Ten quick tips for clinical electroencephalographic (EEG) data acquisition and signal processing

Giulia Cisotto[1,2] and Davide Chicco[1,3]

[1] Dipartimento di Informatica Sistemistica e Comunicazione, Università di Milano-Bicocca, Milan, Milan, Italy
[2] Dipartimento di Ingegneria dell'Informazione, Università di Padova, Padua, Padua, Italy
[3] Institute of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada

ABSTRACT

Electroencephalography (EEG) is a medical engineering technique aimed at recording the electric activity of the human brain. Brain signals derived from an EEG device can be processed and analyzed through computers by using digital signal processing, computational statistics, and machine learning techniques, that can lead to scientifically-relevant results and outcomes about how the brain works. In the last decades, the spread of EEG devices and the higher availability of EEG data, of computational resources, and of software packages for electroencephalography analysis has made EEG signal processing easier and faster to perform for any researcher worldwide. This increased ease to carry out computational analyses of EEG data, however, has made it easier to make mistakes, as well. And these mistakes, if unnoticed or treated wrongly, can in turn lead to wrong results or misleading outcomes, with worrisome consequences for patients and for the advancements of the knowledge about human brain. To tackle this problem, we present here our ten quick tips to perform electroencephalography signal processing analyses avoiding common mistakes: a short list of guidelines designed for beginners on what to do, how to do it, and what not to do when analyzing EEG data with a computer. We believe that following our quick recommendations can lead to better, more reliable and more robust results and outcome in clinical neuroscientific research.

INTRODUCTION

Electroencephalography (EEG) is a convenient and common tool to record the electrical activity of our brain in a non-invasive way, easily, at relatively low cost, and eventually using portable devices. The EEG allows to accurately follow the fast dynamic of the brain and to obtain quantitative measurements of the electrical activity that small portions of our brain produce while we are accomplishing cognitive and motor tasks, as well as while we are in resting-state. The electroencephalogram is a special case of electrogram, and should not be confused with the electrocardiogram (ECG), which is the recording of the electrical activity of the heart.

The EEG differs from other kinds of brain monitoring, for example functional magnetic resonance imaging (fMRI) or magnetoencephalography, to need cheaper costs, lighter and

simpler acquisition procedures, faster acquisition sessions, and to be portable. For these reasons, EEG is one of the most popular methods to investigate the brain and tons of algorithms have already been proposed to analyze this kind of signal.

The first quantitative analysis of an EEG signal dates back to the pioneering work of Hans Berger who, in 1929, employed a Fourier transform of an EEG signal to quantify the spectral distribution of the brain activity under different physiological and stimulation conditions (*Berger, 1929*). Since then, a vast literature flourished and obtained successful achievements in the modeling and classification of EEG data for different clinical and research applications (*Teplan, 2002*). However, the use of electroencephalography presents several challenges, with non-trivial choices to design and implement the acquisition steps, and a critical, yet underestimated, pre-processing phase.

Researchers take advantage of EEG data processing in several scientific fields, and not only in medicine: EGG data in fact are the core of analyses on emotion recognition (*Xu, Guo & Wang, 2022*), sleep stage classification (*Zaman et al., 2024*), motor imagery classification (*Zaman et al., 2024*; *Zancanaro et al., 2021*), neurorehabilitation (*Ang & Guan, 2016*), seizure detection (*Shen et al., 2024*), just to mention a few.

Several articles serve as guides to beginners on how to understand the electroencephalography (EEG). The studies of *Beniczky & Schomer (2020)* and *Biasiucci, Franceschiello & Murray (2019)*, for example, describe the basic features of the EEG, how its data are encoded and how they can be analyzed.

Some studies published in the past described the main challenges and obstacles of EEG signal computational processing (*Khademi, Ebrahimi & Kordy, 2023*; *Sharma et al., 2023*; *Rashid et al., 2020*; *Sun & Mou, 2023*; *Ein Shoka et al., 2023*). These research works capture efficiently the main issues of the EEG computational signal processing scenario, indicating problems that we ourselves describe in the present article as well. However, the authors of these studies do not provide practical solutions for handling these obstacles: we tackle this problem by providing here our ten simple tips to perform EEG signal processing correctly.

So far no article in the scientific literature explains how to perform EEG signal processing correctly avoiding common mistakes: we fill this gap by presenting here our guidelines for this common digital signal processing activity. Our ten simple recommendations, if taken into account, can help researchers avoid common mistakes and perform better, more robust analyses that, in turn, can lead to more reliable results and outcomes (Fig. 1). Even if we wrote these guidelines for beginners, we believe they should be followed by experts, too, in any scientific project involving EEG data.

