in removing complex, non-stationary artifacts generated by tACS and tRNS provides a robust methodology to investigate neural dynamics in neuroimaging applications where analysis was previously confounded by such noise. Likewise, from a computational standpoint, the architecture's foundation on State Space Models lends itself to efficient, parallelizable training, and optimal constant inference; whereas the Wavelet-CNN positions itself as a light-weight, efficient alternative for real-time resource-constrained applications, such as *on-device* denoising.

Furthermore, we must consider the semi-synthetic nature of our data. Our methodology, which superimposes modeled tES artifacts onto real-world, expertly-cleaned EEG signals, is essential for establishing a reproducible benchmark with a definitive ground-truth. Nevertheless, we acknowledge that performance on these datasets, while providing a rigorous and controlled comparison, may not be wholly representative of performance in live clinical settings, which might present additional complexities such as dynamic electrode impedance or other instrumental artifacts.

Table 2
Average EEG Denoising Performance of all SNRs under tACS, tRNS, & tDCS Stimulation. Values are presented as mean ± standard error. For each stimulation type, the best-performing model is shown in **bold** and served as the control for statistical tests. Values marked with ** are not significantly different (Bonferroni-corrected $p > 0.05$) from the control model in that column, based on a pairwise Welch's t-test.

	tDCS			tRNS			tACS		
	RRMSE-Temporal	RRMSE-Spectral	CC	RRMSE-Temporal	RRMSE-Spectral	CC	RRMSE-Temporal	RRMSE-Spectral	CC
Traditional *Shallow* Denoiser									
EMD	$0.313_{\pm 0.001}$	$0.138_{\pm 0.001}$	$0.952_{\pm 0.000}$	$0.648_{\pm 0.005}$	$0.485_{\pm 0.006}$	$0.831_{\pm 0.002}$	$0.635_{\pm 0.005}$	$0.545_{\pm 0.013}$	$0.816_{\pm 0.002}$
Original EEGdenoiseNet Networks									
FCNN	$0.131_{\pm 0.001}$	$0.075_{\pm 0.004}$	$0.993_{\pm 0.000}$	$0.405_{\pm 0.005}$	$0.317_{\pm 0.008}$	$0.904_{\pm 0.001}$	$0.211_{\pm 0.002}$	$0.124^{**}_{\pm 0.001}$	$0.975_{\pm 0.001}$
Simple CNN	$0.021^{**}_{\pm 0.001}$	$0.017^{**}_{\pm 0.003}$	$1.000^{**}_{\pm 0.000}$	$0.600_{\pm 0.011}$	$0.481_{\pm 0.007}$	$0.840_{\pm 0.009}$	$0.189_{\pm 0.003}$	$0.112_{\pm 0.001}$	$0.976_{\pm 0.001}$
Complex CNN	$\mathbf{0.021}_{\pm 0.000}$	$\mathbf{0.010}_{\pm 0.001}$	$\mathbf{1.000}_{\pm 0.000}$	$0.514_{\pm 0.009}$	$0.420_{\pm 0.006}$	$0.864_{\pm 0.004}$	$0.179^{**}_{\pm 0.005}$	$0.106_{\pm 0.003}$	$0.979_{\pm 0.001}$
LSTM	$0.199_{\pm 0.019}$	$0.072_{\pm 0.016}$	$0.981_{\pm 0.003}$	$0.405_{\pm 0.006}$	$0.306^{**}_{\pm 0.008}$	$0.903_{\pm 0.005}$	$0.210_{\pm 0.007}$	$0.101_{\pm 0.007}$	$0.976_{\pm 0.004}$
Alternative Denoising Models									
Novel CNN	$0.195_{\pm 0.007}$	$0.155_{\pm 0.008}$	$0.973_{\pm 0.002}$	$0.357_{\pm 0.004}$	$0.405_{\pm 0.007}$	$0.929^{**}_{\pm 0.001}$	$0.238_{\pm 0.005}$	$0.214_{\pm 0.003}$	$0.964_{\pm 0.001}$
Transformer (EEGDnet)	$0.411_{\pm 0.012}$	$0.335_{\pm 0.007}$	$0.914_{\pm 0.004}$	$0.520_{\pm 0.004}$	$0.408_{\pm 0.001}$	$0.854_{\pm 0.002}$	$0.298_{\pm 0.008}$	$0.179_{\pm 0.005}$	$0.956_{\pm 0.003}$
MMNN-4	$0.066_{\pm 0.003}$	$0.029_{\pm 0.003}$	$0.999_{\pm 0.000}$	$0.414_{\pm 0.002}$	$0.308_{\pm 0.001}$	$0.899_{\pm 0.001}$	$0.181_{\pm 0.001}$	$0.106_{\pm 0.000}$	$0.981_{\pm 0.000}$
Wavelet-CNN	$0.101_{\pm 0.007}$	$0.035_{\pm 0.001}$	$0.996_{\pm 0.001}$	$0.462_{\pm 0.001}$	$0.342_{\pm 0.000}$	$0.884_{\pm 0.000}$	$0.201_{\pm 0.004}$	$0.110_{\pm 0.001}$	$0.977_{\pm 0.001}$
Novel SSM-Based Networks									
Mamba (Original)	$0.042_{\pm 0.004}$	$0.023_{\pm 0.002}$	$1.000^{**}_{\pm 0.000}$	$0.452_{\pm 0.018}$	$0.735_{\pm 0.066}$	$0.925^{**}_{\pm 0.004}$	$0.178_{\pm 0.003}$	$0.132^{**}_{\pm 0.008}$	$0.983^{**}_{\pm 0.002}$
Multi-Modular Macro Mamba (M4)	$0.120_{\pm 0.002}$	$0.067_{\pm 0.001}$	$0.992_{\pm 0.001}$	$\mathbf{0.340}_{\pm 0.000}$	$\mathbf{0.289}_{\pm 0.001}$	$\mathbf{0.927}_{\pm 0.001}$	$\mathbf{0.163}_{\pm 0.001}$	$\mathbf{0.127}_{\pm 0.000}$	$\mathbf{0.989}_{\pm 0.000}$

