Imperial College London

DEPARTMENT OF BIOENGINEERING

Deep Learning classification of EEG responses to multi-dimensional transcranial electrical stimulation

Author:

Alexis Pomares Pastor

Supervisors:

Dr. Gregory Scott
Dr. Ines Ribeiro Violante

Submitted in partial fulfilment of the requirements for the Award of the MRes in Neurotechnology

CID: 01985285

Word Count: 6125

September 2021

Abstract

A major shortcoming of medical practice is the lack of an objective measure of conscious level. Impairment of consciousness is common, e.g. following brain injury and seizures, which can also interfere with sensory processing and volitional responses. This is also an important pitfall in neurophysiological methods that infer awareness via command following, e.g. using functional magnetic resonance imaging or electroencephalography (EEG).

Transcranial electrical stimulation (TES) can be employed to non-invasively stimulate the brain, bypassing sensory inputs, and has already showed promising results in providing reliable indicators of brain state. However, current non-invasive solutions have been limited to transcranial magnetic stimulation, which is not easily translatable to clinical settings. Our long-term vision is to develop an objective measure of brain state that can be used at the bedside, without requiring patients to understand commands or initiate motor responses.

In this study, we demonstrated the feasibility of a framework using Deep Learning (DL) algorithms to classify EEG brain responses evoked by a defined multi-dimensional pattern of TES. We collected EEG-TES data from 11 participants and found that delivering transcranial direct current stimulation (tDCS) to posterior cortical areas targeting the angular gyrus elicited an exceptionally reliable brain response. For this paradigm, our best Convolutional Neural Network models reached a 92% classification F1-score on Holdout data from participants never seen during training, significantly surpassing an estimated human-level performance at 60-70% accuracy.

We hope that our findings will pave the way for further experiments with healthy participants in asleep states as well as with patients in abnormal states of consciousness, and that our insights into TES-evoked brain function will contribute to developing a robust measure of conscious level that could transform clinical decision making in patients with disorders of consciousness.

In this spirit, we documented and open-sourced the project in full—including datasets and commented code—to be used freely by the neuroscience and artificial intelligence research communities, who may replicate our results with free tools like GitHub, Kaggle, and Colaboratory.

Acknowledgements

To Ines and Greg for your deep insights and inestimable help. Thank you for fostering a strong critical scientific spirit in me, and for letting me explore unconventional interesting opportunities while staying on the right track in our research.

To my colleagues at Imperial, and to all the people who got involved in the studies presented here, who have helped me in ways too numerous to list.

To Chris Mihm for being such a splendid mentor in a decisive career stage. Thank you for making our goals your own and connecting me to Tom, who awakened my interest in neuroengineering with Kernel and his contagious enthusiasm.

Finally I thank my dad, mum, and brother for their love and support through the years.

"It is important to realize that if certain areas of science appear to be quite mature,
others are in the process of development,
and yet others remain to be born."

– Ramón y Cajal, 1897

Contents

List of Figures

List of Tables

List of Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Network
DC	Direct Current
DFA	Detrended Fluctuation Analysis
DL	Deep Learning
DoC	Disorder of Consciousness
EEG	Electroencephalography
EGI	Electrical Geodesics, Inc.
fMRI	Functional Magnetic Resonance Imaging
Grad-CAM	Gradient-weighted Class Activation Mapping
GRU	Gated Recurrent Unit
GTEN	Geodesic Transcranial Electrical Neuromodulation *[hardware device]*
IRAS	Integrated Research Application System
LSTM	Long Short-Term Memory
MFF	MetaFile Format
ML	Machine Learning
MNE	Magnetoencephalography and Electroencephalography *[software package]*
Hz	Hertz
mA	Milliamperes
mV	Millivolts
μV	Microvolts
NCC	Neural Correlates of Consciousness
PCI	Perturbational Complexity Index
PMS	Probabilistic Metastable Substates
PSD	Power Spectral Density
RNN	Recurrent Neural Network
ROI	Region of Interest
tACS	Transcranial Alternating Current Stimulation
tDCS	Transcranial Direct Current Stimulation
tPCS	Transcranial Pulsed Current Stimulation
tRNS	Transcranial Random Noise Stimulation
TES	Transcranial Electrical Stimulation
TMS	Transcranial Magnetic Stimulation
Topomap	EEG Topographic Map

1. Introduction

1.1 Project Motivation and Objectives

Brain imaging technologies are a common tool in neuroscience research environments, but they present inherent limiting factors that have prevented wide adoption in clinical settings. Electroencephalography (EEG) provides a good balance between cost-effectiveness, reliability, convenience, and signal quality. Nevertheless, EEG yields a complex signal that can require years of training to master its subjective interpretation, and advanced signal processing and feature extraction techniques to implement into functional applications (1). Clinical practice today relies on one technical specialist to record EEG, and a separate neurophysiologist to analyse it and draw conclusions on patient's brain state, a process that can take as long as several weeks (2).

The overarching long-term goal of our research team is to develop a novel measure of brain state that can be easily used at the patient's bedside, derived from the set of brain responses evoked by a multi-dimensional (i.e. potentially varying in intensity, time and space) 'program' of electrical neurostimulation. We want to leverage Artificial Intelligence (AI) to reliably automate—i.e. expedite and facilitate—the assessment of consciousness level, removing exposure to human error and simplifying the workflow of clinicians.

For this initial pilot study, our aim was to answer the following research question:

Can a Deep Learning algorithm even distinguish the EEG brain responses evoked by different types of electrical stimulation in distant cortical regions?

1.2 How Do We Measure (Un)Consciousness Today?

The lack of a 'gold standard' to measure conscious level is a major shortcoming in medical practice (3,4). Bedside evaluation of consciousness relies on a patient's ability to hear or see, understand, and act on instructions. Unfortunately, impairment of consciousness is common, e.g. following brain injury or seizures (5), and can often be accompanied by impaired sensory processing and volitional responses, a major confound for the behavioural assessment of conscious level (6).

The current approach can therefore miss clinically significant abnormalities in patient's states of consciousness, an important concern that could be solved with an objective measure that can bypass processing of verbal commands and motoric responses.

1.3 Functional Brain Imaging

Researchers sought such a measure by turning to non-invasive functional brain imaging methods, such as functional magnetic resonance imaging (fMRI) and EEG (7). Activity in specific brain regions in response to external stimuli can serve as a proxy of conscious level, e.g. instructing subjects in a vegetative state to perform different mental imagery tasks and monitoring the activation in motor or visual cortices (8).

fMRI measures changes in blood flow concomitant to modulation of neural metabolism, achieving high signal-to-noise ratio and spatial resolution (typically 3–4 mm) (9). Yet, being an indirect measure of neural activity limited by the hemodynamic response time, temporal resolution is comparatively poor, around 1–2 seconds. More importantly, fMRI lacks clinical scalability due to its relatively expensive and complex equipment, the highly specialized personnel required to operate it, and its restrictive limitations in terms of patient compatibility (e.g. ferromagnetic prostheses) (7).

By contrast, EEG measures voltage fluctuations in the scalp as a difference in potential from a designated reference electrode. EEG has poor spatial resolution compared to other imaging methods (due to volume conduction), but an excellent temporal resolution in the order of milliseconds, given it is a direct electrical recording (10). Furthermore, as Amabilis and John (7) eloquently stated:

"EEG is widely available, inexpensive, easy to administer at the bedside, is fairly robust to many artifacts that can cause fMRI data to be unusable, and has virtually no restrictions with regard to patient compatibility and safety. EEG is more easily validated on large groups of subjects and data acquisition times are generally shorter, making it not only more suited for clinical applications but also for the basic research required prior to applications in patients."

1.4 Non-Invasive Brain Stimulation

Neurostimulation is a valuable tool to use in combination with brain imaging methods, not only as a therapeutic approach, but also because it provides a framework to consistently elicit responses in the brain (4). Transcranial Magnetic Stimulation (TMS) uses electromagnetic induction to non-invasively perturb cortical activity, and it has been successfully used in combination with EEG to produce reliable indicators of conscious level (11). Nonetheless, TMS equipment is costly, poorly portable, and complex to setup and use.

Transcranial electrical stimulation (TES) is an umbrella term that encompasses a number of techniques—e.g. direct current (tDCS), alternating current (tACS) or random noise (tRNS)—used to safely pass electrical current through the cortex to alter brain function (12). TES is a portable, cost-effective, non-invasive method of directly stimulating the brain, which makes it a promising alternative to more easily probe brain function (versus fMRI or TMS/EEG) in patients with Disorders of Consciousness (DoC) who are not responsive to external stimuli (6).

In the next Section 1.5, we will review the current literature regarding how these techniques have been previously applied to produce metrics that may be used as proxies for brain activity.

1.5 Previous Work

1.5.1 Consciousness Assessment and Neural Correlates

The lack of an established method to measure conscious level empirically is the foremost confounding factor for diagnostic assessment, which consequentially has hindered the identification of adequate and relevant neural correlates of consciousness (NCC) (3). For this reason bedside evaluation is still largely based on behavioural observations, which have proven an unreliable proxy for consciousness with as much as a 40% clinical misdiagnosis rate (13).

The NCC have been defined as the minimal set of neuronal mechanisms jointly sufficient for any given specific conscious experience (4). They typically are further classified as either full NCC (supporting awareness in general, irrespective of content), content-specific NCC (neural substrate necessary for a particular content of experience—e.g. faces, whether seen, dreamt or imagined) or background conditions (factors that enable consciousness but do not contribute directly to it, such as arousal systems that ensure adequate excitability in the brain).

It has been argued that, in comparison with the amount of research devoted to content-specific NCC (so-called 'local states of consciousness'), full NCC or 'global states of consciousness' have been traditionally neglected (3). Correspondingly, most literature showed a tendency to overly simplify the description of full NCC by using a taxonomy that is taken to be scalable along a single dimension (14). This decades-long trend has been changing in recent years to favour a more descriptive multi-dimensional characterization though, with numerous publications reporting superior performance in systems that combine several different markers of cognition (15,16).

1.5.2 Metrics for Brain Activity

Recently it has been shown that conscious level is intrinsically linked to the complexity of brain activity, and that it can be detected at rest (i.e. in the absence of explicit instructions) (17). States of reduced conscious level are associated with reduced brain complexity, e.g. decreased variability in the repertoire of cortical activity, and a reduction in information transfer between cortical regions. The perturbational complexity index (PCI) is a putative measure of conscious level that quantifies the diversity in EEG responses to TMS pulses, bypassing sensory inputs (18). However, the PCI requires bulky, expensive TMS equipment and expertise, making broad clinical translation difficult. It also remains to be seen whether a uni-dimensional measure is sufficient to characterise the range of abnormal states of consciousness encountered in clinical practice (14).

In addition to complexity, numerous neural connectivity metrics and source estimation algorithms have been proposed, aimed at characterizing the complex dynamics of the cortex to uncover hidden information embedded in EEG and other signals (19,20). Still, in order to achieve consistent and reliable results, thoughtful selection of the brain regions most relevant to the research question is a paramount consideration when employing connectivity metrics (21) or when combining neurostimulation and functional imaging simultaneously (22). A significant amount of research has converged to identify a "posterior cortical hot zone" as a reliable full NCC, right at the intersection between the parietal, occipital and temporal lobes (see Figure 1)

(4,23,24). This is a promising finding that currently stands out as the main candidate to enable objective, accurate measurements of the level of consciousness in humans.

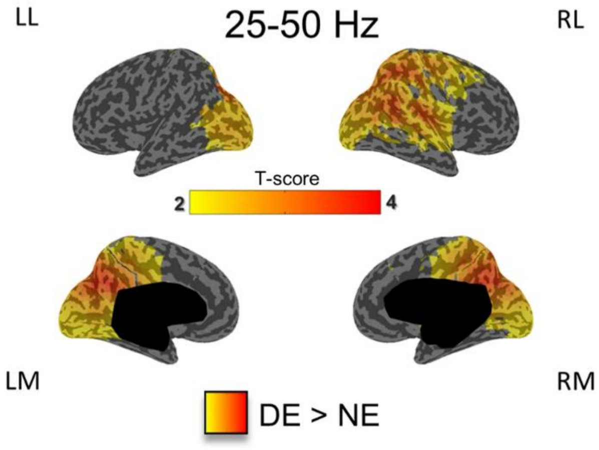


Figure 1. Cortical distribution of the contrast between dreaming experience (DE, i.e. conscious experience during sleep) and no experience (NE, i.e. dreamless sleep) during the non-rapid eye movement stage of sleep (25).

1.5.3 Deep Learning-based EEG Analysis

In the past decade there has been an increasing trend to explore Deep Learning (DL) techniques to analyse neurophysiological signals (Figure 2), most commonly EEG for its convenience and availability (1,26). DL is a subdiscipline within the field of Machine Learning (ML) that has enjoyed remarkable traction due to its excellent performance with relatively simple architectures, consisting of 'artificial neural networks'—loosely modelled on actual biological circuits—which are arranged as a series of nodes and connections (neurons and synapses) that work iteratively to optimally 'learn' the expected outcome for any series of given inputs.

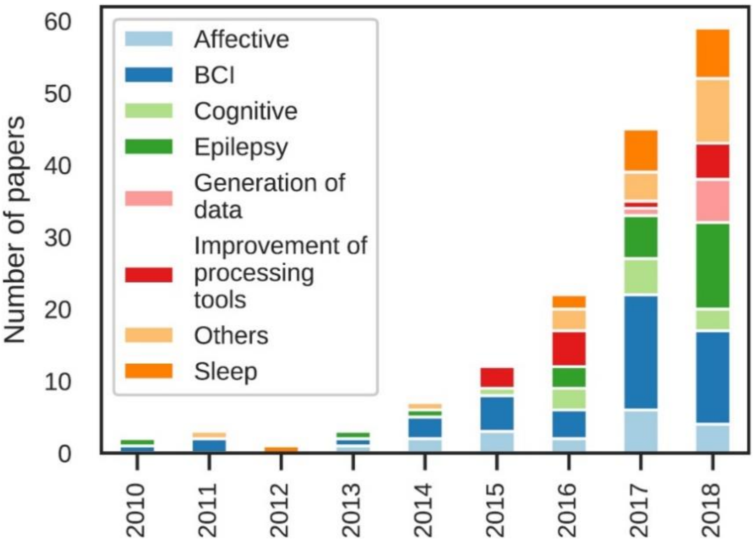


Figure 2. Number of research publications using DL-EEG approaches per domain per year, from 2010 - 2018 (1).

In a systematic review of more than 150 papers conducted by Roy et al. (1), it was found that different DL techniques applied directly to raw EEG time series accomplished a median gain in accuracy of 5.4% compared to more traditional ML methods such as signal processing combined with feature extraction. Lee and colleagues (27) used a Convolutional Neural Network (CNN) in combination with TMS to achieve a 90.9% accurate measure of sleep consciousness, while Ullah et al. (28) developed a DL-based system to consistently detect epileptic seizures with 99.1% minimum accuracy. These results add to a growing body of literature pointing to DL-EEG synergy as a promising tool to solve numerous pain points in clinical environments (26).

1.6 Rationale and Research Proposal

The goal of this study was to investigate whether a combination of EEG, TES, statistical feature extraction, and ML classification (whether DL or classical ML approaches) could render a framework to reliably classify evoked brain responses from non-invasive electrical stimulation.

The rationale to select our brain Regions of Interest (ROIs) for TES was based on the work from Deco et al. (29) where they studied controllability in different brain areas, i.e. the ability of a certain region to, when stimulated, induce transitions between cognitive states in humans. Dr. Deco kindly provided us with additional data containing the numerical values for controllability from the brain regions considered in their publication, as summarized in Annex I.

We used this data to parcellate a brain atlas by controllability values, from which we decided to target two bilateral neuroanatomic ROIs: the angular gyrus (posterior areas, marked with green stars in Figure 3), and the middle frontal gyrus (anterior regions, in blue). We selected these two ROIs for several reasons: they had diametrically opposite controllabilities—the *posterior* zone had the *highest* possible value, while the frontal area approached the lower end of controllability; they were distant from each other, thus simplifying the electrode arrangement for stimulation; the posterior ROI was overlapping with the 'hot zone' described in Section 1.5.2 (providing further evidence of this region as a NCC).

We decided to deliver TES in the two most conventional modalities—tDCS and tACS—across the two spatial conditions of *(i)* posterior parietal 'hot zone' bilaterally, and *(ii)* frontal areas bilaterally. Based on the context presented here in the *Introduction* section, it was our hypothesis that this montage would be best suited to answer the research question.

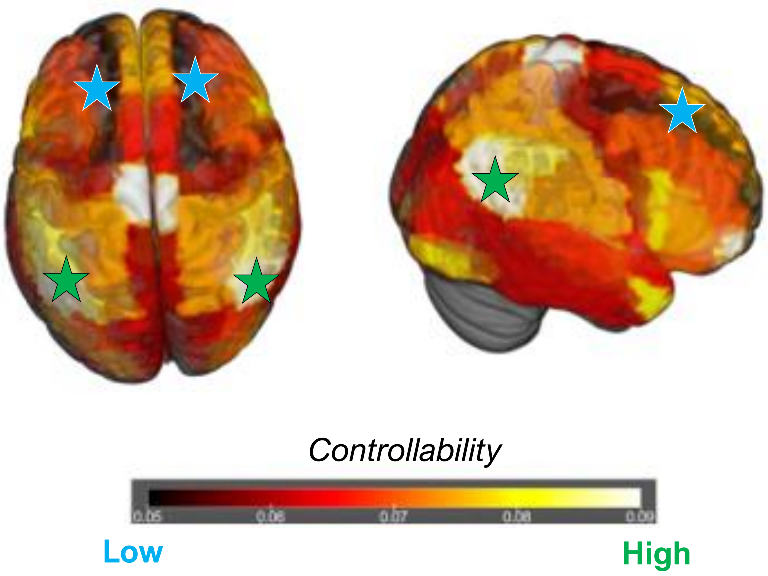


Figure 3. *Controllability values for different brain regions of the Automated Anatomical Labeling (AAL) atlas (30).*

2. Methodology

2.1 Equipment

We used a newly-acquired GTEN 200 neuromodulation system (Electrical Geodesics Inc, EGI) that allows TES—in any combination of tDCS/tACS/tRNS—and high-density EEG to be recorded through the same 256-electrodes cap (Figure 4) (31). The GTEN allows stimulation through any combination of electrodes whilst simultaneously recording EEG from the rest of electrodes.