TIP 1: BEFORE STARTING, CLEARLY DEFINE YOUR SCIENTIFIC QUESTION AND YOUR EXPERIMENTAL PROTOCOL WITH A CLINICAL NEUROSCIENTIST

The first action to take when you have a research topic in mind is to clearly define a *research question*. In this field, where multidisciplinarity is an essential characteristic of the investigation, you must discuss and converge towards a consistent and well-grounded research question together with the clinicians and the other members of the research team
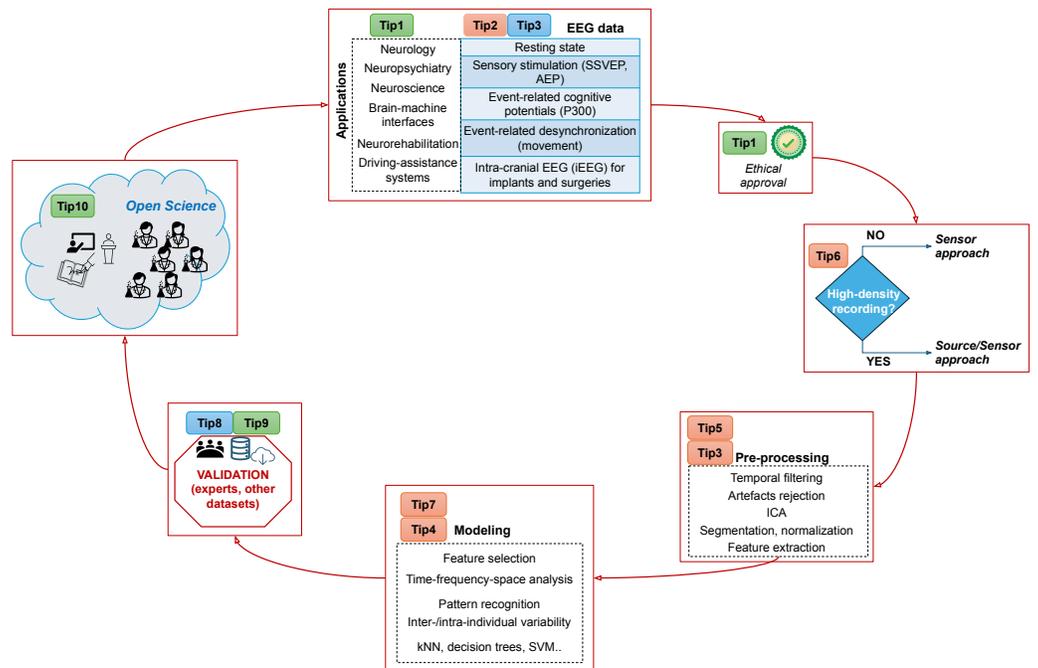
Figure 1   Schematization of the proposed tips as a flowchart. This diagram shows all the steps we suggest to follow in order to facilitate the integration and analysis of electroencephalografic data. We generated the icons through Microsoft Windows PowerPoint.

Full-size ⬜ DOI: 10.7717/peerjcs.2256/fig-1

(it might include neurologists, neuroscientists, physical therapists, cognitive psychologists, data scientists, *etc.*).

The research question must be well-grounded from the current state of the art, innovative, simple to be explained, and clearly stated (Fig. 1).

Once it is formulated, a *research protocol* has to be properly designed. Every in-field experiment has to be supported by a rigorous protocol that defines the methods and the timing of the data collection. The protocol typically includes:

1. A systematic review of the state-of-the-art;
2. The research question and the main objective of the investigation as well as intermediate objectives;
3. The methods to reach the objectives, including the type of study to conduct (for example, a randomized clinical trial), the equipment to collect the data, the statistical analysis and the numerical computation planned to be used, the reporting strategies, and the validation methods;
4. The criteria to include and exclude participants from the data collection and their assignment to the experimental groups (the "control group" typically includes healthy subjects, while the "treatment group" is composed by patients receiving a specific therapeutic intervention);
5. The dissemination and publication modalities at the end of the investigation.

Research protocols ensure the robustness of the investigation (that is, of its outputs), the safety of the participants, and the replicability of the study all over the world.

You can find templates open to be used as a basis to design your own research protocol in *Higgins & Green (2008)* (particularly, its Box 2.2a) and in *The Cochrane Collaboration (2024)*.

If working at the hospital, the research protocol has to be approved before it can be implemented, any research involving humans must be approved by an Institutional Review Board (IRB), and informed consent from the subjects must be collected before participation. Then, it is mandatory that it is presented to the hospital's Ethical Committee, a multidisciplinary committee formed by clinicians, statisticians, experts of medical law and insurance, professionals in health-tech (medical equipment and devices) and possibly other stakeholder representatives. The committee has the role of analyzing all parts of the submitted protocol and guaranteeing its applicability and safety for the involved participants (healthy and patients, as well).

Once you have your experimental protocol approved, you and your team can start recruit patients and collect data (Fig. 1).

TIP 2: KEEP IN MIND THAT PART OF THE SUCCESS OF YOUR SCIENTIFIC PROJECT DEPENDS ON THE QUALITY OF THE EEG RECORDINGS

To ensure proper medical interpretation of the raw EEG data as well as the success of any machine learning model (ML), it is recommended to make the best effort to obtain high-quality EEG recordings (Fig. 1). This can be realized by following these simple guidelines:

1. If available, use high-quality research-grade equipment. This typically provides finer resolution (in time and amplitude) of the EEG data.
2. Carefully and properly set the acquisition parameters. First, the *sampling frequency* should be high enough to allow you to detect the brain behaviour under investigation. If brain reaction is expected a few milliseconds after a given stimulus, the sampling frequency should be higher than 1 kiloHertz (kHz), in order to have at least 1 sample every millisecond. Typical values for the sampling frequency are: 500, 1,000, 2,000 Hz. Pay attention also to have a sufficient number of bit to code the values of your signals: *bit resolution* should be at least 8 bit (check the technical specifications of the EEG device you use). The number of channels is also important: having a sufficiently high number of sensors (also called electrodes) covering the whole scalp surface allows you to monitor the whole brain activity, that is the activity produced by all regions of the brain. On the other hand, if the medical team has strong assumptions on the region where the activity of interest is produced, you can decide to reduce the number of sensors, having a lighter setup with consequent lower time for the preparation of the participant. Finally, there are two special electrodes to place: the ground and the reference one. Our advice is to place the reference electrode in an electrically-neutral location (like the ear lobe or the mastoid). If possible, use the linked-mastoids or

linked-ears function provided by the EEG equipment company, to ensure perfect symmetry in the acquisition. For the ground, you can find indication by the company of which electrode has to be used as ground (check the equipment user manual).