Future Work: Despite research in ML-based EEG denoising getting a great deal of attention lately, several avenues remain unexplored. We discuss below some promising challenges in the field which could yield interesting improvements and open new research paths in psychology and neuroimaging under brain stimulation.

In the first place, a notable limitation in the existing body of work is inherited from the original extensively benchmarked `EEGdenoiseNet` dataset: EEG recordings are single-channel and often restricted to segments no longer than 2 s. Although most of the methods explored in this study are already compatible with *multi-channel* signals and longer sequence lengths, the lack of a *foundational* reference dataset for evaluating artifact removal in more realistic settings severely limits the *practicality* of these methods in real-world studies on neural dynamics. In this regard, while our study establishes strong single-channel benchmarks, future work must focus on adapting and validating these architectures on datasets that are not only multi-channel and of longer duration, but also comprised of a diverse participant population. The ability to process multiple channels simultaneously is critical for capturing the spatial dynamics of neural activity, which is of paramount importance for robust clinical applications and advanced Brain–Computer Interfaces.

Secondly, the demonstrated success of our proposed models on complex, non-stationary stimulation artifacts suggests their potential for broader applications. Future research could explore the generalizability of these advanced SSM-based architectures to other challenging types of EEG noise, such as muscular (EMG), ocular (EOG), or other kinds of artifacts. Investigating the transferability of these models might significantly advance the development of an integrated cross-perturbatory EEG denoising tool.

Likewise, another critical aspect of how EEG denoising is usually approached is the robustness of the current evaluation metrics used to assess artifact removal (i.e., CC and RRMSEs). These metrics assess separately the performance in the temporal and spectral domains, and the correlation between the two signals. For that reason, future studies should investigate and develop more comprehensive evaluation frameworks the hopefully assess and combine the consistency and fidelity of the denoised signals across the different domains. Moreover, studies such as Liu et al. (2023) propose to leverage both temporal and spectral information by extracting network features from both domains.

Additionally, while these metrics are essential for benchmarking against existing literature, they may not fully capture the preservation of subtle but physiologically crucial features of the EEG signal. Future studies should therefore incorporate more functionally relevant metrics to ensure that denoising algorithms do not inadvertently distort underlying neural dynamics. These could include the preservation of power within specific spectral bands (e.g., alpha, beta), the conservation of phase synchrony between channels (in multi-channel extensions of this work), or the integrity of event-related potential (ERP) waveforms. Assessing performance on these dimensions is paramount for validating the use of these algorithms in clinical neuroscience research where the fidelity of the neural signal is the primary outcome of interest.

Furthermore, the application of *state-of-the-art* foundational time-series methods based on Transformers presents another promising direction for future research. Models such as Google's decoder-only `TimesFM` (Das et al., 2023), Amazon's encoder–decoder `CHRONOS` (Ansari et al., 2024), or Salesforce's masked encoder-only `MOIRAI` (Woo et al., 2024) among many others have demonstrated impressive results in forecasting tasks by handling complex temporal dependencies and could potentially bring significant improvements in EEG denoising.

Finally, the exploration of diffusion models for EEG denoising represents an exciting frontier. Diffusion models, which have shown success in various signal processing tasks, could offer novel ways to approach the denoising problem (Yang et al., 2024; Meijer and Chen, 2024). These models work by iteratively refining noisy signals through a learned diffusion process, progressively enhancing the signal quality. Applying diffusion models to EEG data could provide a robust framework for mitigating noise while preserving critical neurodynamic information.