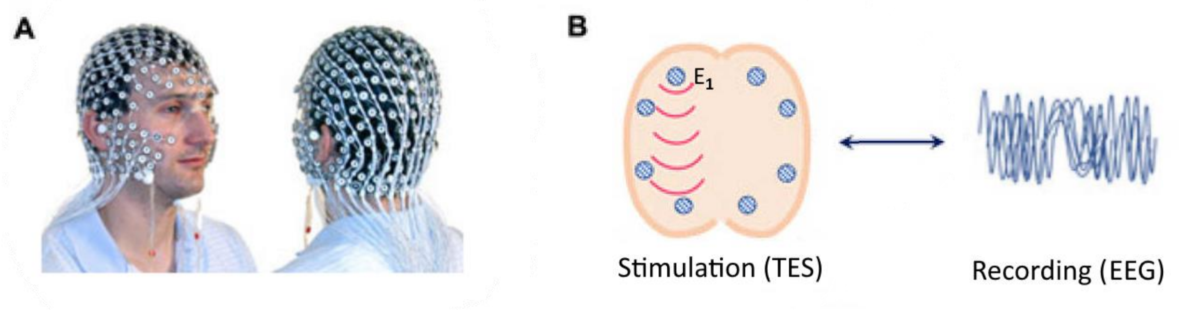


Figure 4. (A) Subject wearing a high-density cap for the GTEN 200 neuromodulation system (31). (B) Any combination of the 256 electrodes can be used for simultaneous—or alternate—brain imaging and stimulation.

2.2 Experimental Design

The experimental design is shown in Table 1 and Figure 5. Each session was split into 10 runs with 14 individual blocks, whereby we alternated stimulation periods of 2mA peak intensity ('Stim' blocks, labelled as 'Posterior/Frontal: tACS/tDCS') with non-stimulation ('Rest') periods. Rest blocks are intervals when EEG was recorded without any stimulation, and can be further classified as either 'M'—for Measure, i.e. blocks coming immediately after stimulation when we expected TES-related dynamics would be observed—or 'C', for Control rest periods following 'M' blocks when TES effects were expected to be minimal or no longer present.

For safety reasons, the GTEN planning software allows for a maximum of 0.2mA current intensity per electrode. We targeted the cortical regions explained in Section 1.6 by designating a narrow 5-channel circular anodal ring directly above each ROI, with a larger concentric cathodal ring surrounding the anodes with a separation ≥ 3 centimetres. This montage was designed to work with the GTEN restriction while ensuring maximal TES current would flow focally through our brain ROIs.

Table 1. Overall experimental design. Sessions were split into 10 runs of 14 alternate stimulation (tACS at 10Hz) and rest blocks, each ending in a final auditory oddball exercise designed to detect any drop in participant's attention.

Block / **Run**	1: 20s	2: 20s	3: 20s	4: 20s	5: 20s	6: 20s	7: 20s	8: 20s	9: 20s	10: 20s	11: 20s	12: 20s	13: 20s	14: 20s	15: 30s
Calibration	C	C	eyes closed			C	◢ 0.5mA	M	◢ 1mA	M	◢ 1.5mA	M	◢ 2mA	M	OB
Run 1	P: tACS	M	C	F: tDCS	M	F: tACS	M	C	P: tACS	M	P: tDCS	M	F: tACS	M	OB
Run 2	F: tDCS	M	P: tACS	M	C	P: tDCS	M	F: tDCS	M	C	F: tACS	M	P: tDCS	M	OB
Run 3	F: tACS	M	P: tDCS	M	F: tDCS	M	P: tACS	M	P: tACS	M	C	F: tDCS	M	C	OB
Run 4	P: tDCS	M	F: tACS	M	C	F: tDCS	M	C	P: tDCS	M	F: tACS	M	P: tACS	M	OB
Run 5	F: tACS	M	C	F: tDCS	M	C	F: tACS	M	P: tACS	M	P: tDCS	M	P: tACS	M	OB
Run 6	F: tDCS	M	P: tDCS	M	F: tDCS	M	P: tACS	M	P: tDCS	M	C	F: tACS	M	C	OB
Run 7	F: tACS	M	F: tACS	M	P: tDCS	M	C	P: tACS	M	F: tDCS	M	P: tACS	M	C	OB
Run 8	P: tDCS	M	C	P: tACS	M	P: tDCS	M	F: tDCS	M	F: tACS	M	C	F: tDCS	M	OB
Run 9	F: tACS	M	P: tDCS	M	C	P: tACS	M	C	F: tDCS	M	P: tACS	M	F: tACS	M	OB
Legend	Symbol	M		C		OB		P: tACS		F: tACS		P: tDCS		F: tDCS	
	Type	**Measure** Rest block		**Control** Rest block		**Oddball** (*no EEG*)		**Stim**: Posterior, AC		**Stim**: Frontal, AC		**Stim**: Posterior, DC		**Stim**: Frontal, DC	

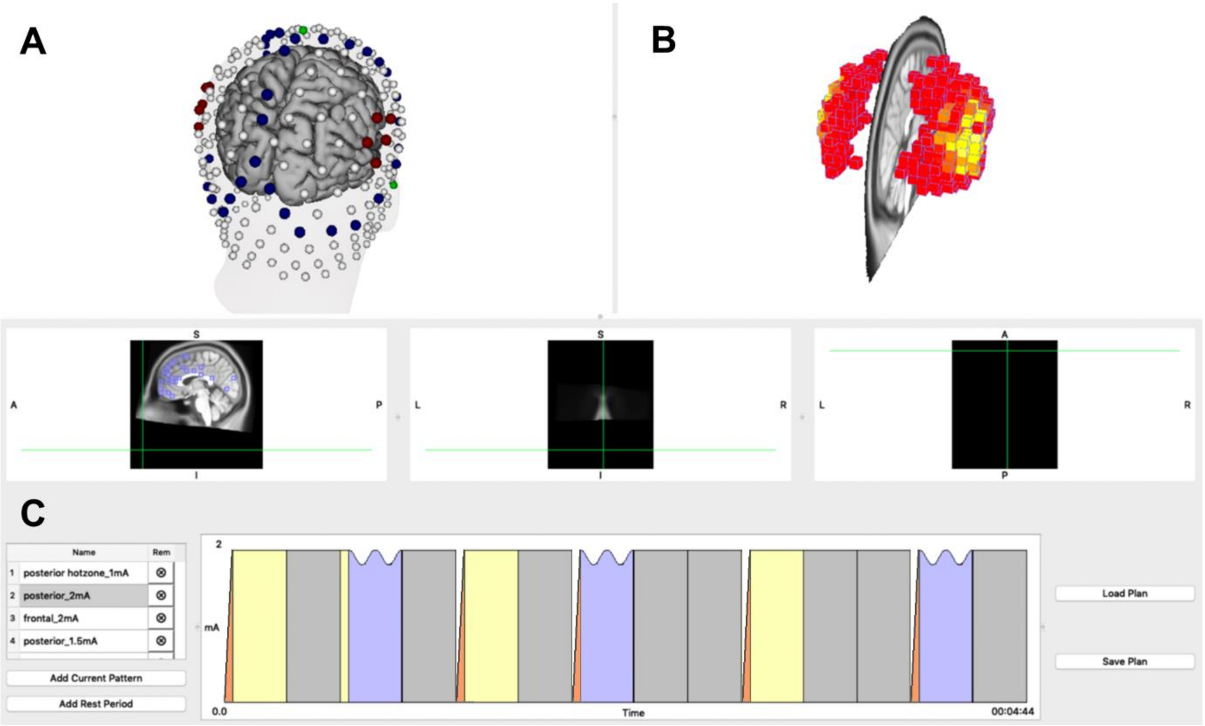


Figure 5. GTEN planning software interface. (A) Designation of electrodes to deliver anodal stimulation (i.e. electrical current flows inwards, in red) and their corresponding cathodes (outwards current flow, in blue). Note the bilateral concentric ring shape, chosen to maximize focal current intensity at our target brain ROIs. (B) Estimated current intensity for EGI's brain atlas, showing maximum intensity (yellow) in narrow cortical areas beneath anode electrodes. (C) Example of a non-calibration run for the experimental sessions. Wider blocks correspond to tDCS (yellow), Rest (grey) and tACS (purple); narrower orange blocks are positive-ramp stimulations designed to minimize discomfort.

The initial '*Calibration'* run was not intended to produce any meaningful data, but to check that the GTEN system was working correctly, as well as to deliver TES blocks of gradually increasing intensity to confirm that all participants were able to undergo TES without discomfort. An extra provision to minimize any potential TES-related sensation was to precede each stimulation block with a 3-seconds positive linear ramp (0 → 2mA) stimulation, as in Figure 5 above.

Furthermore, at the end of each run participants engaged in a 30-seconds auditory oddball exercise designed to measure and detect any drop in their level of attention, which could potentially interfere with the validity of results. The Python code for this oddball paradigm was open-sourced and made available as a GitHub repository (32).

2.3 Data Collection

We enrolled 11 healthy resting awake participants (4 female; ages 20-37, average 25.0±4.6 years old) to conduct 13 separate experimental sessions (subjects *P000* and *P001* participated twice after the results from their first sessions were found to be invalid). The anonymized version of our final raw EEG dataset is available to be used freely by the neuroscience research community as a Kaggle dataset (33).

Participants were instructed to sit awake with eyes open, and blinded to the conditions applied. Following an initial rest period of 120 seconds (including 60 seconds with eyes closed), up to 58 blocks of TES were performed, with a total time of up to ~60 minutes per session as tolerated per the participant. EEG was continuously recorded at a sampling rate of 1000Hz.

Prior to each session all subjects were asked to read and/or listen to a verbal brief of the study and the experiment, including what to expect and any potential adverse effects. After finishing the session, participants were also asked to fill a questionnaire where they reported if and when they had experienced any sensations or adverse effects from the neurostimulation.

All participants gave written informed consent. The study fully conforms to the Declaration of Helsinki, and ethical approval was granted through the local ethics board (IRAS Project ID: 154511).

2.4 EEG Preprocessing

We developed a holistic pipeline in Python to preprocess, visualize and analyse the raw EEG in EGI's Metafile Format (*.mff*), using the MNE library, an open-source package for analysing neurophysiological data (34,35). MNE provides comprehensive, powerful and well-documented functionality that is compatible with multiple languages (Python, Matlab, C, C++) and backends through a unified syntax and layer of abstraction. Our EEG pipeline is available in full as a GitHub repository, alongside our entire DL pipeline and support data required for different modules (36).

First, we removed the reference electrode as well as a subset of 68 'bad' channels (marked in red, Figure 6) that we empirically observed to consistently have bad contact with the scalp of participants. Most electrodes excluded from analysis were situated on non-scalp regions (e.g. neck or face). Although DL algorithms have been proven largely robust to bad data, adding noisy or redundant examples translates into longer and less effective training processes (37). Following this same rationale, we decided to resample our EEG signals from EGI's original 1000Hz to 250Hz, which significantly expedited the training speed of our DL models.

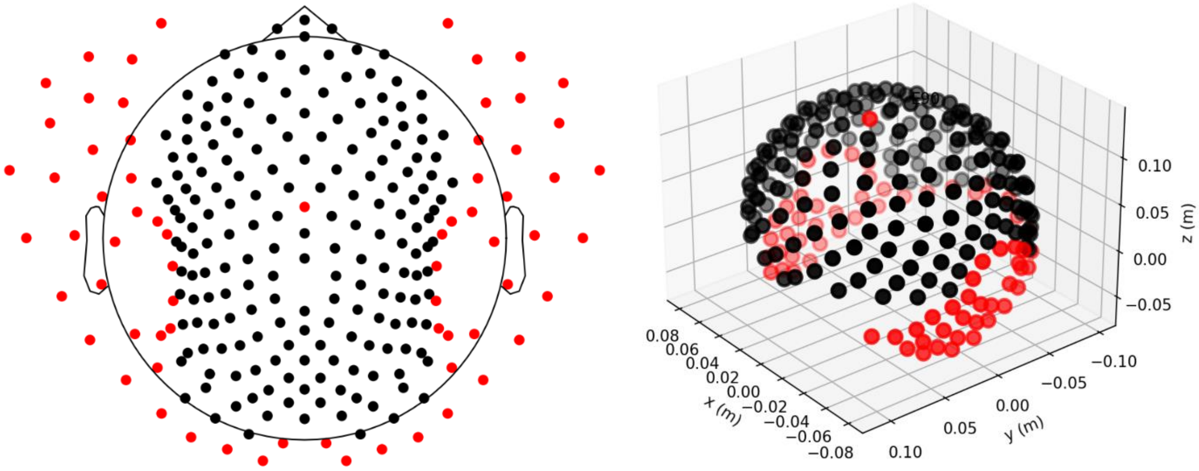


Figure 6. Electrodes around neck and face that showed bad contact with the scalp of participants, as well as the reference electrode, were marked (see red channels) and ignored by the analysis pipeline.

Next, we filtered the downsampled raw data (1Hz high-pass, 80Hz low-pass, 50Hz notch) to reduce or eliminate the impact of common EEG artifacts, such as DC baseline drift, high-frequency background noise, and electrical grid contamination. Then, we epoched the continuous EEG time series into non-overlapping windows of 1000 milliseconds.

2.5 EEG Datasets Generation

In addition to the typical EEG artifacts mentioned above, our study was also affected by electrical contamination from the interleaved TES applied to participants, which had an intensity in the order of mV and effectively corrupted the µV-scale EEG fluctuations in an irretrievable manner. Hence, the only possible solution was to discard all EEG epochs containing data from time intervals that had been impacted by neurostimulation artifacts. This is a well-known issue to the research community, with recent publications reporting successful approaches to mitigate the effects to a moderate extent (38–40).

Once the EEG signal had been cleaned and epoched, we computed a total of 37 electrode-wise statistical features (Table 2) along the time axis for each epoch. The features were selected as a combination of basic statistical measures (mean, median, std, max, min), standard EEG measures of linear and logarithmic power across 5 different EEG bands (delta, theta, alpha, beta, gamma) (41), and the collection of 22 features included in the *catch22* Python package (42,43). This Python library contains a set of high-performance statistical measures that have been selected according to their ability to capture structural information of time series data.

We exported both the epoched timeseries dataset and the computed features into separate folders. Each contained participant-wise directories with Comma Separated Value (*.csv*) files, a platform-neutral format that would later be easily consumed by our DL models. Epochs from the timeseries dataset consisted of rectangular slices of 250 time samples x 188 good electrodes, while for the features dataset each example contained 37 statistics x 188 good electrodes.

Table 2. Statistical features selected to describe the 1-second-long epochs from our EEG time series data.

Index	*Label*
0	Mean (μV)
1	Median (μV)
2	Standard Deviation (μV)
3	Maximum (μV)
4	Minimum (μV)
5	Power in Delta band, 1-3Hz (μV^2 / Hz)
6	Power in Theta band, 4-7Hz (μV^2 / Hz)
7	Power in Alpha band, 8-12Hz (μV^2 / Hz)
8	Power in Beta band, 13-29Hz (μV^2 / Hz)
9	Power in Gamma band, 30-45Hz (μV^2 / Hz)
10	Power in Delta band, 1-3Hz (dB)
11	Power in Theta band, 4-7Hz (dB)
12	Power in Alpha band, 8-12Hz (dB)
13	Power in Beta band, 13-29Hz (dB)
14	Power in Gamma band, 30-45Hz (dB)
15	Mode of z-scored distribution (5-bin histogram)
16	Mode of z-scored distribution (10-bin histogram)
17	Longest period of consecutive values above the mean
18	Time intervals between successive extreme events above the mean
19	Time intervals between successive extreme events below the mean
20	First 1 / e crossing of autocorrelation function
21	First minimum of autocorrelation function
22	Total power in lowest fifth of frequencies in the Fourier power spectrum
23	Centroid of the Fourier power spectrum
24	Mean error from a rolling 3-sample mean forecasting
25	Time-reversibility statistic, $\langle (x_{t+1}-x_t)^3 \rangle_t$
26	Automutual information, $m = 2$, $\tau = 5$
27	First minimum of the automutual information function
28	Proportion of successive differences exceeding 0.04σ
29	Longest period of successive incremental decreases
30	Shannon entropy of two successive letters in equiprobable 3-letter symbolization
31	Change in correlation length after iterative differencing
32	Exponential fit to successive distances in 2-d embedding space
33	Proportion of slower timescale fluctuations that scale with DFA (50% sampling)
34	Proportion of slower timescale fluctuations that scale with linearly rescaled range fits
35	Trace of covariance of transition matrix between symbols in 3-letter alphabet
36	Periodicity measure of (Wang et al. 2007)

Lastly, we produced a third dataset consisting of a simple reshape of the rectangular features to stack them horizontally into single rows (e.g. a shape of 37x10 became 1x370). However, in our study we focused exclusively on the rectangular features and timeseries datasets, and did not use these ‘concatenated’ features for our analysis pipeline. The motivation to produce such dataset was in anticipation of it being necessary for future research work, such as exploring other forms of Machine Learning (e.g. Support Vector Machines) that are typically designed to consume training examples in single-row format.

Anonymized full versions of all our 3 final preprocessed datasets (*timeseries*, *features* and *concatenated_features*) were also made publicly available in Kaggle (44).

2.6 Deep Learning Analysis

We partitioned the data from our 11 participants into *Training*, *Validation* and *Holdout Test* datasets. Training and Validation sets were generated from the same distribution of aggregated data from up to 10 participants, which was shuffled and randomly 95:5% split. Data from the one remaining participant was reserved to constitute our Holdout dataset for leave-one-out cross validation, where we tested whether the algorithm was able to generalize truly in an inter-subject manner and classify data from a person that was entirely never seen during training (that is, from a different data distribution).

Multiple DL models were developed in Python using the TensorFlow library, an end-to-end open source platform widely used by the artificial intelligence community (45). We trained and optimized these algorithms to classify EEG responses (*Rest* blocks) according to their evoking pattern of stimulation (i.e. their preceding *Stim* blocks), as illustrated in Figure 7.