3. Follow the International 10–20 EEG Placement System (or its extension) to correctly place the EEG electrodes on the participant's scalp (*Homan, Herman & Purdy, 1987*). More often, an EEG cap is provided by the EEG producer which integrates all sensors in agreement with the above-mentioned standard. On the other hand, in case you use a portable device, the electrode placement might be different to improve the user comfort. In this case, the reference to the International 10–20 EEG Placement System is useful during the subsequent analysis to compare your results with other studies using standard equipment. Ensure no hair is in between the EEG electrodes and the scalp, or try to scrap them apart. To note, require the participant not to apply any hair foam before the EEG experiment. Bald people might show lower signal because of a thicker scalp. Ensure that cables from the cap to the amplifier are not stretched anywhere.

4. Ensure that the participant sits comfortably on the chair, with the monitor (if any) at about 1 meter apart from them at their eyes' height (so to avoid head movements).

5. To ensure the fairest comparison across different EEG studies, make sure that participants are recorded at the same time during the day (to have consistent phase of the circadian rhythm) and that the room is properly shielded from other electromagnetic interferences.

6. During montage, fill the space between each electrode and the scalp with enough conductive gel to have a sufficiently low impedance that allows a good signal amplitude. You can read the impedance values in the software provided with the equipment. Check the value of every single channel, starting from the ground and the reference ones. Remember: the amplifier takes the voltage difference between each electrode and the reference one, and calibrates itself on the electrical activity acquired at the ground location. This means that if the reference and/or the ground electrodes have problems, all the remaining electrodes will suffer from poor quality. The rule-of-thumb is to keep the impedance value below 20 kiloOhm (but check the equipment user manual, because this threshold might change depending on the electronics of each specific EEG amplifier).

7. Once step 5 is completed, check the raw signals on the visualization panel (equipment software). The signals (from any channel) is expected to have small variations in the range of (max) ±100 microvolts (μV). Figure 2 represents an example of multi-channel EEG recording. If you notice large and slow variations, you should go back to step 6 and check the impedance of the reference, the ground, and the electrodes displaying problems.

8. Before starting the actual data recording, ask the participant to blink their eyes repeatedly, to clench teeth, and, finally, to close eyes and relax. You should verify, respectively, that frontal electrodes show a rhythmic activity corresponding with the blinking rhythm, that temporal and fronto-temporal electrodes show electromyography (EMG) activity overlapping the EEG signal while the participant is clenching, and finally that a small periodic activity (at about 10 Hz, the famous α rhythm) appears at the

parietal and occipital electrodes while the participant is fully relaxing. Figure 3 shows an ideal example of clean EEG recording, with the clear occurrence of some α rhythm events.

9. Finally, instruct the participant to stay as much relaxed as possible during the entire experiment duration. Remember, every movement (for example, head, legs, *etc.*) might induce artefactual components on the EEG recordings. If the experiment takes long, tell them to ask for a break whenever they feel cognitive fatigue or drowsiness.

TIP 3: PROPERLY CLEAN YOUR DATA BEFORE STARTING THE COMPUTATIONAL ANALYSIS: CHECK TIME COURSE AND FREQUENCY SPECTRUM, AND REMOVE NOISE

Before starting any computational step, we strongly suggest you check the raw data. You need to plot the signals both in the time and in the frequency domain. In the time domain, check that the amplitudes do not exceed $\pm 100\,\mu$V (well-established rule of thumb). In the frequency domain, you should recognize the $1/f$ shape of the power spectrum. On the top of the $1/f$ shape, you might recognize some additional narrowband peaks around 5 Hz, 10 Hz and 20 Hz, corresponding to the clinical theta, alpha and beta bands. These frequency bands have a important clinical value. Remember that EEG cannot reliably capture brain frequencies above 90 Hz. Then, if you find large peaks above that frequency, it is much likely that they convey non-brain information and should be eliminated (Fig. 1).

To remove the unwanted components, it is common to use filters. The most general filters you can apply are: a *notch filter* to remove the power line noise (50 or 60 Hz, depending on the country where you work). Then, you need to consult the literature or your colleagues from the clinical side, to decide other filters to apply. To make a few examples, in the case you are studying brain processing in relation to a cognitive task or sensory task, you might heavily filter the signal below 30 Hz (for example, to detect the popular P300 component (*Polich, 1993*)). On the contrary, when dealing with movements and desynchronization of sensorimotor rhythms (*Pfurtscheller & Da Silva, 1999*), we suggest you to apply a band-pass filtering between 3 and 45 Hz (or 55 Hz, in case power line is at 60 Hz).