CRediT authorship contribution statement

Miguel Fernandez-de-Retana: Writing – review & editing, Writing – original draft, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Pablo Matanzas-de-Luis:** Investigation, Data curation. **Javier Peña:** Supervision. **Aitor Almeida:** Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been supported by DEUSTEK5 – Human-Centric Computing for Smart Sustainable Communities and Environments, Basque Universities' System's research group, with code IT1582-22.

References

Ansari, A.F., Stella, L., Turkmen, C., Zhang, X., Mercado, P., Shen, H., Shchur, O., Rangapuram, S.S., Arango, S.P., Kapoor, S., et al., 2024. Chronos: Learning the language of time series. arXiv preprint arXiv:2403.07815.

Antal, A., Boros, K., Poreisz, C., Chaieb, L., Terney, D., Paulus, W., 2008. Comparatively weak after-effects of transcranial alternating current stimulation (tACS) on cortical excitability in humans. Brain Stimul. 1 (2), 97–105.

Barban, F., Chiappalone, M., Bonassi, G., Mantini, D., Semprini, M., 2021. Yet another artefact rejection study: an exploration of cleaning methods for biological and neuromodulatory noise. J. Neural Eng. 18 (4), 0460c2.
Baxter, B.S., Edelman, B., Zhang, X., Roy, A., He, B., 2014. Simultaneous high-definition transcranial direct current stimulation of the motor cortex and motor imagery. In: 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, pp. 454–456.
Borse, S., 2015. EEG de-noising using wavelet transform and fast ICA. IJISET-International J. Innov. Sci. Eng. & Technol. 2 (7), 200–205.
Camacho-Conde, J.A., Gonzalez-Bermudez, M.d.R., Carretero-Rey, M., Khan, Z.U., 2022. Brain stimulation: a therapeutic approach for the treatment of neurological disorders. CNS Neurosci. Ther. 28 (1), 5–18.
Cohen, M.X., 2017. Where does EEG come from and what does it mean? Trends Neurosci. 40 (4), 208–218.
Dao, T., Gu, A., 2024. Transformers are SSMs: Generalized models and efficient algorithms through structured state space duality. arXiv preprint arXiv:2405.21060.
Das, A., Kong, W., Sen, R., Zhou, Y., 2023. A decoder-only foundation model for time-series forecasting. arXiv preprint arXiv:2310.10688.
Gu, A., 2023. Modeling Sequences with Structured State Spaces. Stanford University.
Gu, A., Dao, T., 2023. Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752.
Gu, A., Goel, K., Ré, C., 2021. Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396.
Hahn, C., Rice, J., Macuff, S., Minhas, P., Rahman, A., Bikson, M., 2013. Methods for extra-low voltage transcranial direct current stimulation: current and time dependent impedance decreases. Clin. Neurophysiol. 124 (3), 551–556.
Haslacher, D., Nasr, K., Robinson, S.E., Braun, C., Soekadar, S.R., 2021. Stimulation artifact source separation (SASS) for assessing electric brain oscillations during transcranial alternating current stimulation (tACS). Neuroimage 228, 117571.
Huang, N.E., Shen, Z., Long, S.R., Wu, M.C., Shih, H.H., Zheng, Q., Yen, N.-C., Tung, C.C., Liu, H.H., 1998. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 454 (1971), 903–995.
Kasten, F.H., Maess, B., Herrmann, C.S., 2018. Facilitated event-related power modulations during transcranial alternating current stimulation (tACS) revealed by concurrent tACS-MEG. Eneuro 5 (3).
Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
Kohli, S., Casson, A.J., 2019. Removal of gross artifacts of transcranial alternating current stimulation in simultaneous EEG monitoring. Sensors 19 (1), 190.
Kohli, S., Casson, A.J., 2020. Machine learning validation of EEG+ tACS artefact removal. J. Neural Eng. 17 (1), 016034.
Liu, X.H., Jiang, W.B., Zheng, W.L., Lu, B.L., 2023. Two-stream spectral-temporal denoising network for end-to-end robust EEG-based emotion recognition. In: International Conference on Neural Information Processing. Springer, pp. 186–197.
Meijer, C., Chen, L.Y., 2024. The rise of diffusion models in time-series forecasting. arXiv preprint arXiv:2401.03006.
Merrill, W., Petty, J., Sabharwal, A., 2024. The illusion of state in state-space models. arXiv preprint arXiv:2404.08819.
Mert, A., Akan, A., 2014. EEG denoising based on empirical mode decomposition and mutual information. In: XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013: MEDICON 2013, 25-28 September 2013, Seville, Spain. Springer, pp. 631–634.
Miniussi, C., Brignani, D., Pellicciari, M.C., 2012. Combining transcranial electrical stimulation with electroencephalography: a multimodal approach. Clin. EEG Neurosci. 43 (3), 184–191.
Nitsche, M.A., Paulus, W., 2000. Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. J. Physiol. 527 (Pt 3), 633.
Pu, X., Yi, P., Chen, K., Ma, Z., Zhao, D., Ren, Y., 2022. EEGDnet: Fusing non-local and local self-similarity for EEG signal denoising with transformer. Comput. Biol. Med. 151, 106248.
Sharma, R., et al., 2017. EEG signal denoising based on wavelet transform. In: 2017 International Conference of Electronics, Communication and Aerospace Technology, Vol. 1. ICECA, IEEE, pp. 758–761.
Simonsmeier, B.A., Grabner, R.H., Hein, J., Krenz, U., Schneider, M., 2018. Electrical brain stimulation (tES) improves learning more than performance: A meta-analysis. Neurosci. Biobehav. Rev. 84, 171–181.
Terney, D., Chaieb, L., Moliadze, V., Antal, A., Paulus, W., 2008. Increasing human brain excitability by transcranial high-frequency random noise stimulation. J. Neurosci. 28 (52), 14147–14155.
Thut, G., Bergmann, T.O., Fröhlich, F., Soekadar, S.R., Brittain, J.-S., Valero-Cabré, A., Sack, A.T., Miniussi, C., Antal, A., Siebner, H.R., et al., 2017. Guiding transcranial brain stimulation by EEG/MEG to interact with ongoing brain activity and associated functions: a position paper. Clin. Neurophysiol. 128 (5), 843–857.
Woo, G., Liu, C., Kumar, A., Xiong, C., Savarese, S., Sahoo, D., 2024. Unified training of universal time series forecasting transformers. arXiv preprint arXiv:2402.02592.
Yang, Y., Jin, M., Wen, H., Zhang, C., Liang, Y., Ma, L., Wang, Y., Liu, C., Yang, B., Xu, Z., et al., 2024. A survey on diffusion models for time series and spatio-temporal data. arXiv preprint arXiv:2404.18886.
Yavari, F., Jamil, A., Samani, M.M., Vidor, L.P., Nitsche, M.A., 2018. Basic and functional effects of transcranial electrical stimulation (tES)—An introduction. Neurosci. Biobehav. Rev. 85, 81–92.
Yu, L., 2009. EEG de-noising based on wavelet transformation. In: 2009 3rd International Conference on Bioinformatics and Biomedical Engineering. IEEE, pp. 1–4.
Żebrowska, M., Dzwiniel, P., Waleszczyk, W.J., 2020. Removal of the sinusoidal transorbital alternating current stimulation artifact from simultaneous EEG recordings: effects of simple moving average parameters. Front. Neurosci. 14, 735.
Zhang, H., Wei, C., Zhao, M., Liu, Q., Wu, H., 2021a. A novel convolutional neural network model to remove muscle artifacts from EEG. In: ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 1265–1269.
Zhang, Z., Yu, X., Rong, X., Iwata, M., 2022. A novel multimodule neural network for EEG denoising. IEEE Access 10, 49528–49541.
Zhang, H., Zhao, M., Wei, C., Mantini, D., Li, Z., Liu, Q., 2021b. EEGdenoiseNet: a benchmark dataset for deep learning solutions of EEG denoising. J. Neural Eng. 18 (5), 056057.