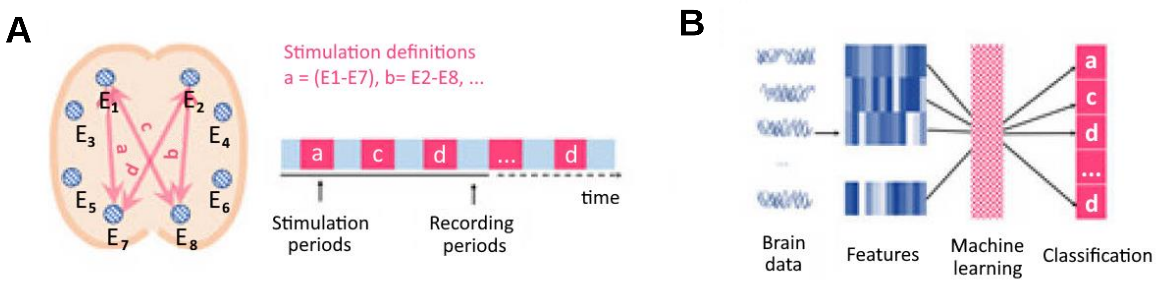


*Figure 7. **(A)** GTEN system executes a preset program of brain stimulations—covering a spatially-distributed set of electrode configurations—and records the EEG responses following each stimulation. **(B)** Collected brain data is analysed and fed to a DL algorithm that classifies each signal according to its specific evoking TES pattern.*

Several architectures were explored, starting with a Convolutional Neural Network (CNN) model inspired by the work of Jodie Ashford and colleagues (46). Other notable DL architectures explored were GRU (Gated Recurrent Unit) and LSTM (Long Short-Term Memory) types of Recurrent Neural Network (RNN) (47), as well as an adaptation of VGG-16, a well-known model widely used for visual recognition and image classification tasks (48).

We selected the best versions of our own CNN models after numerous iterations, which were developed by trial-and-error of core components as well as by finer, systematic hyperparameter tuning. Eventually we defined 2 separate architectures for our final CNN algorithms (36), each trained on either of our preprocessed datasets (timeseries or features).

3. Results

3.1 Experimental Data Collection and Subject Reports

Participants reported mild to moderate sensations of itching, skin redness, visuals, and fatigue/sleepiness. The visual experiences were always reported as phosphenes, i.e. participants perceived quickly flickering lights (at 10Hz in our case) that were caused by tACS current 'leaking' from the areas of stimulation to the retina (49). Phosphenes may be a contaminant because the measured brain response could correspond to the visual sensations perceived, rather than to the neural dynamics induced directly by the electrical impulse (50).

During calibration runs, participants received stimulation of increasing peak amplitudes: 0.5mA, 1mA, 1.5mA, and a maximum of 2mA intensity. All subjects reported no serious discomfort for the 2mA intensity with the exception of participant *P004*, who felt an intense itching sensation for the 2mA and 1.5mA intensities. Consequently, this subject received a peak amplitude of 1mA that would be easily tolerated for the entire session.

On the contrary, participant *P002* reported feeling no effects whatsoever during the entire session even with the maximum 2mA intensity, even after we verified TES current was being applied and flowing normally. An empirical observation was that while *P002* had an abundant amount of hair, *P004* had recently shaved himself, and hence his scalp was more easily exposed to the GTEN electrodes. In line with similar research on the issue, our finding suggests that hair acts as a mild insulator and participants with a certain voluminosity / density of hair may be preferred (51,52).

Lastly, an analysis of the results from the auditory oddball exercise at the end of each session run showed that the reaction times from participants remained stable within a range of 0.37 ± 0.12 seconds (see Annex II). This relatively narrow range indicates that the conditions of our experiment did not cause significant fatigue or loss of attention, and hence brain state can be considered relatively stable across the experiment.

3.2 EEG Analysis

We used the MNE Python package to curate, filter, trim, annotate and visualize all raw EEG signals. Some of these steps are depicted in Figure 8, where we observe the progression from a waveform heavily affected by electrical artifacts, to a clean and clearly segmented EEG signal.

Sections with red annotations (and 'bad/…' labels on top) mark the time-locked intervals affected by TES-related noise. Since these electrical artifacts had an intensity in the order of mV against the µV of the EEG, they inextricably corrupted the true signal and forced us to altogether discard those segments, resulting in an effective loss of approximately 52% of our collected data. This is a well-known issue to the research community, with some novel techniques being developed in recent years that allow circumventing this need to discard EEG segments, such as using adaptive filtering or sawtooth tACS (38–40).

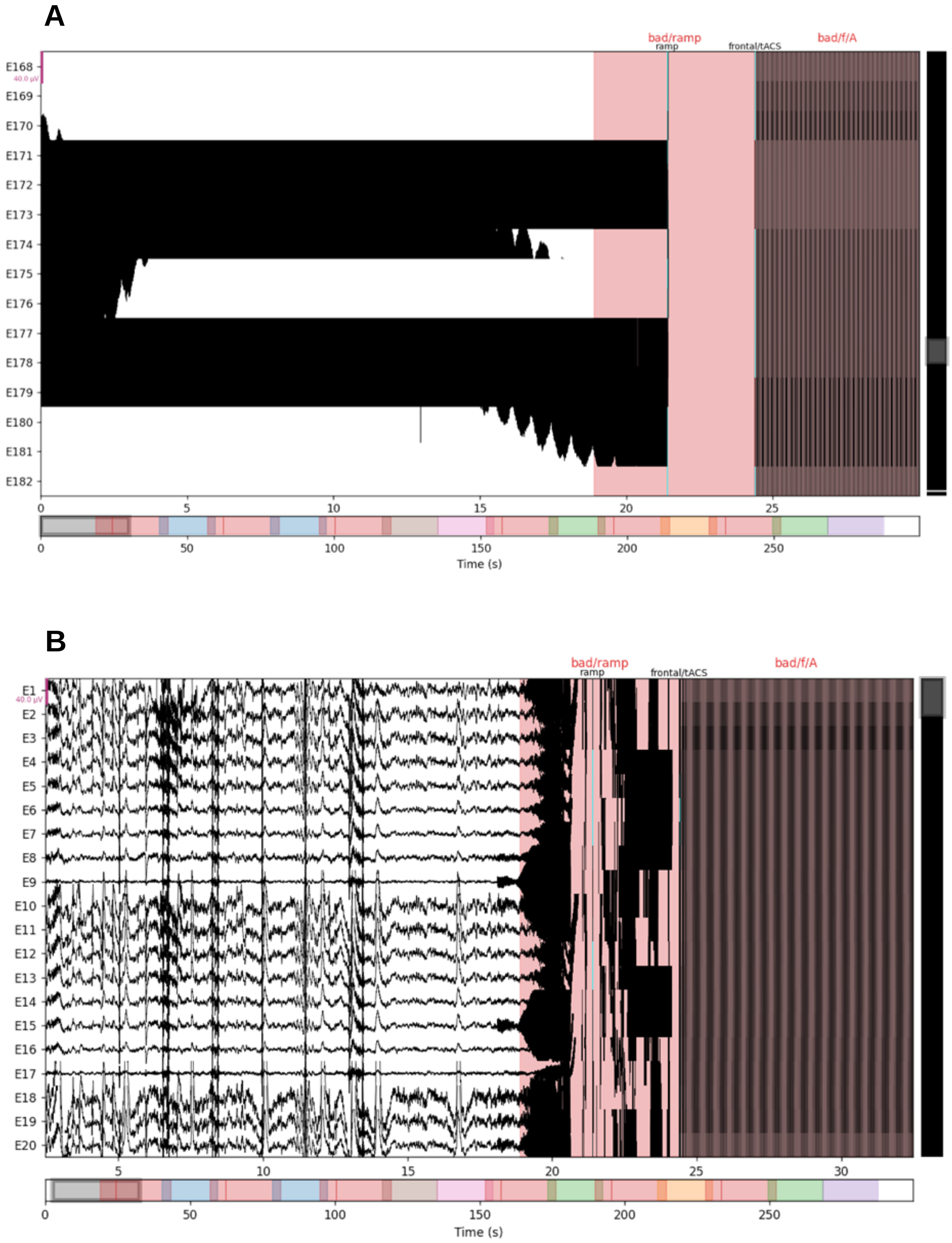


*Figure 8a. (A) Visualization of raw EEG files from EGI's custom MFF format. Note how the signal is heavily contaminated by various sources of electrical artifacts, mainly by channel-dependent DC baseline drift (low frequency << 1Hz) and by power grid noise (50Hz in the United Kingdom). **(B)** Applying the right filters (high-pass at 1Hz, notch at 50Hz) eliminates the distortions introduced by the artifacts mentioned before, resulting in a clean EEG signal at least for the 'rest' intervals (periods when TES was being applied remain completely noisy, as seen in the red segments).*

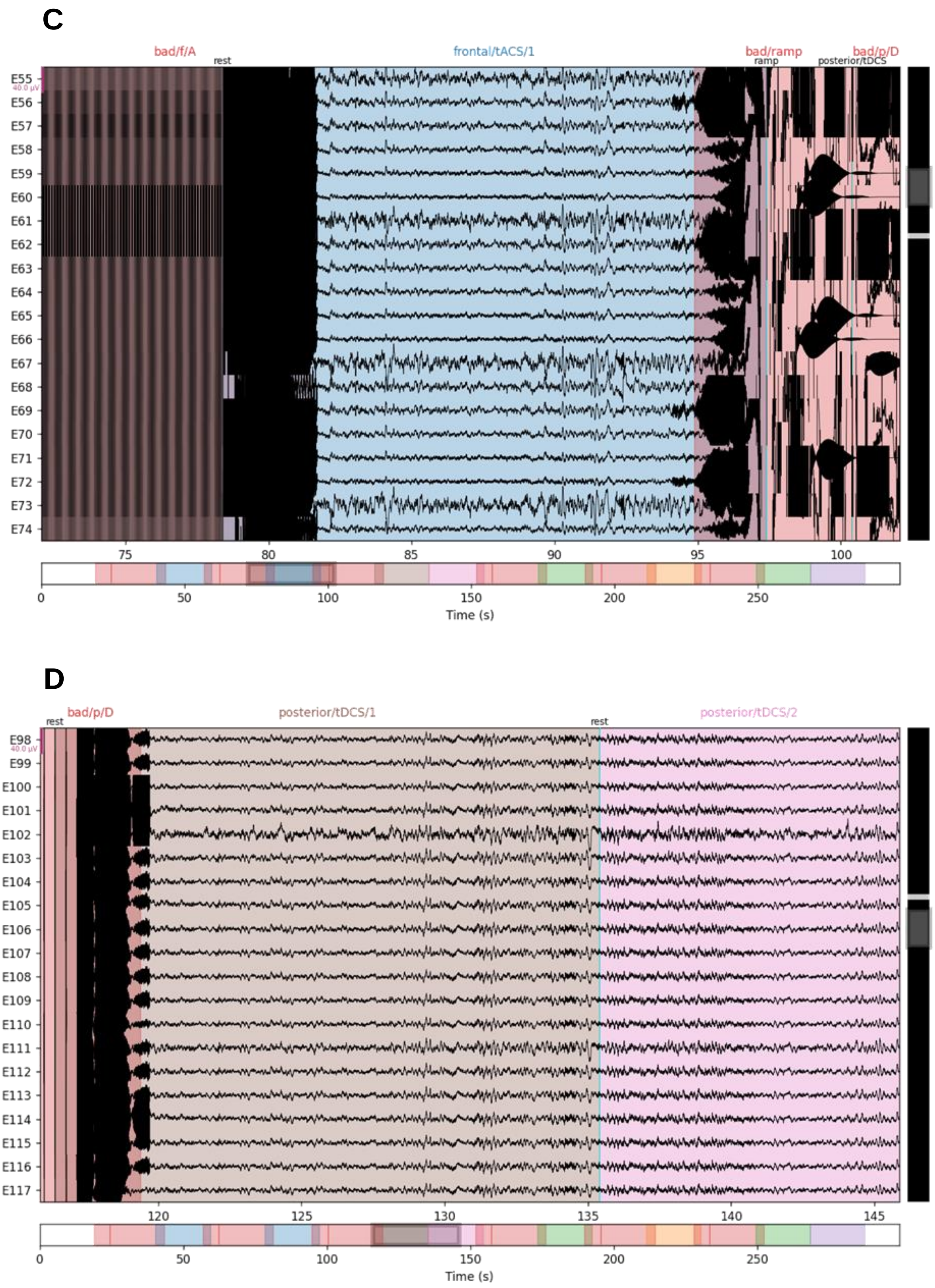


*Figure 8b [continued]. **(C)** Example of a single 'Measure' type of rest block (in blue), preceded by frontal/tACS and followed by posterior/tDCS stimulations (red). Note how the EEG signal is affected by the TES artifact even for a short span before and after the stimulation events (shown in red letters at the top). This effect is caused by the GTEN amplifier preparing to deliver (or recovering from) the electrical impulse, and was cancelled programmatically (see how the 'bad' red area extends 4 seconds further than the time-locked events marking the beginning and end of artifacts). **(D)** Example of a stimulation interval followed by 2 rest blocks: 'Measure' type (in brown) and a 'Control' rest block (pink).*

The impact of our preprocessing pipeline is also readily apparent when looking at the Power Spectral Density (PSD) before and after filtering the raw EEG (see Figure 9). The PSD describes the energy distribution per time unit of the different frequency components that constitute a time series signal. PSD is a common method to characterize EEG signals (53,54), allowing researchers to divide the time series into its composing frequencies and analyse them separately.

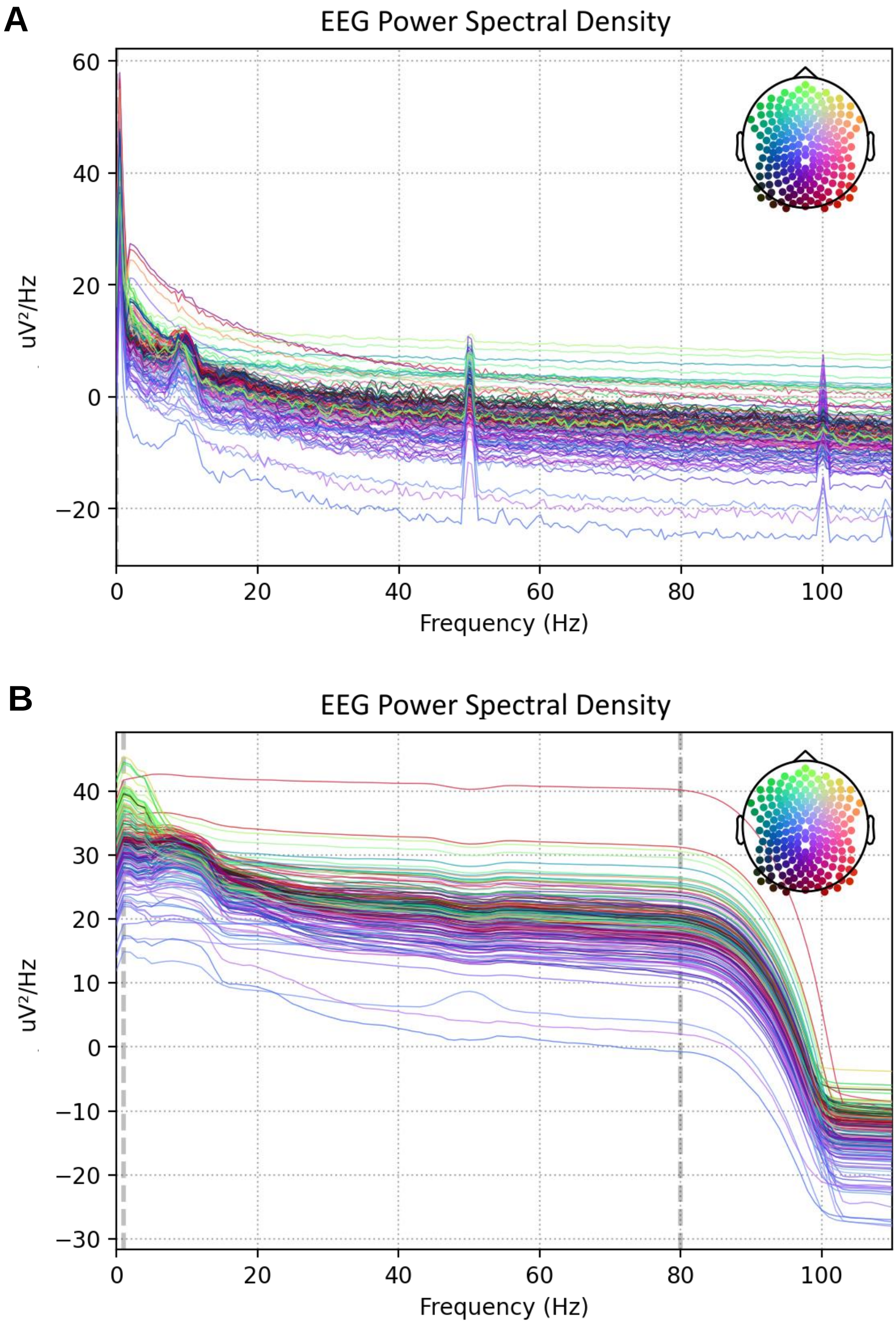


*Figure 9. (A) PSD of the original EEG signal, before any filtering or preprocessing was applied. Note the sharp increase at the electrical grid frequency (50Hz in the United Kingdom) and its harmonics (100Hz, 150Hz…), as well as a pronounced peak at very low frequencies (0-1Hz). **(B)** PSD of the filtered EEG signal. We can observe how the power density is now more evenly distributed across all frequencies, with the vertical dotted grey lines marking the cut-off frequencies for our 1Hz high-pass filter (left) and the 80Hz low-pass (right). The effect of the 50Hz notch filter is also evident, with the power density on that frequency having almost vanished in comparison to panel (A).*

Once the continuous EEG signal had been cleaned-up, we segmented it into non-overlapping windows (epochs), using the EEG annotations (shown as different colours in Figure 8) to ensure that we discarded all segments affected by any electrical contamination. The final result is presented in Figure 10, where we observe that only artifact-free 1-second-long epochs remain in our discontinuous time series dataset.