To increase the spatial specificity of the signal acquired from one location, you can also apply spatial filters such as the common average (CAR) filter, the small Laplacian filter, or the large Laplacian filter. We recommend not to use the CAR, unless you are sure to have clean signals from all electrodes. Spatial filtering is more beneficial in high-density recordings (more than 64 channels), where the electrodes spacing is short.

In the latter cases, you might also decide to exclude the signal from a particularly corrupted channel and substitute it with the interpolated version of its surrounding channels. The latter contain similar information to the rejected channel, given their short distance from it.

Finally, it is suggested to run independent component analysis (ICA) (*Makeig et al., 1995*) to decompose the multi-channel recording into independent components, identify those most likely associated with artefacts (for example, eye blinks), and then recompose the signal by using only clean components. To perform this procedure, the help of domain
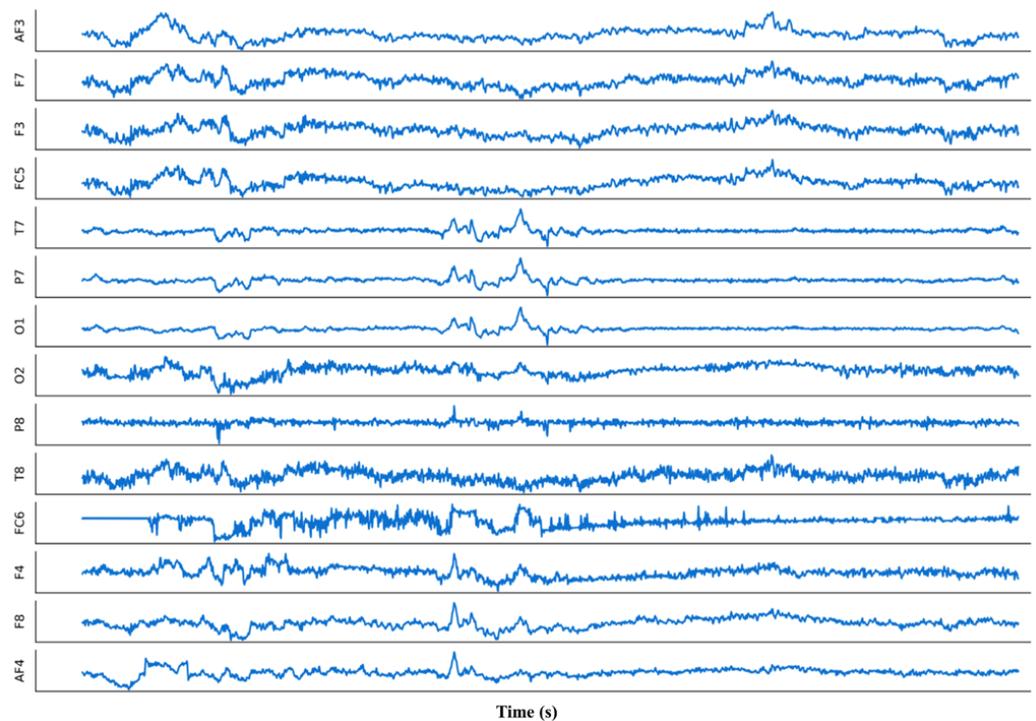
Figure 2 An example of EEG multi-channel recordings. Image released on Wikimedia Commons under the Creative Commons CC0 1.0 Universal Public Domain Dedication (*Krol, 2024*). Each signal represents one EEG recording from one single channel (channel's name on *y* axis) along time (*x* axis).
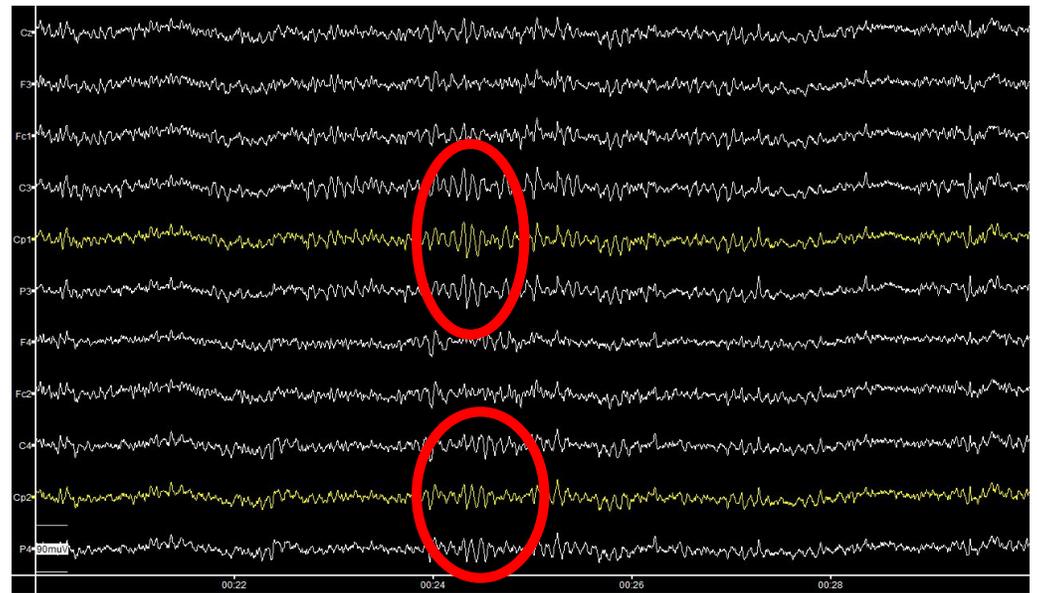
Full-size ⊡ DOI: 10.7717/peerjcs.2256/fig-2