Miguel Fernandez-de-Retana is affiliated with the Basque Center for Applied Mathematics (BCAM) and the Faculty of Engineering at the University of Deusto. His main research interests lie in mathematical modeling with multidisciplinary applications, particularly in biologically-motivated and health-related problems. His current work focuses on the application of Bayesian methods and Hamiltonian Monte Carlo for identifying potential genetic biomarkers responsible for developing resistance to anti-cancer endocrine therapies. His research interests also include Causal ML, Geometric Deep Learning, as well as the fields of Natural Language Processing, Mathematical Biology, and Bioinformatics.

Pablo Matanzas-de-Luis is affiliated with the Faculty of Engineering at the University of Deusto. His main interests include computer programming and automation in engineer technology.

Javier Peña is an Associate Professor in the Department of Methods and Experimental Psychology at the University of Deusto, where he directs the Doctoral Program in Psychology. He has participated in twenty projects, some in institutions such as the Michael J. Fox Foundation, the National Institute of Mental Health, and the Johns Hopkins Medical Institution. He has published 41 scientific articles in international journals.

Aitor Almeida is a researcher and lecturer in the Faculty of Engineering at the University of Deusto. His research interests include Machine Learning, the use of Artificial Intelligence in healthcare applications, the analysis of user behavior in smart environments, and the study of user activity and discourse on social media. Aitor has participated in multiple European research projects related to these areas and has published more than 125 papers on these topics.