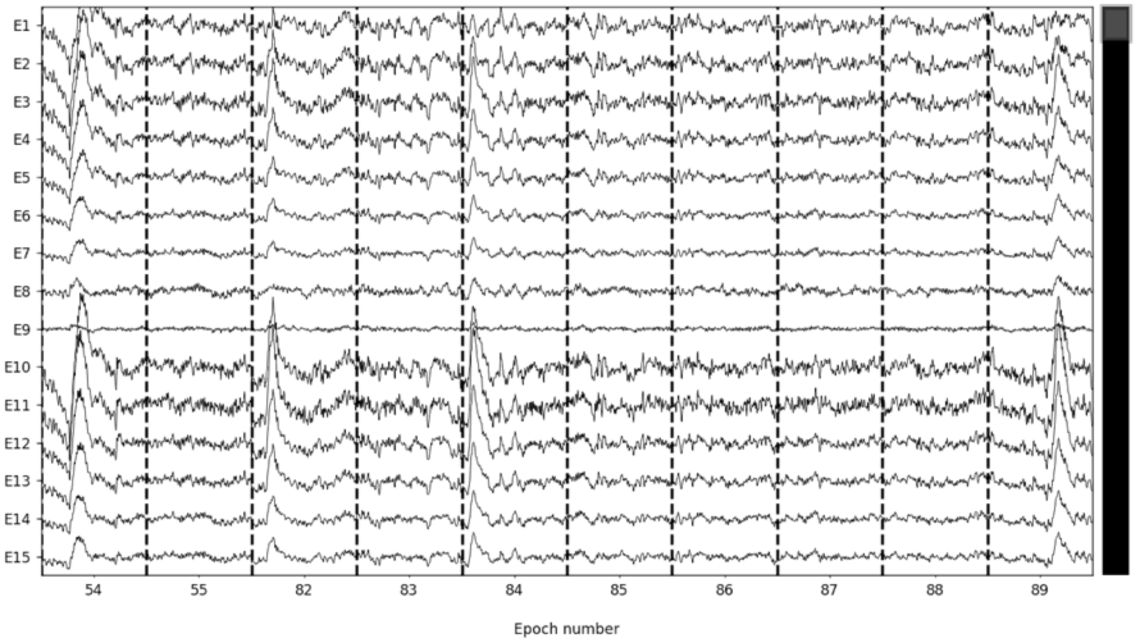
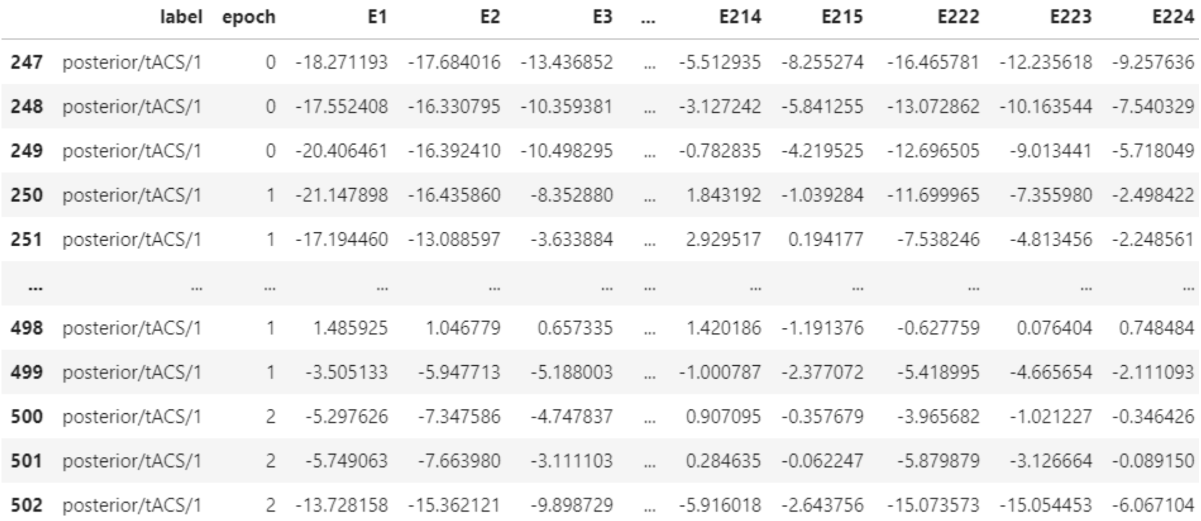


Figure 10. Example of 10 EEG epochs, to be used for all downstream EEG and DL analysis. Note the absence of electrical artifacts and the discontinuity between epochs number 55 and 82, as well as in the bottom bar, caused by the elimination of all epochs that corresponded to the interleaved TES segments.

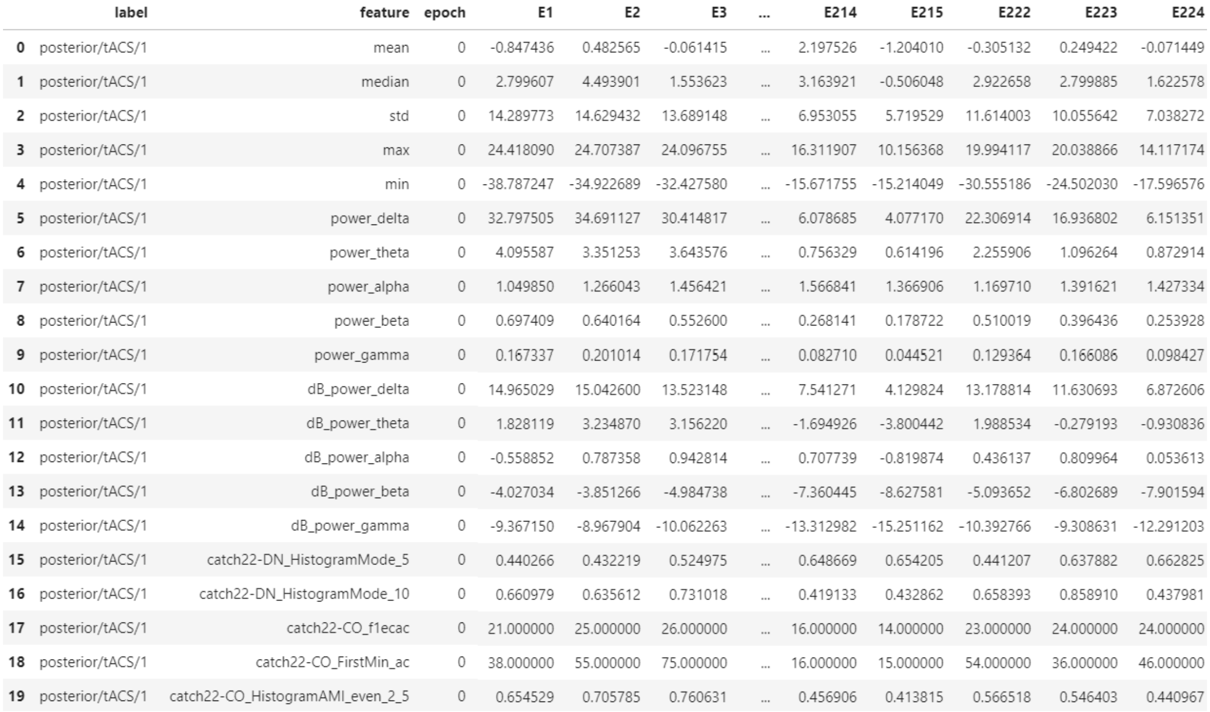
Next, we compute a set of 37 different features across the time axis of EEG epochs. At this point, we had produced 2 separate datasets that essentially contained different information about the same TES-elicited brain responses in participants: our timeseries dataset (Table 3) captured the EEG fluctuations directly as electrode-wise voltage differences (in μV) with respect to the reference electrode, while the features dataset (Table 4) condensed this information into a reduced set of descriptive statistical measures.

Table 3. Example of 10 rows from our timeseries dataset, including a transition between different 1-second epochs of the same stimulation interval. Training examples that would later be fed to our DL algorithms were obtained by slicing this dataset into subsets of shape 250 rows (i.e. EEG resampling rate) x 188 columns (number of electrodes).



	label	epoch	E1	E2	E3	...	E214	E215	E222	E223	E224
247	posterior/tACS/1	0	-18.271193	-17.684016	-13.436852	...	-5.512935	-8.255274	-16.465781	-12.235618	-9.257636
248	posterior/tACS/1	0	-17.552408	-16.330795	-10.359381	...	-3.127242	-5.841255	-13.072862	-10.163544	-7.540329
249	posterior/tACS/1	0	-20.406461	-16.392410	-10.498295	...	-0.782835	-4.219525	-12.696505	-9.013441	-5.718049
250	posterior/tACS/1	1	-21.147898	-16.435860	-8.352880	...	1.843192	-1.039284	-11.699965	-7.355980	-2.498422
251	posterior/tACS/1	1	-17.194460	-13.088597	-3.633884	...	2.929517	0.194177	-7.538246	-4.813456	-2.248561
...	...	...	...	...	...	...	...	...	...	...	...
498	posterior/tACS/1	1	1.485925	1.046779	0.657335	...	1.420186	-1.191376	-0.627759	0.076404	0.748484
499	posterior/tACS/1	1	-3.505133	-5.947713	-5.188003	...	-1.000787	-2.377072	-5.418995	-4.665654	-2.111093
500	posterior/tACS/1	2	-5.297626	-7.347586	-4.747837	...	0.907095	-0.357679	-3.965682	-1.021227	-0.346426
501	posterior/tACS/1	2	-5.749063	-7.663980	-3.111103	...	0.284635	-0.062247	-5.879879	-3.126664	-0.089150
502	posterior/tACS/1	2	-13.728158	-15.362121	-9.898729	...	-5.916018	-2.643756	-15.073573	-15.054453	-6.067104

Table 4. Example of the first 20 rows from our features dataset. Training examples in this case were slices of shape 37 rows (i.e. number of distinct statistical features) x 188 columns (number of channels). The smaller sizes for training examples translated into computations almost 7 times faster when compared to the timeseries dataset.



	label	feature	epoch	E1	E2	E3	...	E214	E215	E222	E223	E224
0	posterior/tACS/1	mean	0	-0.847436	0.482565	-0.061415	...	2.197526	-1.204010	-0.305132	0.249422	-0.071449
1	posterior/tACS/1	median	0	2.799607	4.493901	1.553623	...	3.163921	-0.506048	2.922658	2.799885	1.622578
2	posterior/tACS/1	std	0	14.289773	14.629432	13.689148	...	6.953055	5.719529	11.614003	10.055642	7.038272
3	posterior/tACS/1	max	0	24.418090	24.707387	24.096755	...	16.311907	10.156368	19.994117	20.038866	14.117174
4	posterior/tACS/1	min	0	-38.787247	-34.922689	-32.427580	...	-15.671755	-15.214049	-30.555186	-24.502030	-17.596576
5	posterior/tACS/1	power_delta	0	32.797505	34.691127	30.414817	...	6.078685	4.077170	22.306914	16.936802	6.151351
6	posterior/tACS/1	power_theta	0	4.095587	3.351253	3.643576	...	0.756329	0.614196	2.255906	1.096264	0.872914
7	posterior/tACS/1	power_alpha	0	1.049850	1.266043	1.456421	...	1.566841	1.366906	1.169710	1.391621	1.427334
8	posterior/tACS/1	power_beta	0	0.697409	0.640164	0.552600	...	0.268141	0.178722	0.510019	0.396436	0.253928
9	posterior/tACS/1	power_gamma	0	0.167337	0.201014	0.171754	...	0.082710	0.044521	0.129364	0.166086	0.098427
10	posterior/tACS/1	dB_power_delta	0	14.965029	15.042600	13.523148	...	7.541271	4.129824	13.178814	11.630693	6.872606
11	posterior/tACS/1	dB_power_theta	0	1.828119	3.234870	3.156220	...	-1.694926	-3.800442	1.988534	-0.279193	-0.930836
12	posterior/tACS/1	dB_power_alpha	0	-0.558852	0.787358	0.942814	...	0.707739	-0.819874	0.436137	0.809964	0.053613
13	posterior/tACS/1	dB_power_beta	0	-4.027034	-3.851266	-4.984738	...	-7.360445	-8.627581	-5.093652	-6.802689	-7.901594
14	posterior/tACS/1	dB_power_gamma	0	-9.367150	-8.967904	-10.062263	...	-13.312982	-15.251162	-10.392766	-9.308631	-12.291203
15	posterior/tACS/1	catch22-DN_HistogramMode_5	0	0.440266	0.432219	0.524975	...	0.648669	0.654205	0.441207	0.637882	0.662825
16	posterior/tACS/1	catch22-DN_HistogramMode_10	0	0.660979	0.635612	0.731018	...	0.419133	0.432862	0.658393	0.858910	0.437981
17	posterior/tACS/1	catch22-CO_f1ecac	0	21.000000	25.000000	26.000000	...	16.000000	14.000000	23.000000	24.000000	24.000000
18	posterior/tACS/1	catch22-CO_FirstMin_ac	0	38.000000	55.000000	75.000000	...	16.000000	15.000000	54.000000	36.000000	46.000000
19	posterior/tACS/1	catch22-CO_HistogramAMI_even_2_5	0	0.654529	0.705785	0.760631	...	0.456906	0.413815	0.566518	0.546403	0.440967

3.3 Deep Learning Metrics and Classification

After keeping only the EEG epochs that met the selection criteria outlined in Section 2.4—as well as in the previous section—we were left with a grand total of 166 minutes (2.8 hours) of valid EEG training data. This volume of data rendered almost 10,000 training examples for our DL pipeline, originated from resampling all our EEG recordings to 250Hz and dividing them into 0% overlap, 1-second windows. The full timeseries dataset contained 2.5 million rows at about 3.69GB total, while our features dataset consisted of 0.36 million rows and 0.56GB size in memory.

From the different DL architectures explored to reliably classify the EEG responses induced by TES, for both timeseries and features tasks we regularly observed poor performance with out of the box RNN topologies (47) as well as in famous models for image classification such as VGG-16 (48). In neither of these cases were we able to produce accuracies above chance level, which was approximately ~28% for our imbalanced 5-labels classification task (Table 5).

Table 5. Balance of classes for an arbitrary session run. The highest occurrence ratio of all labels (in this case, the 'rest' blocks) sets the chance level accuracy at approximately 28% for the classification tasks.

Class Label	Occurrences	Weight (ratio)
rest	90736	0.28
posterior/tACS	63031	0.19
frontal/tACS	62234	0.19
frontal/tDCS	56647	0.17
posterior/tDCS	56314	0.17

On the other hand, most CNN architectures that we tested based on the DL-EEG systematic review of Roy et al. (1) showed great promise out of the box. We quickly reached an accuracy of 50% with no modifications other than adapting the input layer to consume our data format. Subsequently, we decided to follow this line of work for our study, as presented in Figure 11.

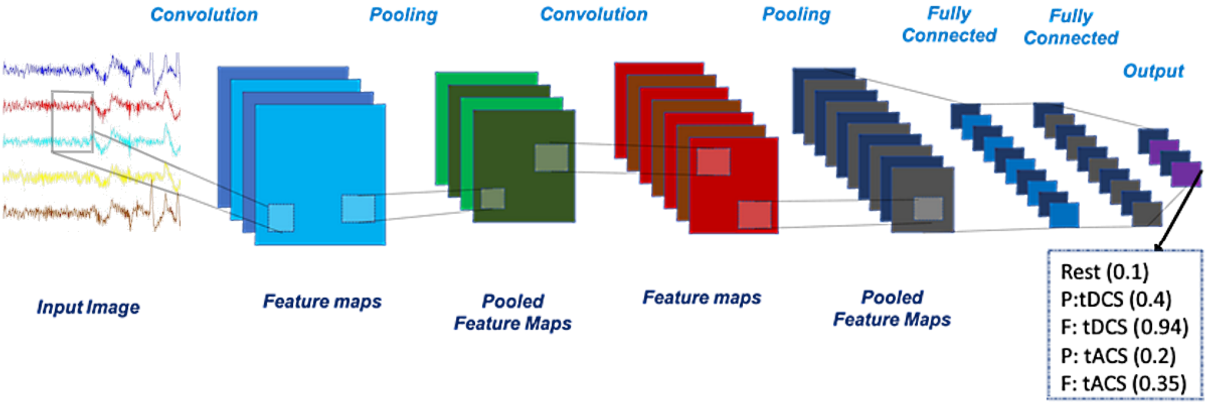
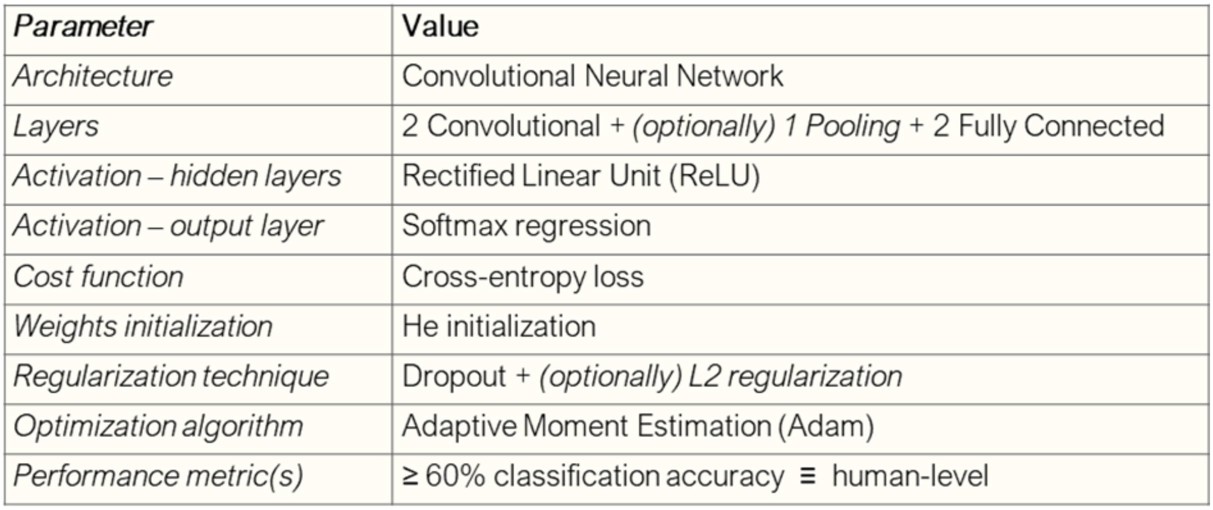


Figure 11. CNN base architecture with highest out-of-the-box observed accuracies (illustrated for timeseries classification task). Input EEG (whether timeseries or features) passes through 2 convolutional and (optionally) pooling layers, then through 2 fully connected layers that result in a 5-class softmax output, from where we pick the highest confidence prediction to classify our label.