experts or other online resources (for example *Pion-Tonachini, Makeig & Kreutz-Delgado, 2024*; *Pion-Tonachini, Makeig & Kreutz-Delgado, 2017*) is needed to properly select which components to discard. With this method, you can reliably remove the power line noise, blinks and, to some extent, also physiological interferences (for example, muscular or heart activity). Once you obtain the new multi-channel EEG recording after ICA, check the quality of the new dataset.

As a side suggestion, keep track of the signals or the components you discard. Pre-processing in EEG is often not fully replicable, so having notes of what pieces of the signal have been eliminated and why might be important in the case of article review or to reuse the same dataset in the future (Tip 10) (*Karimzadeh & Hoffman, 2018*; *Vardigan, Heus & Thomas, 2008*; *Rasmussen & Blank, 2007*; *Fabris et al., 2022*).

TIP 4: ALWAYS PAY ATTENTION TO TIME, FREQUENCY, AND SPACE DOMAINS, SUBJECT-SPECIFICITY, AND INTER-SUBJECT VARIABILITY OF THE EEG SIGNALS

Assuming you have a clean EEG dataset, it is of key importance to analyze it in the three main domains: time, frequency, and space (Fig. 1). The time domain is essential to identify specific brain responses to stimuli that might be included in the experimental protocol (a

Figure 3 **An example of clean EEG recordings with visible α ryhthm events (in the red circles).**
Full-size ◩ DOI: 10.7717/peerjcs.2256/fig-3

typical example is the P300, explained in Tip 3). The frequency domain is needed especially
when you expect a particular engagement of specific brain waves. Common examples are
the steady-state visual evoked potentials (SSVEP) or the event-related desynchronization
(ERD) paradigms: in SSVEP, a visual stimulus is repeatedly administered to the subjects
letting their brain synchronize the occipital activity to the same frequency of the stimulation
(*Beverina et al., 2003*). In the ERD paradigm (task-oriented), while the subject is moving,
movement-related brain areas decrease their activity in two frequency bands around 10 Hz
and 20 Hz. In these cases, comparing the power spectra before and during stimulation/task
performance will make these frequency responses clearly appearing. The space domain is
fundamental to find the regions where the above time and frequency responses are more
evident. As an example, ERD is mostly found over the sensorimotor cortex (for example,
C3, Cz, C4 electrodes) (*Dugué et al., 2020*), SSVEP over the parietal and occipital cortex
(*Ding, Sperling & Srinivasan, 2006*), while P300 is mostly visible over the centro-parietal
midline (Cz, Pz).

To optimize the analysis efforts towards the most informative direction, it is critical
to have prior knowledge or hypotheses coming from the clinical side or neuroscientific
literature. Depending on the task performed by the individual (motor imagery, sleep,
rest with open eyes, *etc.*), the brain electrical activity can highly vary. Also, keep in mind
that other two factors can change the shape of the EEG recordings: even using the same
experimental protocol (Fig. 1), that is, asking the participants to perform the same task with
the same timing, you might see differences in their EEG signals due to possible pathology
affecting them, and also due to the inherent inter-subject variability that characterizes the
human brain.

In the analysis of EEG data, identifying common patterns across different subjects and investigating subject-specific responses are both important.

TIP 5: PROPERLY PREPARE YOUR DATASET FOR THE COMPUTATIONAL ANALYSIS

To identify those time-frequency-space patterns associated with a task or a pathological brain behaviour, machine learning (ML) models are typically employed. However, it is common to run a few pre-processing steps to prepare the clean dataset for being input to the model (Fig. 1). First, *segmentation* is performed. This means to cut the long-time EEG recording, into smaller chunks that will be then independently treated. Choosing the correct segment duration is a non-trivial step. We would suggest to choose the duration in relation to the experimental protocol: for example, if a movement task is asked to be repeatedly performed every 10 seconds, then a good option is to cut the EEG signals to have 10s-segments. On the other hand, if no external time scheduling is imposed to the subject (in sleep or resting state studies), you might consider 2-seconds segments. The latter is an *empirical* numerical choice representing a popular choice in the EEG community, as it should allow to ensure stationarity (that is, the same statistical properties) across different segments and, thus, proper analysis. Besides, we strongly recommend you to segment *after* filtering, to avoid filter border effects to appear in every segment, attenuating a good amount of initial and final segment samples.

Second, *normalization* can be applied if you need to level out differences in those factors that you don't want to influence your analysis. For example, if you are interested in the general brain response to a specific new stimulus without considering individual differences, then normalizing EEG signals within each subject can be a good choice.

Lastly, for most ML models you need to extract a number of well-established *features*: a rich list of possible features can be found in *Cisotto et al. (2020b)*. Features can be extracted from every single electrode, both in the time and in the frequency domain. Then, *feature selection* based on prior expert knowledge or using an automatic algorithm (*Cisotto et al., 2022*; *Cisotto et al., 2020a*) might be beneficial to let the machine learning model learn more effectively and quickly. Once you have transformed your raw EEG dataset into the features domain, then normalization of each single feature (across EEG samples) might be useful to help the model learn.

If you use deep learning techniques, then feature extraction could be skipped, and you can decide to input raw segments of EEG data into the model: however, this choice depends on the specific architecture you use. There is no golden choice, you need to carefully study previous literature addressing similar analysis.

However, if your objective is to prove the effectiveness of a new analytical method, then use the same input size and preprocessing steps before compare your results with previous literature results.

TIP 6: CAREFULLY CHOOSE BETWEEN THE "SENSORS" OR A "SOURCES" APPROACH

When talking about EEG analysis, you can decide to work with the data directly coming from the electrodes (*sensor analysis*), or to infer the brain activity of deeper brain regions using specific models to transform the recorded EEG data into the *real* brain sources (*source approach*) (Fig. 1). The latter operation lets you find the most probable sources of the electrical activity you have just measured at the surface of the head.

In the following, we explain how to carry source analysis and we discuss the opportunity and the disadvantages to evaluate when deciding to work with sources or sensors. This tip has to be a bit more technical to allow you understand the complexity, but also the advantages, of the source approach.

In some clinical applications, it might be useful to operate the so-called electrical source imaging (*He et al., 2011*; *Michel & He, 2019*), a transformation aimed to map the electrical activity measured at the scalp level (by the EEG sensors) to a set of brain sources that might have produced those superficial measurements. This processing is particularly useful in clinical applications like the identification of the sources of epileptic foci (*Bénar et al., 2006*), of the α rhythm activity during a resting-state period (*Cuspineda et al., 2009*), of the sleep waves (spindles) (*Del Felice et al., 2014*), and in examinations where other imaging methods (MRI, fMRI, CT, *etc.*) are available and can be used in co-registration with EEG (*Bénar et al., 2006*).

However, implementing this transformation is a non-trivial task. There is an entire subfield of computational neuroscience addressing the challenges related to this transformation: in fact, a number of hypotheses has to be assumed. First, you need to have a model of the head layers (*Plummer, Harvey & Cook, 2008*). Second, prior knowledge should let you select the number and the distribution of the sources expected to explain the superficial activity (*He et al., 2011*; *Michel & He, 2019*). Third, a variety of possible models to actually realize the mapping are available: to name a few, the weighted minimum norm (WMN) (*de Peralta Menendez et al., 2004*) , the popular low-resolution electromagnetic tomography algorithm (LORETA) (*Pascual-Marqui, Michel & Lehmann, 1994*), the multiple signal classification (MUSIC) (*Mosher & Leahy, 1999*). The interested reader might refer to *Kaur et al. (2022)*, *Asadzadeh et al. (2020)* for recent reviews on this topic.

Nevertheless, if you have less than 60 electrodes, it is *not* recommended to apply this transformation: motivated by the fundamental theorems of linear algebra and mathematical modeling, you need to have a sufficiently high number of sensors to reverse the modelling problem and obtain the sources of the observed activity. Empirically, 60 sensors is a well-accepted minimum value to have a reliable source localization within the brain from scalp EEG recordings (*Michel et al., 2004*).

On the contrary, if you have more than 60 electrodes and you opt for the source approach, you can find a number of software tools and libraries to help you apply this transformation, implementing different modeling methods. To name a few: CARTOOL

(_Michel & Brunet, 2019_), EEGLAB (_Delorme et al., 2011_), and FieldTrip (_Oostenveld et al., 2011_).

Finally, it is worth to mention that models based on deep learning have been also recently proposed, showing promising results (_Liang et al., 2023_; _Cui et al., 2019_; _Pantazis & Adler, 2021_).

TIP 7: FOR THE COMPUTATIONAL ANALYSIS, START WITH THE SIMPLEST METHODS, AND USE MORE COMPLEX METHODS ONLY IF IT IS NECESSARY

When the dataset you would like to analyze is finally pre-processed and ready to be used, you come to an important decision which regards all the computational researchers in any field: which computational method should I start with? We have a straight piece of advice for this choice: start with the simplest method available. Period.

A simple technique (for example, linear regression), in fact, gives you the possibility to understand how its statistical model works, to comprehend its functioning, and to interpret its results and why they were generated that way. Complex methods, instead, are difficult to implement and hard to interpret, and should be utilized only if necessary. If you obtain sufficient results with a simple method, stick with them; otherwise, of course feel free to use more complex methods. But do not start with complicated methods: start with simple algorithms.

If you can achieve sufficiently relevant results with basic statistics tools (for example, mean, standard deviation, median, minimum and maximum) or with traditional biostatistics tests (such as Mann–Whitney U test (_MacFarland et al., 2016_), Student's t test (_Mishra et al., 2019_), Kruskal–Wallis test (_Ostertagova, Ostertag & Kováč, 2014_), chi-squared test (_McHugh, 2013_, _etc._), then go with them. For probability values, we suggest to use the $p < 0.005$ significance threshold, as indicated by _Benjamin et al. (2018)_.

If your analysis involves supervised machine learning, we suggest you to start with Decision Trees, and not to start with complex or large deep learning models (_Chicco, 2017_; _Zancanaro et al., 2021_; _Cisotto et al., 2023_). If your analysis includes an unsupervised machine learning phase, we advise you to start with k-means clustering and not with spectral clustering.