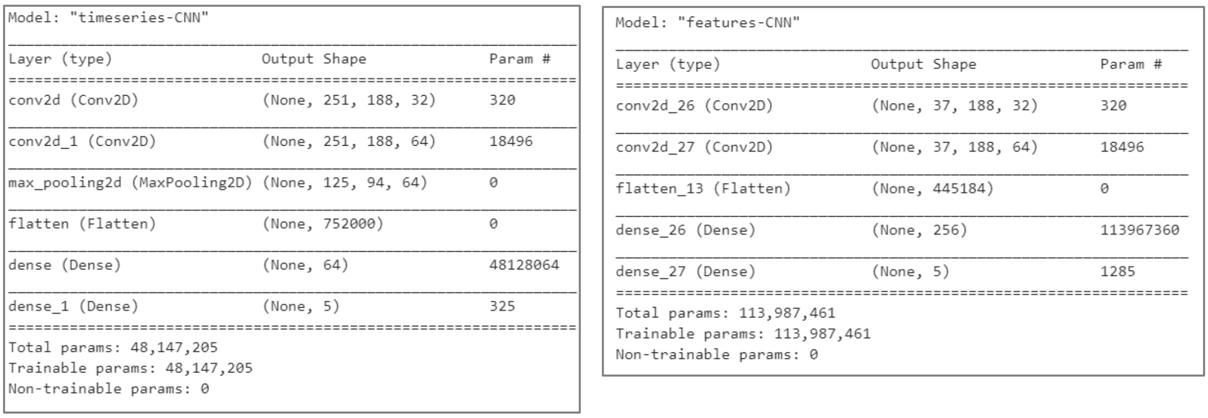
After an extensive process of hyperparameter tuning and architecture optimization, we concluded by defining our final neural networks as in Table 6 for both the timeseries and features tasks, with each of their corresponding parameters also summarized in Table 7.

Table 6. Summary of architecture, methods and functions chosen for our final CNN models. Note the choice of performance metric as equal or higher than human-level performance, as estimated by the rate of 30-40% clinical misdiagnosis reported by Giacino et al. (13).



Parameter	**Value**
Architecture	Convolutional Neural Network
Layers	2 Convolutional + *(optionally) 1 Pooling* + 2 Fully Connected
Activation – hidden layers	Rectified Linear Unit (ReLU)
Activation – output layer	Softmax regression
Cost function	Cross-entropy loss
Weights initialization	He initialization
Regularization technique	Dropout + *(optionally) L2 regularization*
Optimization algorithm	Adaptive Moment Estimation (Adam)
Performance metric(s)	≥ 60% classification accuracy ≡ human-level

Table 7. Network topologies and parameters used for the timeseries (left) and features (right) classification tasks. For specific definitions, please refer to the TensorFlow documentation (45).



```
Model: "timeseries-CNN"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 251, 188, 32)      320
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 251, 188, 64)      18496
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 125, 94, 64)       0
_________________________________________________________________
flatten (Flatten)            (None, 752000)            0
_________________________________________________________________
dense (Dense)                (None, 64)                48128064
_________________________________________________________________
dense_1 (Dense)              (None, 5)                 325
=================================================================
Total params: 48,147,205
Trainable params: 48,147,205
Non-trainable params: 0
```

```
Model: "features-CNN"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d_26 (Conv2D)           (None, 37, 188, 32)       320
_________________________________________________________________
conv2d_27 (Conv2D)           (None, 37, 188, 64)       18496
_________________________________________________________________
flatten_13 (Flatten)         (None, 445184)            0
_________________________________________________________________
dense_26 (Dense)             (None, 256)               113967360
_________________________________________________________________
dense_27 (Dense)             (None, 5)                 1285
=================================================================
Total params: 113,987,461
Trainable params: 113,987,461
Non-trainable params: 0
```

Using these models we were able to achieve >90% accuracy on both Training and Validation datasets, as plotted in Figure 12 for the features classification task. We also computed the confusion matrix to discover what were the label-wise misclassification rates. The results (Figure 13) clearly show that the DL algorithm classified tDCS conditions almost perfectly, while a small but significant percentage of tACS were mislabelled as 'rest' (i.e. resting state 'C' type blocks).

This fact suggests that at some level the EEG responses to tACS stimulation resemble resting state neural dynamics. We further explore how these two conditions compare in Section 3.4.

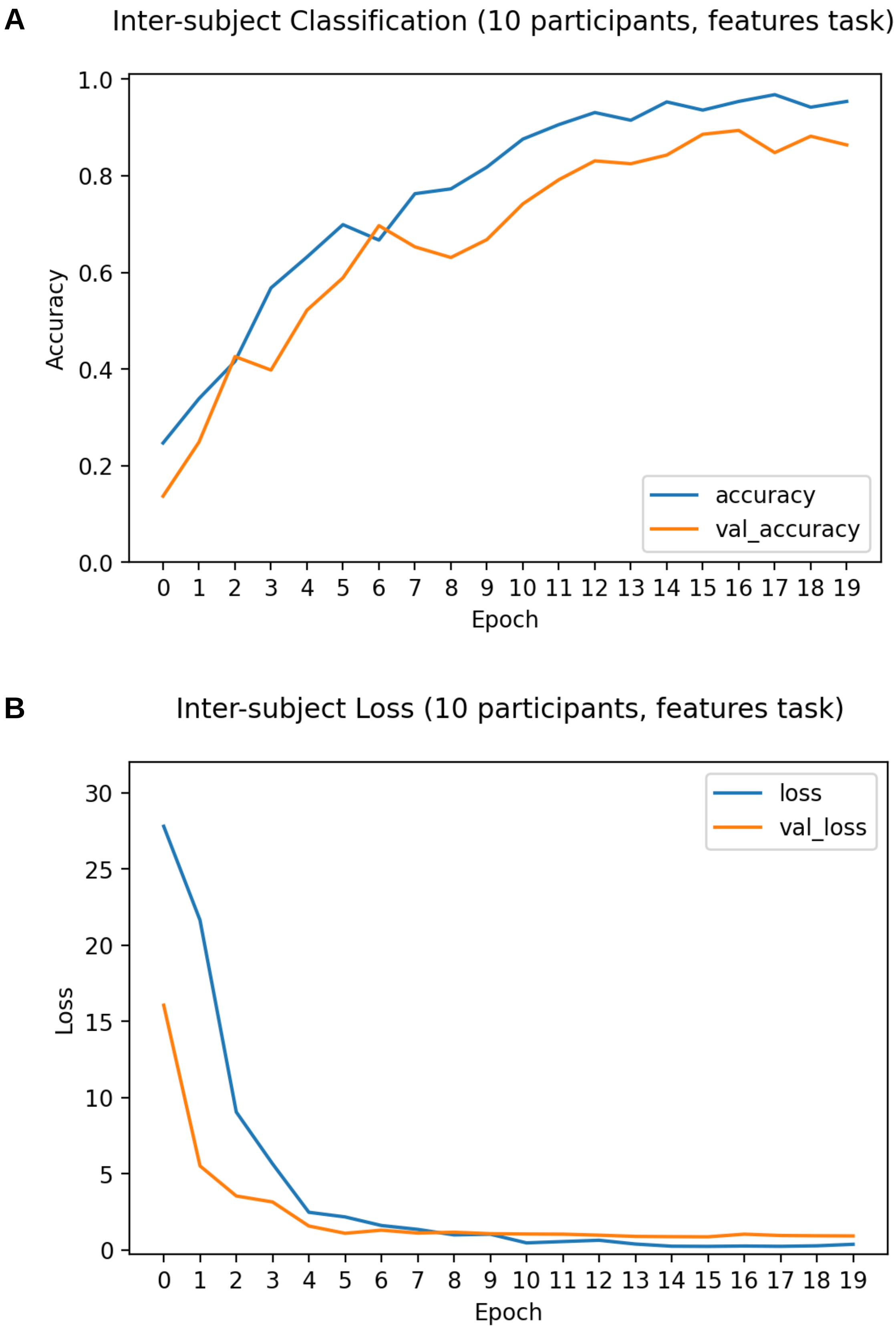
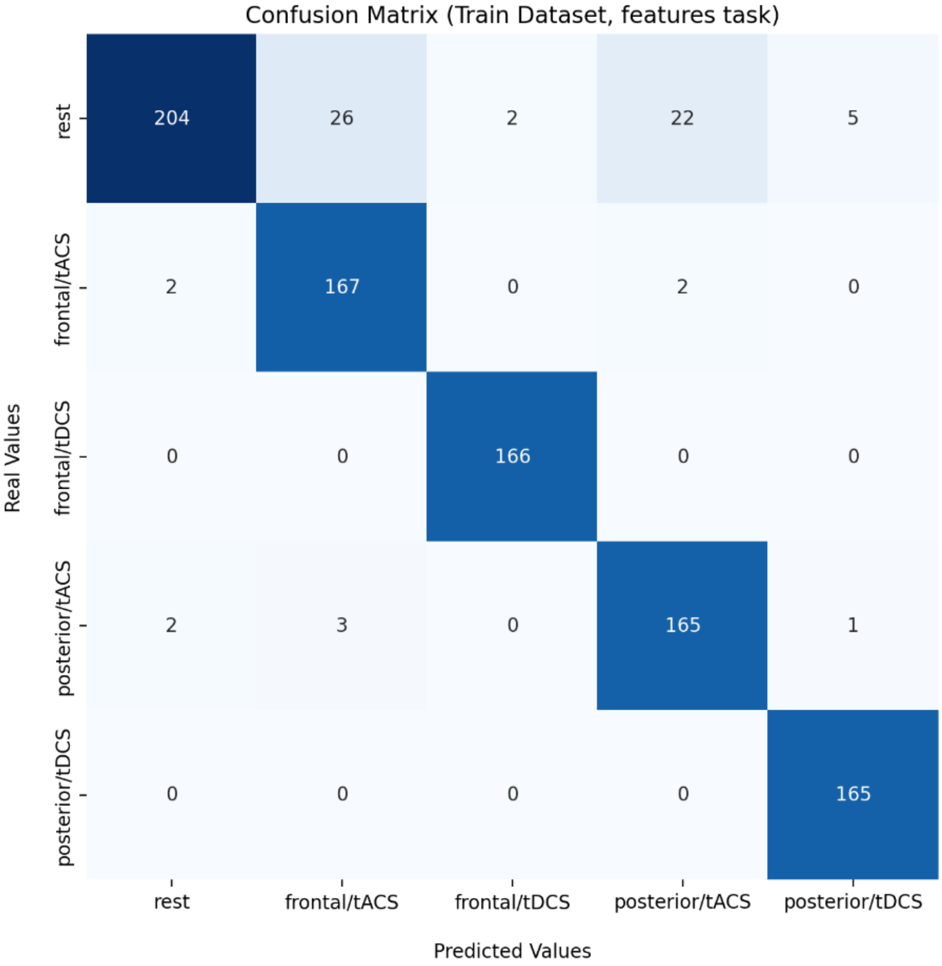
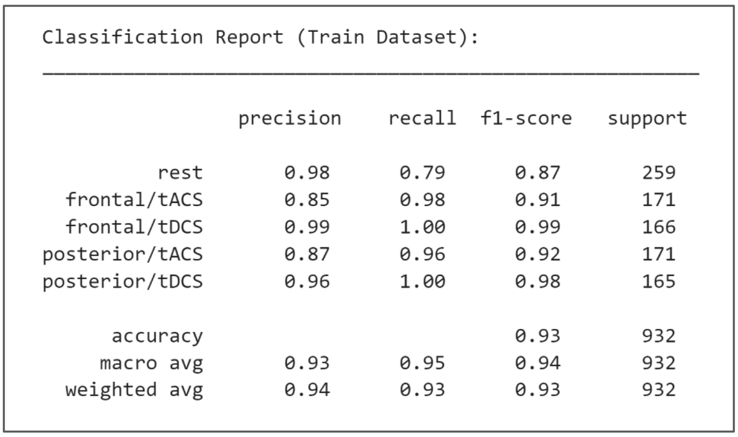


Figure 12. (A) Plot of the highest inter-subject accuracy progression observed during training of our algorithms, achieved using the aggregated EEG statistical features from 10 participants. (B) Plot of the loss evolution for the same CNN as (A). Note that despite the high accuracies of these models, post-hoc analysis revealed they had significantly overfitted to the distribution of 10 participants, even after only a few training epochs. Including data from more participants will be an essential aspect to avoid this issue in future works.





```
Classification Report (Train Dataset):
________________________________________________________

                precision    recall  f1-score   support

          rest       0.98      0.79      0.87       259
  frontal/tACS       0.85      0.98      0.91       171
  frontal/tDCS       0.99      1.00      0.99       166
posterior/tACS       0.87      0.96      0.92       171
posterior/tDCS       0.96      1.00      0.98       165

      accuracy                           0.93       932
     macro avg       0.93      0.95      0.94       932
  weighted avg       0.94      0.93      0.93       932
```

Figure 13. Confusion matrix (top) and Classification Report (bottom) for the EEG Training dataset and CNN model considered in the previous Figure 12. Note how tDCS is classified almost perfectly with an impressive 98% F1-score, while performance drops moderately for resting state and tACS conditions, with 87% and 91% respective F1-scores.

Yet, the golden metric to answer our research question would in fact be the accuracy attained on the Holdout dataset: how reliably can our best algorithm classify data from a participant that has been *altogether never seen* during training (in contrast to the Validation dataset of *unseen* data from *seen* participants)?

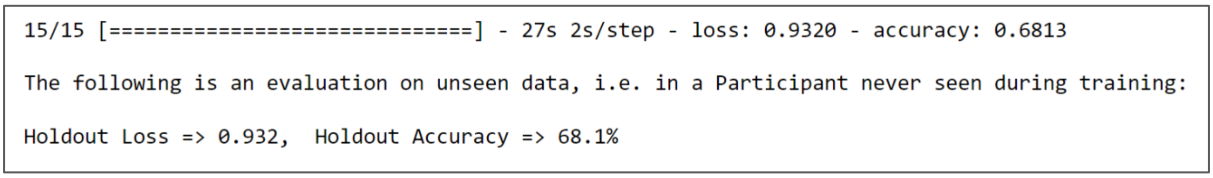
Answering this question posed a significant challenge, because while Holdout accuracy may be the most relevant metric, it is also the trickiest. DL algorithms are able to learn highly accurate statistical representations and generalize to new data coming from the same distribution seen during training, but they struggle to learn features that may make them more generalizable to data from an entirely new distribution, even if it is from a similar source (55). This is a however an inherent feature, since learning these out-of-distribution representations would come at the cost of steering the algorithms away from their lowest possible training losses (≅ highest accuracies), which is what the models are ultimately optimizing for (56). For all these reasons, initial tests on the Holdout dataset resulted in poor performance at ~40% accuracy, which was not significantly above the 28% mark for chance level accuracy.

Substantial research has been conducted on different tools used to prevent models from overfitting and losing their ability to generalize, but still, methods to promote generalization always carry an implicit trade-off for accuracy (57). We studied the effects of 2 different techniques: Dropout, which randomly switches off ('drops') neurons from the network during training to prevent them from adopting excessively fixed configurations (58); and L2 Regularization, which introduces a penalty in the training loss function (called regularization term) that favours setting less important neurons to 0 and prevents the network parameters from growing too large (59).

Nevertheless, implementing these methods did not have a positive impact overall. The best outcome resulted from choosing small Dropout values (5-10% after each convolutional layer), and even those translated in a sharp drop of Training and Validation accuracies from around 90% to ~65%, with Holdout accuracy increasing only slightly to ~50%. For stronger regularization parameters and techniques, accuracies were always decreased to slightly above chance level.

However, simply reducing the size of the CNN architecture resulted in the best regularization technique for our particular case, with the only downside that training times doubled in order to reach highest accuracies. Prior to this discovery, we would observe our relatively large models (up to 300 million trainable parameters) exhibited a steady increase in accuracy for the first 8-12 training epochs, after which training accuracy would keep on growing while validation accuracy would start to drop rapidly, signalling that the model had clearly begun overfitting. With the fairly smaller models that constitute our final optimized CNNs (previously summarized in Table 7), overfitting was displaced to the 20th epoch and beyond, while Training and Validation accuracies remained above >85% and Holdout values crossed the >60% mark.

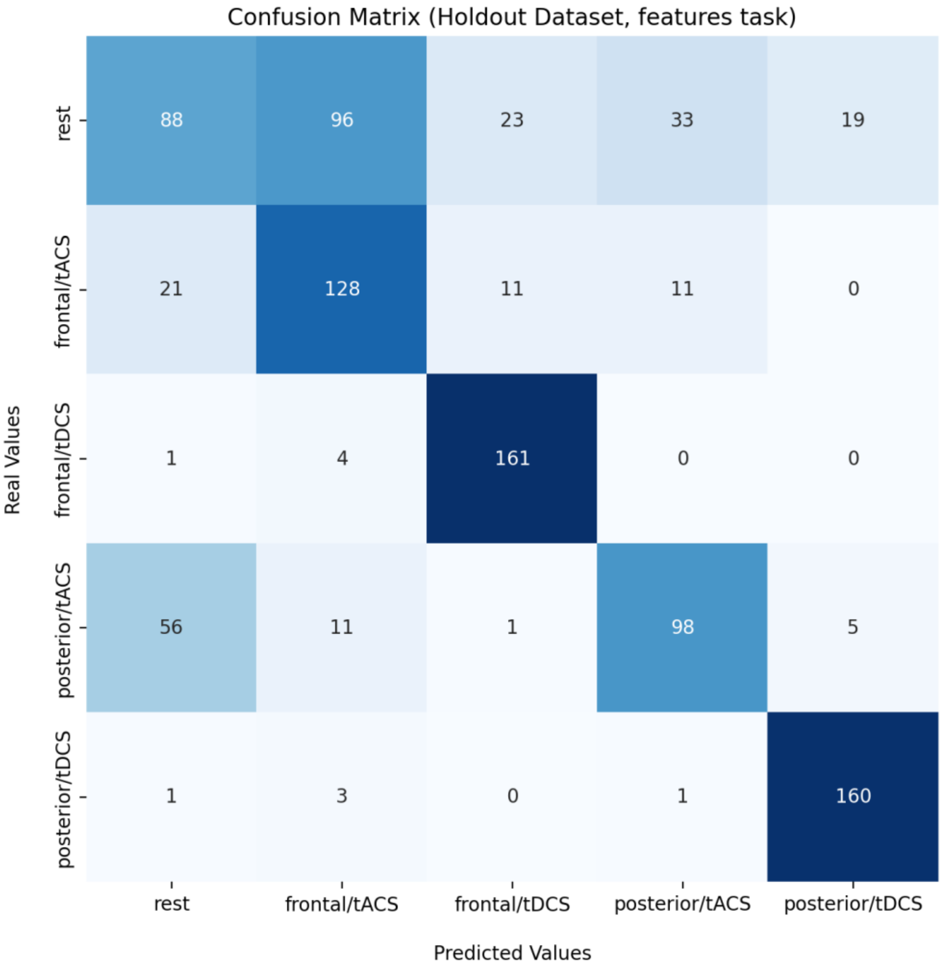
The best performance was achieved with the statistical features dataset, reaching a Holdout accuracy of 68.1% on a CNN model that had learned through 17 training epochs (Figure 14). For the time series classification task, the highest accuracy was 60.5% after 23 full-pass epochs.