Leonardo Da Vinci used to say: "Simplicity is the ultimate sophistication". It is true also for EEG signal processing.

TIP 8: LOOK FOR A VALIDATION COHORT DATASET ONLINE AND REPEAT YOUR ANALYSIS ON IT

Once a researcher has performed a computational analysis on an electroencephalographic dataset and has discovered something relevant regarding the subjects involved, they might tend to think that their job is done and they can focus on paper writing. Actually, scientific findings obtained on a single dataset, although possibly useful and interesting, result being clearly dataset-specific and lack generalizability.

To make a scientific study sounder and more robust, a confirmatory analysis on a validation cohort dataset is necessary (Fig. 1). Finding a dataset that is compatible and similar to the primary cohort dataset employed in a project might be difficult, but fortunately there are multiple online resources in the Internet for this purpose.

So, here is our tip: when you finished your computational analysis on the primary dataset, look for a validation dataset online on resources such as the following ones:

- Google Dataset Search (_Google, 2024_)
- re3data.org (_re3data, 2024_)
- PhysioNet (_PhysioNet, 2024_)
- OpenNeuro.org (_Markiewicz et al., 2021_)
- EEGLAB Wiki (_Delorme et al., 2019_; _Martínez-Cancino et al., 2021_)
- Zenodo (_Zenodo, 2024_)
- Kaggle (_Kaggle, 2024a_)
- University of California Irvine Machine Learning Repository (_University of California Irvine, 2024_)
- Figshare (_Figshare, 2024_)
- Brain-Computer Interface (BCI) Competition IV (_Brain-Computer Interface (BCI) Competition IV Organizers, 2024_)

You might also find a suitable dataset among the articles published in the _Scientific Data_ journal (_Babayan et al., 2019_; _Cao et al., 2019_; _Shin et al., 2018_; _Luciw, Jarocka & Edin, 2014_; _Hollenstein et al., 2018_; _Won et al., 2022_; _Nieto et al., 2022_; _Pernet et al., 2019_; _Grootswagers et al., 2022_; _Mikulan et al., 2020_; _Ma et al., 2022_; _Valdes-Sosa et al., 2021_; _Stevenson et al., 2019_). On Google Scholar, this search can be performed by using the following query: `EEG source:"Scientific Data"`

The possibility to repeat your computational analysis on an alternative, independent dataset, and perhaps to find similar outcomes to the ones you found on the primary dataset would make your scientific study more reliable and relevant for the scientific community.

TIP 9: ASK A CLINICAL NEUROSCIENTIST TO ASSESS YOUR RESULTS

As we mentioned early (Tip 1), a sound medical project always starts with a clear, feasible scientific question designed by clinical neuroscientists. A scientific question is well posed if, once solved, its solution improves the knowledge of neurological research, and consequently can influence medical knowledge and neurological therapies.

So here is our piece of advice. When you generate results on the EEG data you analyzed, we suggest you to knock again on the door of the clinical neuroscientists that you met at the beginning of your journey, and ask them to assess and review your results. They will evaluate your outcomes and findings, providing precious feedback on what to do next, on what to repeat, and on how to write the scientific paper about your electroencephalography project (Fig. 1). A sound medical research project starts in the hospital and ends in the hospital (Fig. 4).
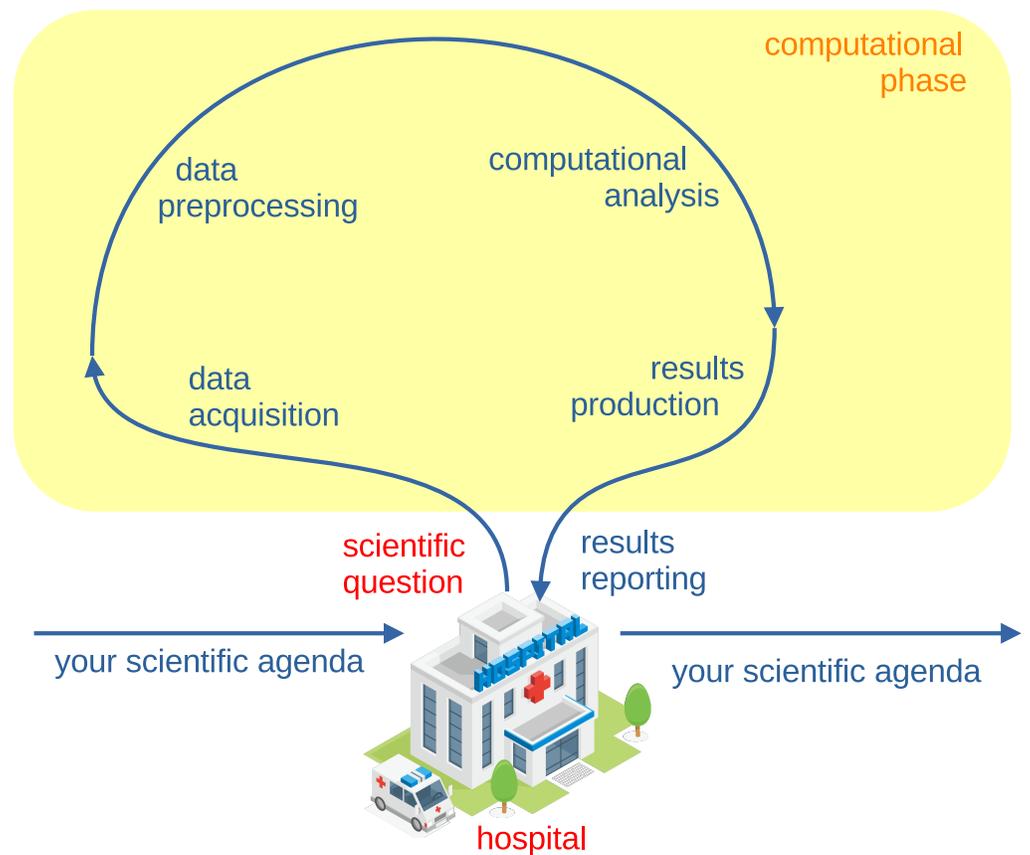
Figure 4 Schematic representation of a sound biomedical research project cycle. A relevant scientific question is born in a hospital from medical doctors or clinical neuroscientists who identify a current lack or problem in biomedical research or in clinical practice. A scientific question invented by the biomedical engineers or by health informatics researchers, without the help of clinical neuroscientists, is probably badly posed or misleading. Scientific researchers take the medical question in custody from the hospital, then collect the data and pre-process them for the computational analysis. They use computational methods to infer new knowledge on these data, and eventually deliver their scientific results back to the clinical neuroscientists of the hospital where the scientific question was born at the beginning. The clinical neuroscientists review the results, provide feedback, comments, prompts, and insights, and possibly change their strategy on treatments and therapies for patients. The hospital building image was released under the Creative Commons 4.0 BY-NC DEED license on PngAll.com.