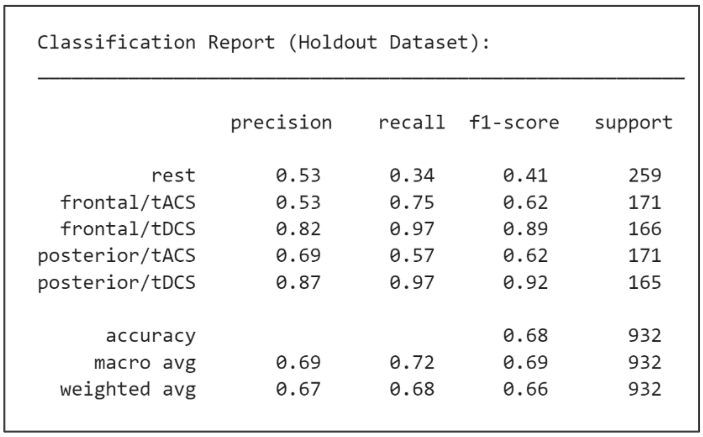
```
15/15 [==============================] - 27s 2s/step - loss: 0.9320 - accuracy: 0.6813

The following is an evaluation on unseen data, i.e. in a Participant never seen during training:

Holdout Loss => 0.932,  Holdout Accuracy => 68.1%
```

Figure 14. Results from evaluating the CNN model that achieved highest Holdout classification accuracy (68.1%).

To further analyse the underlying reasons behind the difference in Holdout and Training/Validation accuracies, we compared their confusion matrices (Figure 15 & Figure 13, respectively). The results confirmed our previous intuitions: the algorithms struggle to distinguish between resting state and tACS conditions, while they generalize fairly well to classify tDCS responses. In the next Section 3.4, we develop a deeper insight on the differences between these conditions in the spatio–temporal frequency domain.



	rest	frontal/tACS	frontal/tDCS	posterior/tACS	posterior/tDCS
rest	88	96	23	33	19
frontal/tACS	21	128	11	11	0
frontal/tDCS	1	4	161	0	0
posterior/tACS	56	11	1	98	5
posterior/tDCS	1	3	0	1	160



```
Classification Report (Holdout Dataset):
_______________________________________________________

                precision    recall  f1-score   support

          rest       0.53      0.34      0.41       259
  frontal/tACS       0.53      0.75      0.62       171
  frontal/tDCS       0.82      0.97      0.89       166
posterior/tACS       0.69      0.57      0.62       171
posterior/tDCS       0.87      0.97      0.92       165

      accuracy                           0.68       932
     macro avg       0.69      0.72      0.69       932
  weighted avg       0.67      0.68      0.66       932
```

Figure 15. Confusion matrix (top) and Classification Report (bottom) for the EEG Holdout dataset. Compared to Figure 13, the gaps between tDCS (89% F1-score), tACS (62%) and rest (41%) become much wider.

3.4 Artificial Neural Network Interpretability and Evoked Brain Dynamics

We studied the brain responses elicited in each of our 5 experimental conditions (‘rest’, ‘frontal/tDCS’, ‘frontal/tACS’, ‘posterior/tDCS’, ‘posterior/tACS’) across three dimensions:

- *Time-frequency analysis*: can we observe the effects of TES-induced neural dynamics in a spectrogram of whole EEG 20-seconds blocks? Do we find any TES-related remnants that may have confounded/misguided the DL classification? If so, for how long?
- *Feature importance*: which statistical descriptors had the most significant impact in the output of our models, for each of the 5 classes? How do they compare to each other?
- *Spatial distribution*: what are the channel importances for the DL classification? Are the most important channels adjacent to TES areas, or do we find them at distant locations?

Our time-frequency analysis on the EEG data rendered similar results for about 15 different session runs examined. We computed the spectrograms—shown in Figure 16 for an arbitrary participant/run combination—of entire 20-seconds ‘M’ and ‘C’ type blocks (i.e. EEG only, no TES). For more information regarding the meaning of these block labels, please refer to Section 2.2.

It should be noted that while the original lengths for both ‘M’ and ‘C’ rest blocks were defined to be exactly 20 seconds, our EEG preprocessing pipeline trimmed contaminated periods of 3-4 seconds immediately before and after any TES intervals, which corresponded to the recovery times for the GTEN amplifier (as evidenced in Figure 8). Consequently, the time lengths for the blocks considered in Figure 16 were effectively never longer than 16 seconds.

We observed that for resting state the averaged spectrogram presented a relatively flat distribution of frequency power, which seems fairly similar to tACS time-frequency figures (in both posterior and frontal areas). On the other hand, tDCS conditions showed well-defined spikes of frequency power, which appeared almost exclusively in the first half (<= 8 seconds) of the time intervals. It seems plausible that the results presented in Figure 13 and Figure 15 were derived from the DL algorithms picking up this characteristic or a signature of a similar kind, making the classification of tDCS conditions a relatively easier task compared to the distinction between rest and tACS classes.

Another interesting realization from Figure 16 is the apparent absence of any remnant electrical artifacts. Previous research reported durations for TES-related artifacts that extended in time for up to 2 seconds (60,61), leading us to believe that the margin of 4 seconds trimmed during our EEG preprocessing pipeline was sufficient to cancel all artifacts introduced by the GTEN amplifier.

Also, we had been previously concerned with the fact that our EEG preprocessing pipeline would not eliminate or even attenuate any electrical residue from our tACS stimulation, which we thought could have caused a remarkable peak at the 10Hz range in the order of mV (against our µV-range EEG signal). The lack of such artifact provides further proof that our choice of preprocessing parameters constitute a valid approach to eliminate preceding and successive TES-related electrical contamination, although it remains unanswered whether our chosen 20-seconds duration for the EEG blocks lead to the most optimal framework.

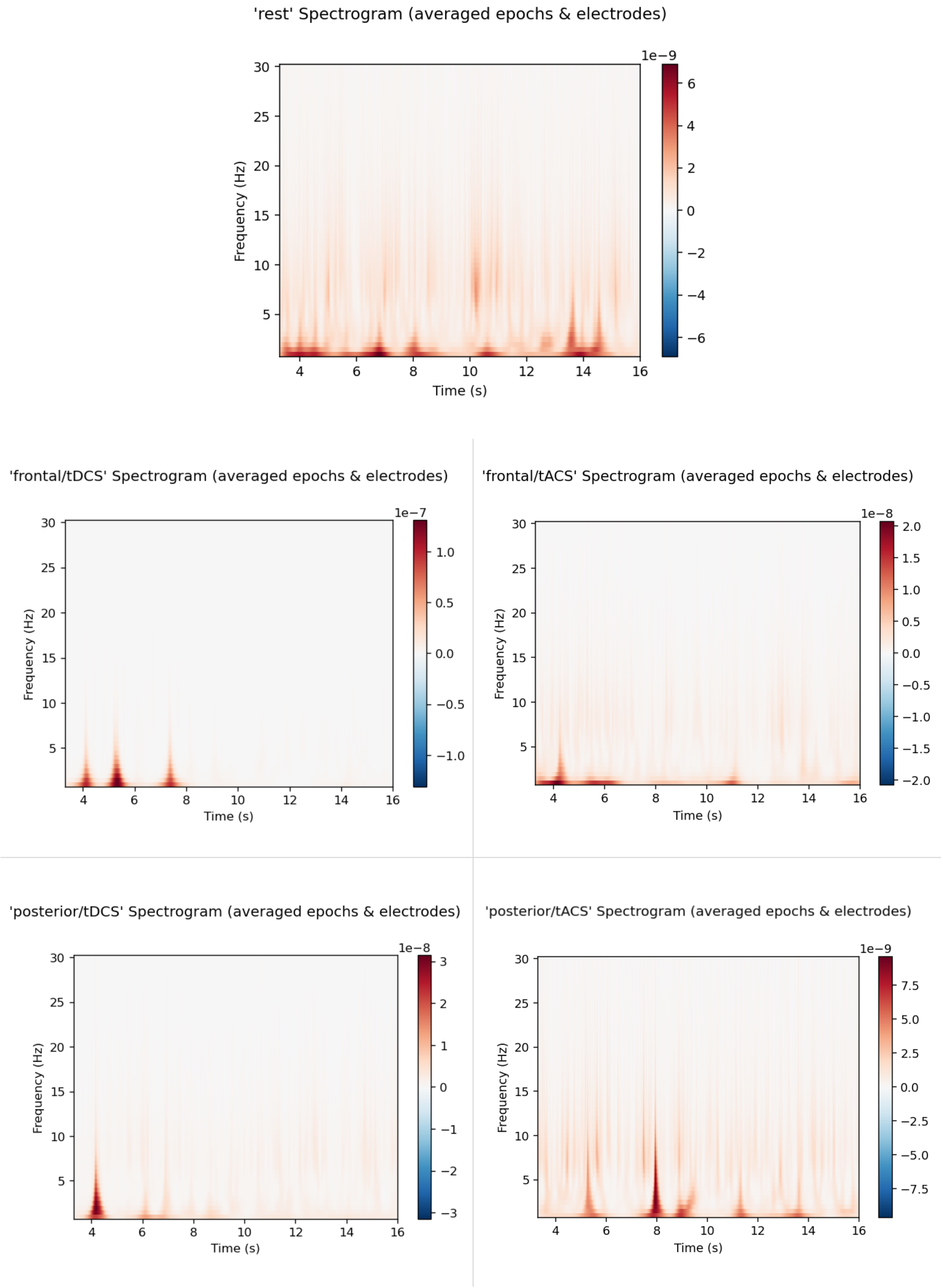


Figure 16. Spectrograms for whole-blocks corresponding to each of our 5 experimental conditions, with power values averaged across all electrodes and normalized. Resting state (top) shows a widespread low-magnitude (~10^{-9} W/Hz) spectrum, while posterior and frontal tDCS (left) contain localized spikes at the beginning of blocks (< 8 seconds), and tACS (right) also presents a widespread distribution of frequencies that somewhat resembles the 'rest' condition. These results align with our conclusions from the confusion matrices on the previous Section 3.3.

After analysing the time-frequency distribution of our EEG data, we sought to explore the internal structure of our CNN algorithms: given a particular input dataset, what was the process through which our models produced the corresponding set of predictions?

A well-known trade-off with DL algorithms is that while they typically exhibit superb performance for a wide array of problems, they do so by behaving as a black-box prediction model that makes it very difficult for humans to understand why and how a given decision was made (62). Fortunately, the AI research community has placed intensive effort in tackling this issue, with numerous solutions becoming readily available for DL researchers in recent years (63).

To gain better insights on the learning process of the CNN algorithms considered in this study, we used the open-source *tf-explain* package (64), which implements several interpretability methods designed specifically for TensorFlow neural networks. Since we were interested in visualizing which parts of the DL training examples (e.g. which statistical features) had the most impact in the classification outputs of our models, we decided to use the Gradient-weighted Class Activation Mapping (Grad-CAM) interpretability approach (65).

The outcomes from this analysis are a series of 2D activation maps with identical shape to their corresponding 2D training examples (i.e. 250x188 slices for the timeseries task, or 37x188 maps for the features classification). We computed these activation grids for every training example contained in each of our 3 EEG datasets, then averaged them and performed Min-Max normalization (66). An example of the outcome is illustrated in Figure 17.

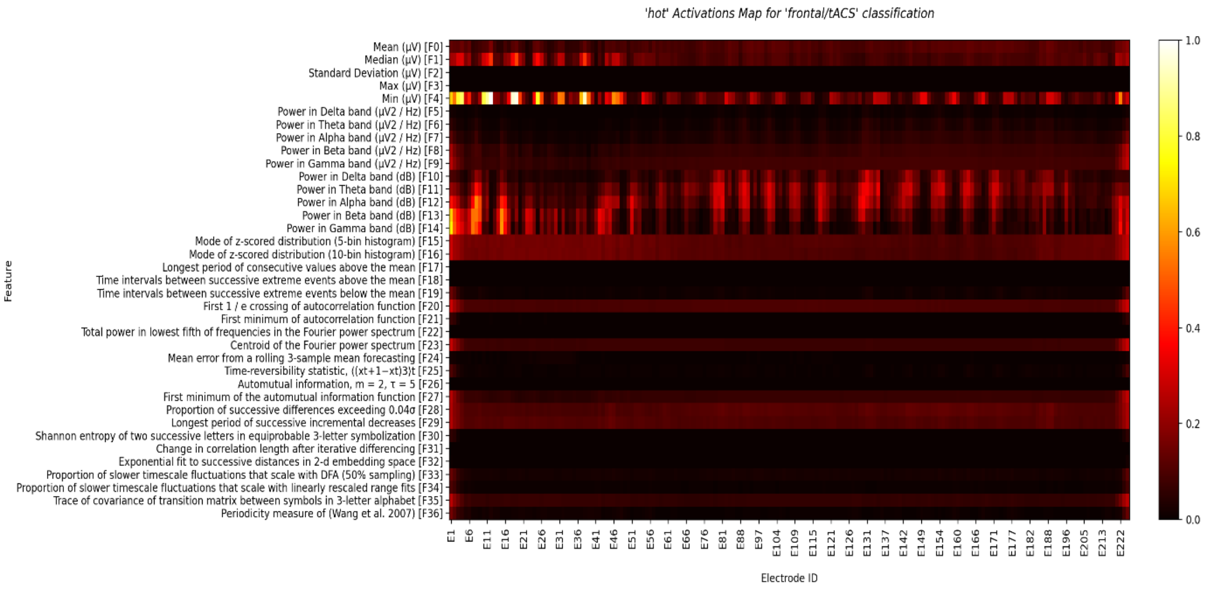


Figure 17. Example of an activation map produced with the Grad-CAM technique, containing normalized averages for activations in our Holdout dataset. This figure illustrates how our DL models make 'decisions', with max-intensity pixels corresponding to feature+electrode combinations that had the largest impact in predictions (65).

Crucially, from these activation maps we were able to infer the importance of each feature (Figure 18) and channel (also shown later in Figure 19) by computing their normalized average across the x- and y-axis, respectively.

We found that the most important features for our models with top-3 performance were common statistics (median, minimum and mean), the mode of z-scored distributions for 5-bin and 10-bin histograms, and the power in alpha and theta EEG bands (expressed in log scale of decibels). First, this observation indicates that different DL models converge to identify a similar set of 5-10 'preferred' features (i.e. above >0.5 normalized activations), which are used for every classification label. Second, this set of preferred descriptors appears to provide information mostly about two aspects of our data: *(i)* the *shape* of the EEG time series signal—mode, median, minimum, mean—and *(ii)* the logarithmic *distribution of power* across different frequency bands.

Other statistical measures that have been shown to produce good results in other types of time series data—such as those included in the *catch22* (43) Python package—seemed to not contribute significantly to the classification output, ranking below <0.5 on a normalized scale.

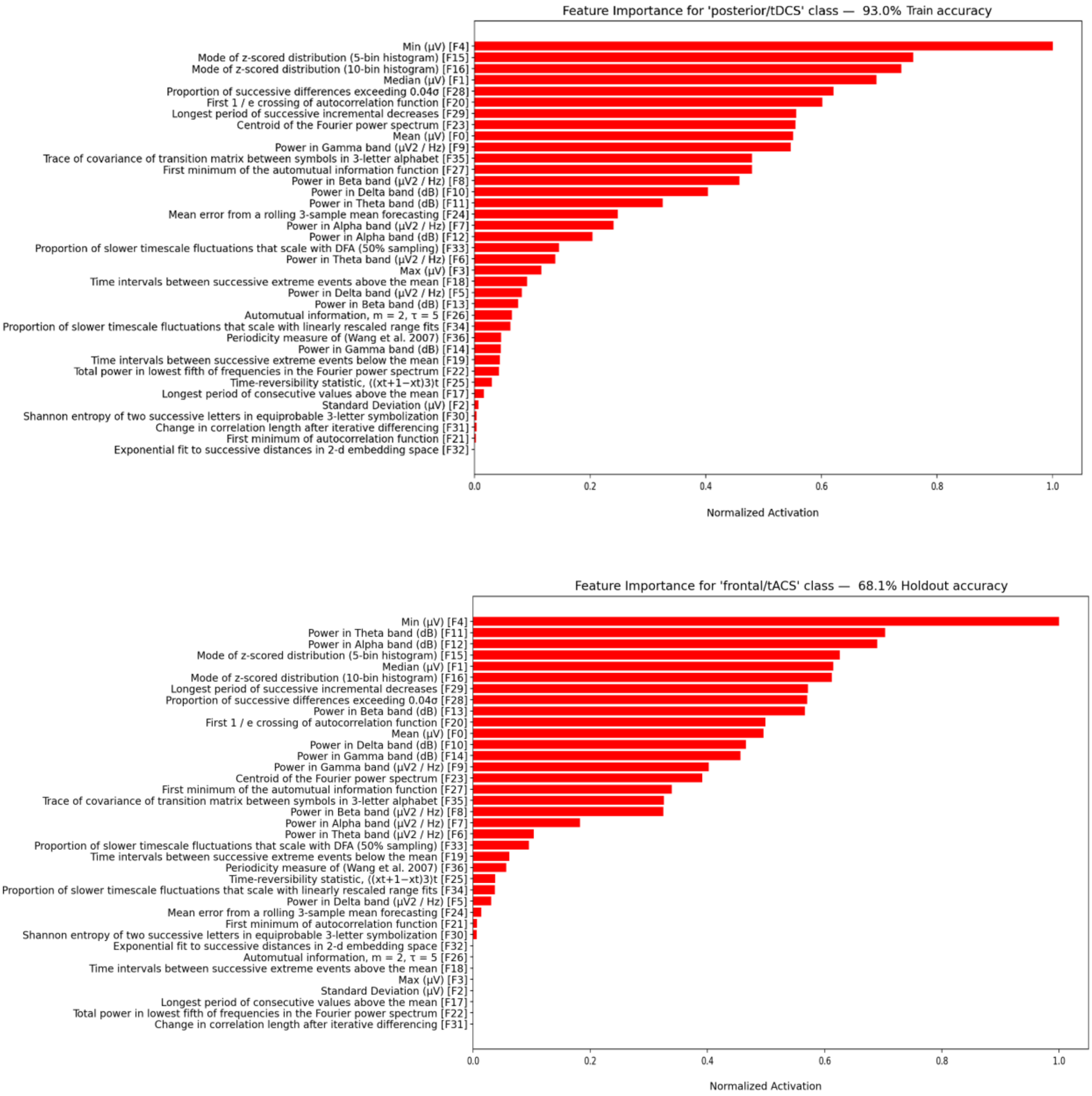


Figure 18. Ranking of normalized Feature Importances, on a range from 0 to 1. For this example we illustrated the 'posterior/tDCS' and 'frontal/tACS' conditions, using the same DL model and EEG datasets considered in Figure 13 (top panel in this figure) and Figure 15 (bottom).

Following a similar approach to the features, we averaged the activation maps to produce EEG topographic maps (topomaps) that illustrated the distribution of channel importances. We expected this information would provide insights around the spatial dynamics evoked after different stimulations. However, these topomaps did not deliver a clear pattern of localized brain features (Figure 19).

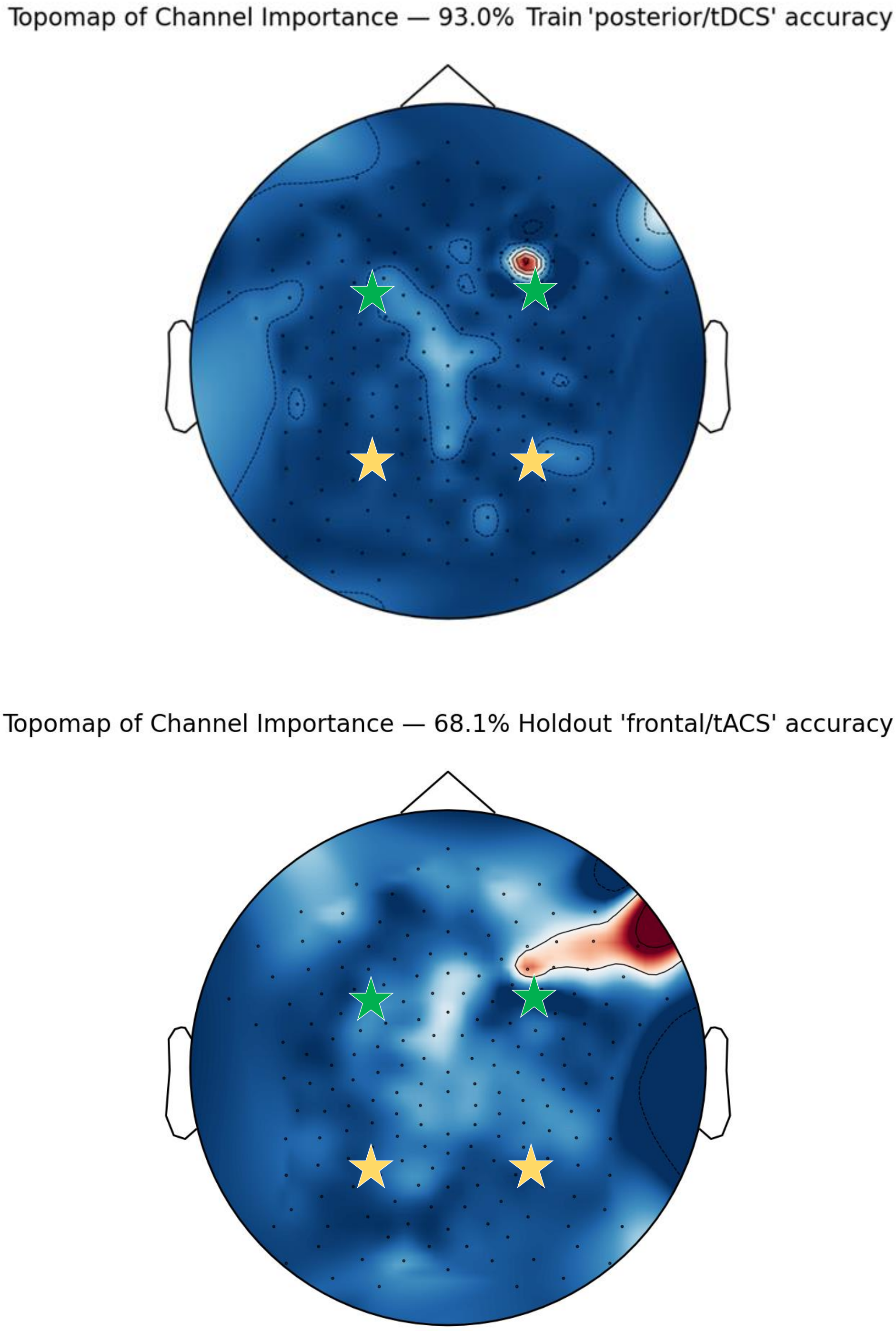


Figure 19. Spatial distribution of channel importances for our model's classification, overlaid over an illustration of EGI's 256-electrodes cap. Stimulation areas have been highlighted in green (frontal) and yellow (posterior). Note that the labels, model, and datasets considered here correspond to those from Figure 18.

3.5 Discussion, Limitations and Future Work

We collected, preprocessed and analysed EEG from 11 participants, across 4 stimulation and 1 control conditions. We found that delivering tDCS pulses in posterior cortical areas provided the most classifiable brain responses, with an F1-score of 92% for the features classification task on the Holdout dataset (i.e. EEG data from a participant unseen on training). Comparatively, Holdout performance only reached 62% and 41% F1-scores for tACS and resting state conditions.

It is possible that the effects induced by tACS may be more transient for relatively short durations (12,67), such as the 20-second blocks administered in our study. This phenomenon would provide a sound explanation as to why our DL algorithms struggled to learn a well-defined distinction between resting state and tACS stimulation. To further investigate this line of reasoning, it would be interesting to study the effects of changing the stimulation/rest block duration from 20 seconds to other values. How would the performance of our DL algorithms be affected by choosing a 5-, 10-, 15- or 30-seconds interval instead?

The top-5 most important features were found to be common summary statistics—namely the minimum, mode, and median values—as well as the decibel power in alpha and theta EEG bands.

In contrast, our analysis for electrode importance was non-conclusive due to wildly varying spatial distributions in the topomaps. It is likely that our DL algorithms learned an intricate internal structure of the EEG training data which is not readily available for human interpretation. Future work could try to find a meaningful pattern in the spatial distribution of most important channels by reproducing such figures using a substantially larger number of participants, or even by establishing a comparison with other topomaps produced via traditional ML algorithms.

With these considerations in mind, we would advise that future studies choose tDCS stimulation over tACS (perhaps investigating an additional type of stimulation such as tRNS or tPCS). With respect to selecting cortical regions for TES delivery, we found posterior areas resulted in slightly higher (+3%) classification accuracies, which matches our initial hypothesis from Section 1.6 regarding brain controllability (29).

Overall, our CNN algorithms successfully classified evoked EEG responses from preceding periods of stimulation, with accuracies of 60% and 68% on Holdout data for timeseries- and features-based datasets, respectively. For the Training and Validation sets, our best models reached an F1-score in excess of >90%.

However, given the relatively small size of our final dataset (2.8 hours total of aggregated eligible EEG), the full extent of the validity and applicability of our results remains an open question. While the outcomes of this pilot study have revealed a promising line of research, clinical translation would require replicating our results in a much larger dataset—at least by an order of magnitude in terms of number of participants and hours of valid EEG data.

This estimate is based on a comprehensive systematic review from Roy et al. (1) where they analysed 154 papers on DL-EEG approaches, providing valuable insights on different methods and their corresponding results. From the data summarized in Figure 20, we estimated that a robust dataset size for clinically viable performance would be in excess of >100 hours of valid training EEG data.

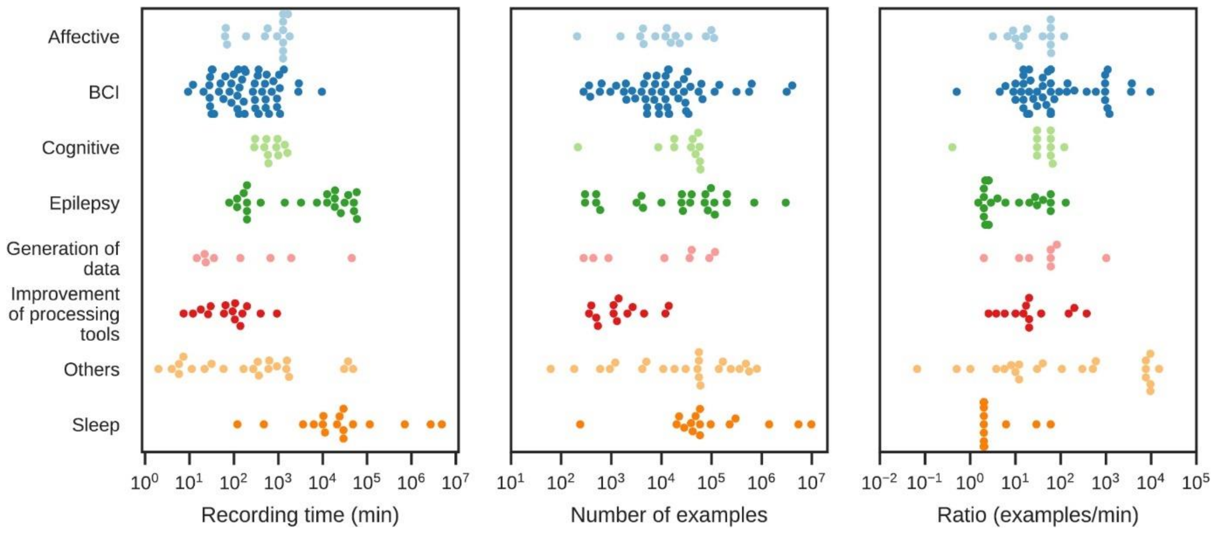


Figure 20. Amounts of EEG data used in more than 150 DL-EEG publications, applied to different neuroscience domains over the span of 2010 – 2018 (1).

Increasing the size of our dataset will provide numerous advantages, mainly by delivering EEG from a substantial number of participants—i.e. data distributions—that favour the DL model learning the 'true signal' corresponding to exclusively the TES-evoked brain responses. A larger amount of data will also enable applying regularization techniques such as Dropout (58) or L2 Regularization (59), aimed at making models generalize to unseen data from any new subject.

A larger-scale analysis will also enable soundly accepting or rejecting our hypothesis that different DL models converge to identify a set of 4-5 most important statistical features (as listed in Section 3.4). For future work one could study the impact of these preferred features by performing A/B testing, i.e. removing/adding features and analysing the corresponding change in performance. For example, what would happen if we removed the top-3 most important statistical features? Does the model find another set of preferred features? Or does it simply fail to learn an accurate statistical representation of the training data, causing accuracy to drop?

Another promising avenue for future work would consist of performing comprehensive 11-fold cross validation, with leave-N-out rotating participants for our Holdout Dataset. It would be interesting to ascertain any correlation between classification accuracy and variance in the training data. How does performance change with the dataset size and number of participants? Do outliers (e.g. *P004,* who was the only subject receiving 1mA max. amplitude) show any negative impact? Or perhaps they contribute most to the generalization capabilities of the DL algorithms?

Finally, for future work the author recommends relying on high-quality open-source solutions rather than on private ones. A major obstacle that severely hindered the progress of this project was the unreliability of the High-Performance Computing service chosen to host our EEG datasets and train our DL models. The suggested solution is using fully open-source platforms with a track record of great community adoption: GitHub to host code and small files, Kaggle for large files (such as EEG recordings), and Google Colaboratory to run code and train any DL algorithms using TensorFlow.

4. Conclusions

This study demonstrated the feasibility of a framework using DL algorithms to classify EEG brain responses evoked by a defined pattern of TES neurostimulation. Our best CNN models reached a 68.1% Holdout classification accuracy for the average of all stimulation/control conditions, which is comparable to an estimated human-level performance at 60-70%, based on the clinical misdiagnosis rate reported by Giacino et al (13). Further research and larger EEG-TES datasets will be required to ascertain whether general Holdout performance could be increased to approach overall Training and Validation accuracies of >90%.

More importantly, we found that delivering tDCS pulses in posterior cortical areas—located inside the consciousness 'hot zone' identified in numerous recent publications (4,23,24)—elicited an exceptionally reliable brain response. Our CNN models achieved for this particular condition a remarkable 92% F1-score (87% precision; 97% recall) on Holdout data from participants that were never seen during training, allowing us to confidently accept the central hypothesis of this thesis: *Can a Deep Learning algorithm even distinguish the brain responses evoked from different types of electrical stimulation in distant cortical regions?*

Finally, in the spirit of contributing to the progress of a relatively young DL-EEG discipline, this project was open-sourced in full, to be used freely by the neuroscience and artificial intelligence research communities. The code may be downloaded from GitHub (32,36), while both our EEG time series and features-based datasets can be found in Kaggle (33,44). The author will keep updating these links in coming months with improved comments and metadata, with the intention of providing step-by-step guidance to potential researchers that may be interested in using the code and/or datasets here presented.

References

1. Roy Y, Banville H, Albuquerque I, Gramfort A, Falk TH, Faubert J. Deep learning-based electroencephalography analysis: a systematic review. J Neural Eng. 2019 Aug;16(5):051001.

2. Harati A, López S, Obeid I, Picone J, Jacobson MP, Tobochnik S. The TUH EEG CORPUS: A big data resource for automated EEG interpretation. In: 2014 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). 2014. p. 1–5.

3. Overgaard M. The Status and Future of Consciousness Research. Frontiers in Psychology. 2017;8(1719).

4. Koch C, Massimini M, Boly M, Tononi G. Neural correlates of consciousness: progress and problems. Nature Reviews Neuroscience. 2016 May;17(5):307–21.

5. Cavanna AE, Monaco F. Brain mechanisms of altered conscious states during epileptic seizures. Nature Reviews Neurology. 2009 May;5(5):267–76.

6. Overgaard M. How can we know if patients in coma, vegetative state or minimally conscious state are conscious? In: Laureys S, Schiff ND, Owen AM, editors. Progress in Brain Research. Elsevier; 2009. p. 11–9. (Coma Science: Clinical and Ethical Implications; vol. 177).

7. Harrison AH, Connolly JF. Finding a way in: A review and practical evaluation of fMRI and EEG for detection and assessment in disorders of consciousness. Neuroscience & Biobehavioral Reviews. 2013 Sep 1;37(8):1403–19.

8. Owen AM, Coleman MR, Boly M, Davis MH, Laureys S, Pickard JD. Detecting awareness in the vegetative state. Science. 2006 Sep 8;313(5792):1402.

9. Glover GH. Overview of Functional Magnetic Resonance Imaging. Neurosurg Clin N Am. 2011 Apr;22(2):133–9.

10. Sturzbecher MJ, de Araujo DB. Simultaneous EEG-fMRI: Integrating Spatial and Temporal Resolution. In: Rao AR, Cecchi GA, editors. The Relevance of the Time Domain to Neural Network Models. Boston, MA: Springer US; 2012. p. 199–217. (Springer Series in Cognitive and Neural Systems).

11. Napolitani M, Bodart O, Canali P, Seregni F, Casali A, Laureys S, et al. Transcranial magnetic stimulation combined with high-density EEG in altered states of consciousness. Brain Injury. 2014 Aug 1;28(9):1180–9.

12. Reed T, Cohen Kadosh R. Transcranial electrical stimulation (tES) mechanisms and its effects on cortical excitability and connectivity. J Inherit Metab Dis. 2018;41(6):1123–30.

13. Giacino JT, Fins JJ, Laureys S, Schiff ND. Disorders of consciousness after acquired brain injury: the state of the science. Nat Rev Neurol. 2014 Feb;10(2):99–114.

14. Bayne T, Hohwy J, Owen AM. Are There Levels of Consciousness? Trends in Cognitive Sciences. 2016 Jun 1;20(6):405–13.

15. Sergent C, Faugeras F, Rohaut B, Perrin F, Valente M, Tallon-Baudry C, et al. Multidimensional cognitive evaluation of patients with disorders of consciousness using EEG: A proof of concept study. NeuroImage: Clinical. 2017 Jan 1;13:455–69.

16. Sitt JD, King J-R, El Karoui I, Rohaut B, Faugeras F, Gramfort A, et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. Brain. 2014 Aug 1;137(8):2258–70.

17. Demertzi A, Tagliazucchi E, Dehaene S, Deco G, Barttfeld P, Raimondo F, et al. Human consciousness is supported by dynamic complex patterns of brain signal coordination. Science Advances. 2019 Feb 1;5(2):eaat7603.

18. Casali AG, Gosseries O, Rosanova M, Boly M, Sarasso S, Casali KR, et al. A Theoretically Based Index of Consciousness Independent of Sensory Processing and Behavior. Science Translational Medicine. 2013 Aug 14;5(198):198ra105.

19. Kida T, Tanaka E, Kakigi R. Multi-Dimensional Dynamics of Human Electromagnetic Brain Activity. Front Hum Neurosci. 2015;9:713.

20. Bassett DS, Sporns O. Network neuroscience. Nature Neuroscience. 2017 Mar;20(3):353–64.

21. Bai Y, Xia X, Li X. A Review of Resting-State Electroencephalography Analysis in Disorders of Consciousness. Frontiers in Neurology. 2017;8.

22. Hill AT, Rogasch NC, Fitzgerald PB, Hoy KE. TMS-EEG: A window into the neurophysiological effects of transcranial electrical stimulation in non-motor brain regions. Neuroscience & Biobehavioral Reviews. 2016 May 1;64:175–84.

23. Boly M, Massimini M, Tsuchiya N, Postle BR, Koch C, Tononi G. Are the Neural Correlates of Consciousness in the Front or in the Back of the Cerebral Cortex? Clinical and Neuroimaging Evidence. J Neurosci. 2017 Oct 4;37(40):9603–13.

24. Koch C, Massimini M, Boly M, Tononi G. Posterior and anterior cortex — where is the difference that makes the difference? Nature Reviews Neuroscience. 2016 Oct;17(10):666–666.

25. Siclari F, Baird B, Perogamvros L, Bernardi G, LaRocque JJ, Riedner B, et al. The neural correlates of dreaming. Nature Neuroscience. 2017 Jun;20(6):872–8.

26. Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (EEG) classification tasks: a review. J Neural Eng. 2019 Apr;16(3):031001.

27. Lee M, Yeom S, Baird B, Gosseries O, Nieminen JO, Tononi G, et al. Spatio-temporal analysis of EEG signal during consciousness using convolutional neural network. In: 2018 6th International Conference on Brain-Computer Interface (BCI). 2018. p. 1–3.

28. Ullah I, Hussain M, Qazi E-H, Aboalsamh H. An automated system for epilepsy detection using EEG brain signals based on deep learning approach. Expert Systems with Applications. 2018 Oct 1;107:61–71.

29. Deco G, Cruzat J, Cabral J, Tagliazucchi E, Laufs H, Logothetis NK, et al. Awakening: Predicting external stimulation to force transitions between different brain states. PNAS. 2019 Sep 3;116(36):18088–97.

30. Rolls ET, Huang C-C, Lin C-P, Feng J, Joliot M. Automated anatomical labelling atlas 3. Neuroimage. 2020 Feb 1;206:116189.

31. Geodesic Transcranial Electrical Neuromodulation 200 (GTEN 200) [Internet]. Electrical Geodesics, Inc. [cited 2021 Sep 9]. Available from: https://www.egi.com/gten-100-research-neuromodulation-system/neuromodulation-eeg-systems

32. Pomares Pastor A. Auditory Oddball framework in Python [Internet]. GitHub. [cited 2021 Sep 8]. Available from: https://github.com/alexispomares/auditory-oddball

33. Pomares Pastor A. DISSERTATION — Raw EEG Dataset [Internet]. Kaggle. [cited 2021 Sep 9]. Available from: https://kaggle.com/alexispomares/dissertation-raw

34. Gramfort A, Luessi M, Larson E, Engemann D, Strohmeier D, Brodbeck C, et al. MNE software for processing MEG and EEG data. Neuroimage. 2014 Feb 1;86:446–60.

35. Open-source Python package for exploring, visualizing, and analyzing human neurophysiological data. [Internet]. MNE. [cited 2021 Sep 11]. Available from: https://mne.tools/stable/index.html

36. Pomares Pastor A. Deep Learning Classification of EEG Responses to Transcranial Electrical Stimulation [Internet]. GitHub. [cited 2021 Sep 11]. Available from: https://github.com/alexispomares/DL-EEG-TES

37. Rolnick D, Veit A, Belongie S, Shavit N. Deep Learning is Robust to Massive Label Noise. arXiv:170510694 [cs] [Internet]. 2018 Feb 26; Available from: http://arxiv.org/abs/1705.10694

38. Dowsett J, Herrmann CS. Transcranial Alternating Current Stimulation with Sawtooth Waves: Simultaneous Stimulation and EEG Recording. Frontiers in Human Neuroscience. 2016;10:135.

39. Kohli S, Casson AJ. Removal of Gross Artifacts of Transcranial Alternating Current Stimulation in Simultaneous EEG Monitoring. Sensors. 2019 Jan;19(1):190.

40. Alagapan S, Shin HW, Fröhlich F, Wu H. Diffusion geometry approach to efficiently remove electrical stimulation artifacts in intracranial electroencephalography. J Neural Eng. 2019 Apr;16(3):036010.

41. Bandpower of an EEG signal [Internet]. [cited 2021 Sep 13]. Available from: https://raphaelvallat.com/bandpower.html

42. Lubba CH. catch22 - CAnonical Time-series CHaracteristics [Internet]. GitHub. 2021 [cited 2021 Sep 11]. Available from: https://github.com/chlubba/catch22

43. Lubba CH, Sethi SS, Knaute P, Schultz SR, Fulcher BD, Jones NS. catch22: CAnonical Time-series CHaracteristics. Data Min Knowl Disc. 2019 Nov 1;33(6):1821–52.

44. Pomares Pastor A. DISSERTATION — Preprocessed EEG Dataset [Internet]. Kaggle. [cited 2021 Sep 9]. Available from: https://kaggle.com/alexispomares/dissertation-preprocessed

45. TensorFlow, an end-to-end open source platform for machine learning [Internet]. TensorFlow. [cited 2021 Sep 11]. Available from: https://www.tensorflow.org/

46. Ashford J, Bird JJ, Campelo F, Faria DR. Classification of EEG Signals Based on Image Representation of Statistical Features. In: Ju Z, Yang L, Yang C, Gegov A, Zhou D, editors. Advances in Computational Intelligence Systems. Cham: Springer International Publishing; 2020. p. 449–60. (Advances in Intelligent Systems and Computing).

47. Chen JX, Jiang DM, Zhang YN. A Hierarchical Bidirectional GRU Model With Attention for EEG-Based Emotion Classification. IEEE Access. 2019;7:118530–40.

48. Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv:14091556 [cs] [Internet]. 2015 Apr 10; Available from: http://arxiv.org/abs/1409.1556

49. Schwiedrzik C. Retina or visual cortex? The site of phosphene induction by transcranial alternating current stimulation. Frontiers in Integrative Neuroscience. 2009;3:6.

50. Kar K, Krekelberg B. Transcranial electrical stimulation over visual cortex evokes phosphenes with a retinal origin. Journal of Neurophysiology. 2012 Oct 15;108(8):2173–8.

51. Khadka N, Bikson M. Role of skin tissue layers and ultra-structure in transcutaneous electrical stimulation including tDCS. Phys Med Biol. 2020 Nov;65(22):225018.

52. Horvath J, Carter O, Forte J. Transcranial direct current stimulation: five important issues we aren’t discussing (but probably should be). Frontiers in Systems Neuroscience. 2014;8:2.

53. Wang R, Wang J, Yu H, Wei X, Yang C, Deng B. Power spectral density and coherence analysis of Alzheimer’s EEG. Cogn Neurodyn. 2015 Jun;9(3):291–304.

54. Carrier J, Land S, Buysse DJ, Kupfer DJ, Monk TH. The effects of age and gender on sleep EEG power spectral density in the middle years of life (ages 20–60 years old). Psychophysiology. 2001 Mar;38(2):232–42.

55. Jin P, Lu L, Tang Y, Karniadakis GE. Quantifying the generalization error in deep learning in terms of data distribution and neural network smoothness. Neural Networks. 2020 Oct 1;130:85–99.

56. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015 May;521(7553):436–44.

57. Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning (still) requires rethinking generalization. Commun ACM. 2021 Mar;64(3):107–15.

58. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. :30.

59. Phaisangittisagul E. An Analysis of the Regularization Between L2 and Dropout in Single Hidden Layer Neural Network. In: 2016 7th International Conference on Intelligent Systems, Modelling and Simulation (ISMS). 2016. p. 174–9.

60. Miniussi C, Brignani D, Pellicciari MC. Combining Transcranial Electrical Stimulation With Electroencephalography: A Multimodal Approach. Clin EEG Neurosci. 2012 Jul 1;43(3):184–91.

61. Liu A, Vöröslakos M, Kronberg G, Henin S, Krause MR, Huang Y, et al. Immediate neurophysiological effects of transcranial electrical stimulation. Nat Commun. 2018 Nov 30;9(1):5092.

62. Rai A. Explainable AI: from black box to glass box. J of the Acad Mark Sci. 2020 Jan 1;48(1):137–41.

63. Li X, Xiong H, Li X, Wu X, Zhang X, Liu J, et al. Interpretable Deep Learning: Interpretation, Interpretability, Trustworthiness, and Beyond. arXiv:210310689 [cs] [Internet]. 2021 May 11; Available from: http://arxiv.org/abs/2103.10689

64. Meudec R. tf-explain: Interpretability methods to ease understanding of neural networks in TensorFlow. [Internet]. GitHub. [cited 2021 Sep 13]. Available from: https://github.com/sicara/tf-explain

65. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. Int J Comput Vis. 2020 Feb;128(2):336–59.

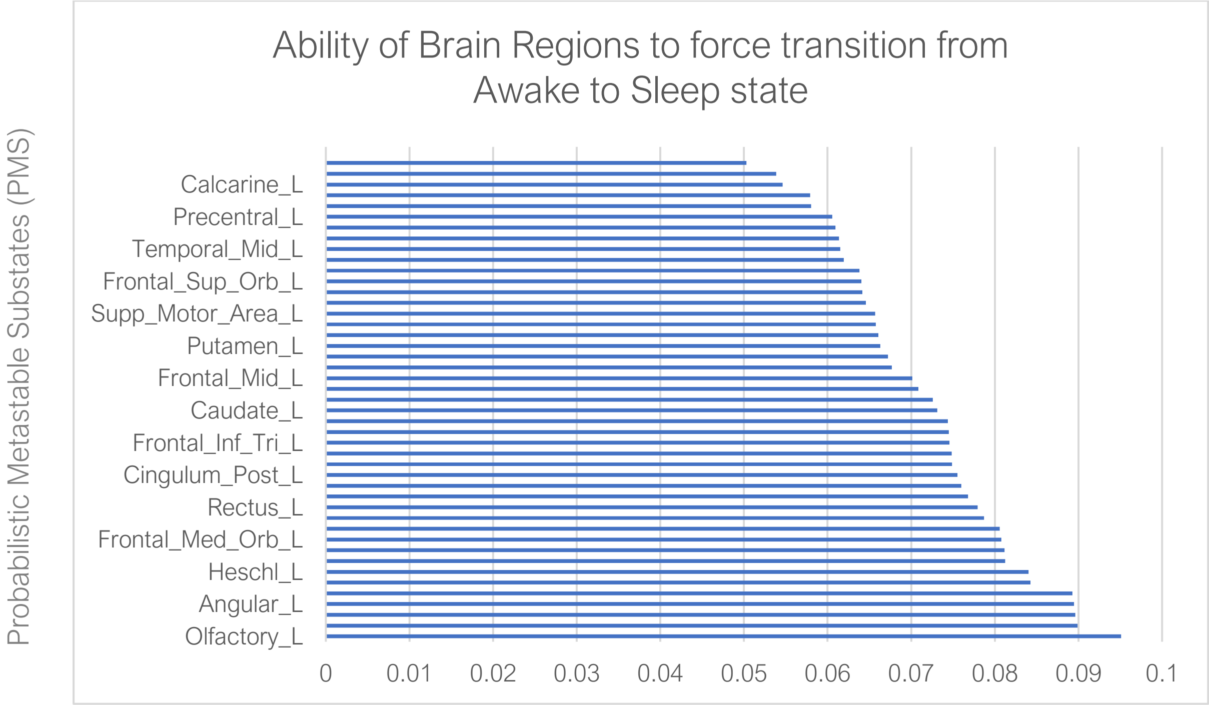
66. Patro SGK, Sahu KK. Normalization: A Preprocessing Stage. arXiv:150306462 [cs] [Internet]. 2015 Mar 19; Available from: http://arxiv.org/abs/1503.06462

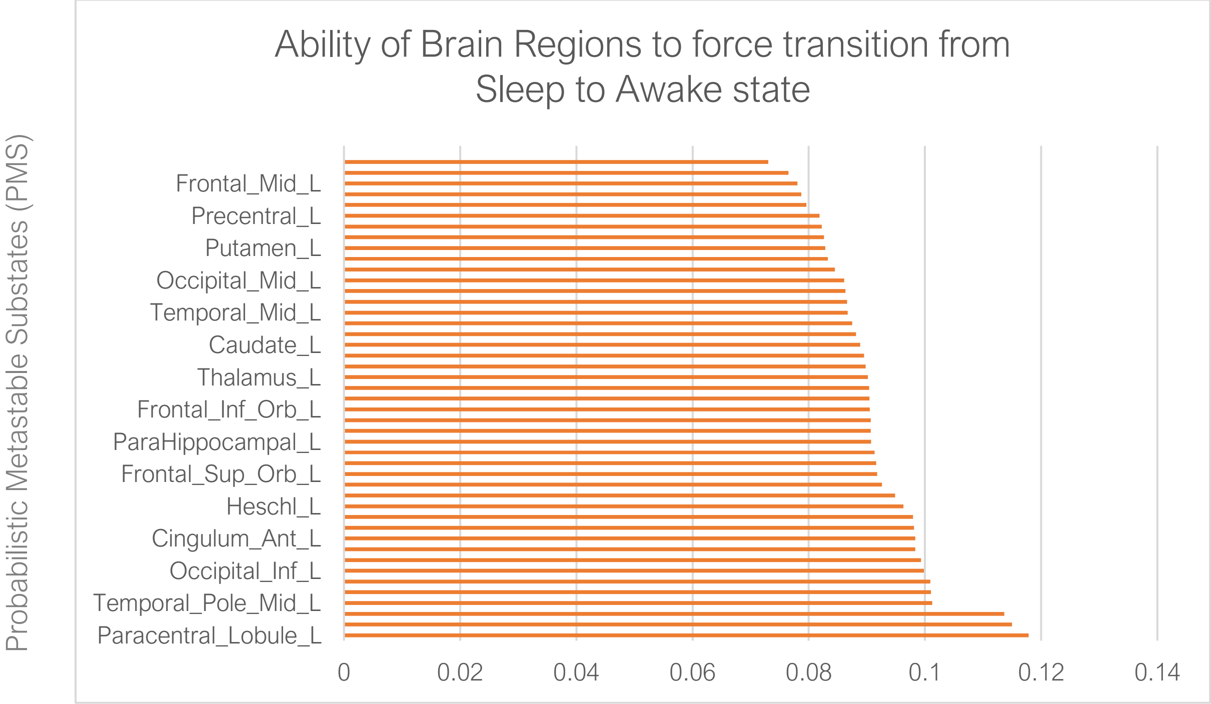
67. Moreno-Duarte I, Gebodh N, Schestatsky P, Guleyupoglu B, Reato D, Bikson M, et al. Chapter 2 - Transcranial Electrical Stimulation: Transcranial Direct Current Stimulation (tDCS), Transcranial Alternating Current Stimulation (tACS), Transcranial Pulsed Current Stimulation (tPCS), and Transcranial Random Noise Stimulation (tRNS). In: Cohen Kadosh R, editor. The Stimulated Brain [Internet]. San Diego: Academic Press; 2014. p. 35–59. Available from: https://www.sciencedirect.com/science/article/pii/B9780124047044000028

Annexes

Annex I

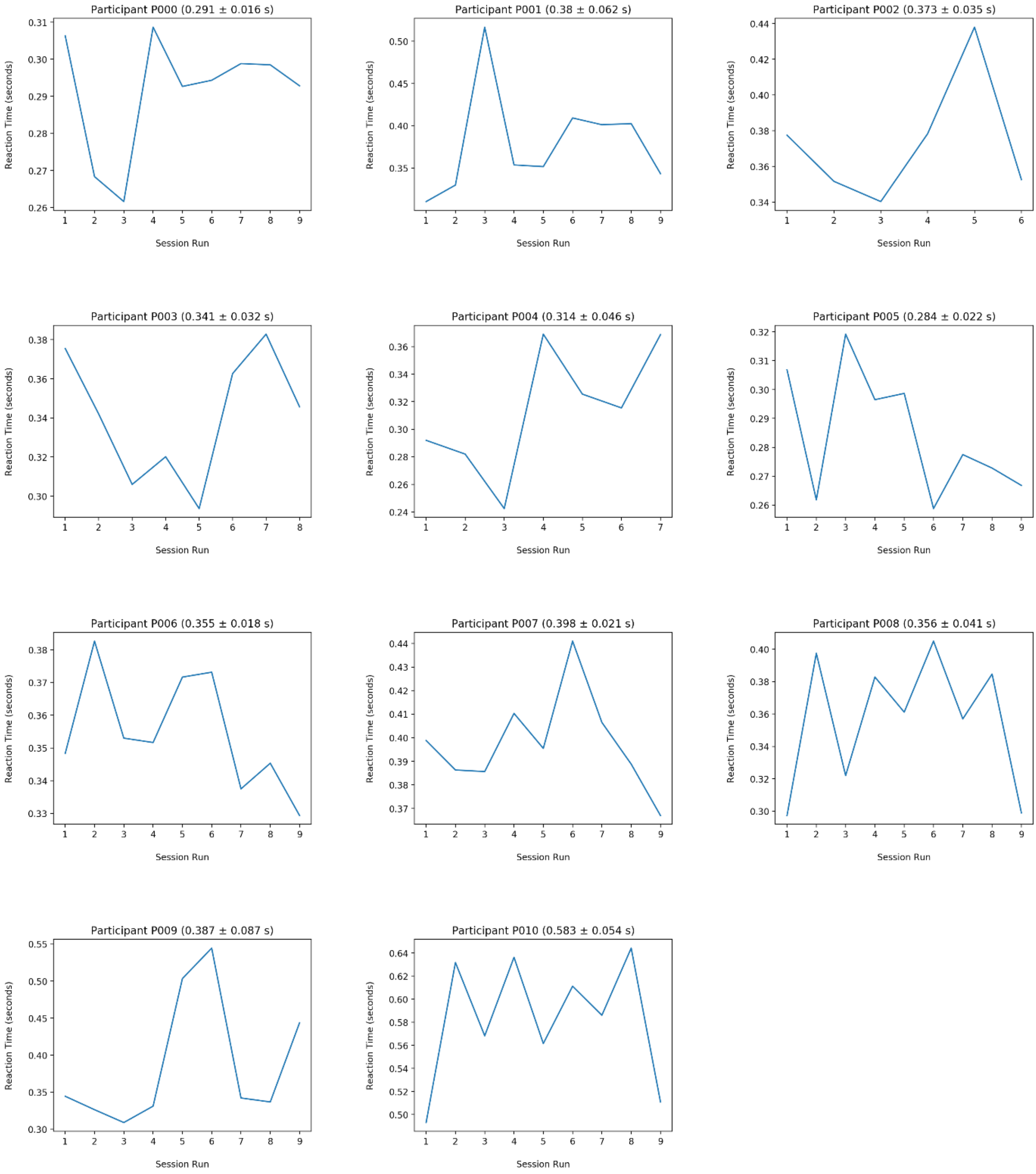
Figure 21. Summary of the data kindly provided by Dr. Deco from his work on transitions between brain states. Note that in both paradigms, small PMS values correspond to high abilities to promote transitions between states.





Annex II

Figure 22. Participant-wise results from the oddball paradigm at the end of every experimental run.


Participant P000 (0.291 ± 0.016 s)
Reaction Time (seconds)
Session Run
Participant P001 (0.38 ± 0.062 s)
Reaction Time (seconds)
Session Run
Participant P002 (0.373 ± 0.035 s)
Reaction Time (seconds)
Session Run
Participant P003 (0.341 ± 0.032 s)
Reaction Time (seconds)
Session Run
Participant P004 (0.314 ± 0.046 s)
Reaction Time (seconds)
Session Run
Participant P005 (0.284 ± 0.022 s)
Reaction Time (seconds)
Session Run
Participant P006 (0.355 ± 0.018 s)
Reaction Time (seconds)
Session Run
Participant P007 (0.398 ± 0.021 s)
Reaction Time (seconds)
Session Run
Participant P008 (0.356 ± 0.041 s)
Reaction Time (seconds)
Session Run
Participant P009 (0.387 ± 0.087 s)
Reaction Time (seconds)
Session Run
Participant P010 (0.583 ± 0.054 s)
Reaction Time (seconds)
Session Run