Full-size 🖾 DOI: 10.7717/peerjcs.2256/fig-4

In your results delivery document, we advise you to write your report in a simple scientific language that can be understood by clinical neuroscientists, too: avoid jargon, avoid words which have different meanings in different fields, and have a lots of patience when working with medical doctors (*Chicco & Jurman, 2023*). If possible, ask your clinical colleagues to meet and discuss your results in person.

In this case, put some effort to carefully prepare nice and intuitive figures and plots for the clinicians. Remember to put clear labels, units, and titles on every figure and to describe them in simple words (this can be a useful preparation to have your nice figures for the scientific manuscript you want to write, as well). If possible, arrange the figures in

a similar way as done in the papers you use as main references for your study. It will help your colleagues get the meaning and impact of your results.

Our personal experiences showed that, when medical doctors were involved both for the study design and for the results' assessment, the final results turned out to be sounder and more reliable (*Cerono, Melaiu & Chicco, 2023*; *Chicco et al., 2023*).

TIP 10: FOLLOW THE PRINCIPLES OF OPEN SCIENTIFIC RESEARCH AND DOCUMENT EVERYTHING

In the history of humanity, scientific progress has happened only when information was shared openly between people. Scientific research, as well, works better when it is open.

We can describe open scientific research through five pillars: usage of open source software code; open software code release; open data release; open access publication; open and complete documentation (for a more detailed, and general, approach to open and reproducible science, also check (*Cabitza & Campagner, 2021*)).

1. Use only open source programming languages and platforms, such as Python and R. Using an open programming language will allow the sharing of software scripts between collaborators, without any issues regarding licenses. Moreover, it will allow the reproducibility of the experiments, allowing anyone in the world with a computer to install the desired software packages and re-run the tests you did. Python is the most used programming language in the world, according to the PYPL index (*PYPL, 2024*), to the TIOBE (*TIOBE, 2024*) and to the Kaggle survey (*Kaggle, 2024b*). Moreover, the Python package index *pypi* contains around fifty software libraries for electroencephalography signal processing in stable or mature development status (*pypi, 2024*). Among them, MNE-Python (*Gramfort et al., 2013*; *Gramfort et al., 2014*), Py-EEG (*Bao, Liu & Zhang, 2011*), and NeuroKit2 (*Makowski et al., 2021*) are the widely used. R is another open source programming language and platform commonly utilized in health informatics and bioinformatics. A few R software packages for EEG signal processing exist in the Comprehensive R Archive Network (CRAN): eegkit (*EEGkit, 2024*), MedicalImaging (*MedicalImaging, 2024*), and eegUtils (*EEGUtils, 2024*).

2. Release your software code openly online. To enforce the reproducibility of your study, you can also consider publishing your software code openly on GitHub, GitLab, or SourceForge. Moreover, if your software code can be used as a package, you can consider submitting it to a central repository of software libraries, such as *pypi* for Python, CRAN for R, Julia Packages for Julia or Crates.io for Rust, for example. Additionally, anyone will be able to assess your software scripts and understand if any mistakes were made, making your study methodology more transparent and robust (*Barnes, 2010*).

3. Release your dataset online, if you are authorized. If the subjects of the EEG experiments gave consent, and the bioethical committee of your institution agreed, go on and publish your EEG data openly online on public data repositories such as PhysioNet (*PhysioNet, 2024*), OpenNeuro.org (*Markiewicz et al., 2021*), EEGLAB Wiki (*Delorme et al., 2019*; *Martínez-Cancino et al., 2021*), Zenodo (*Zenodo, 2024*), Kaggle (*Kaggle,*

2024a), University of California Irvine Machine Learning Repository (*University of California Irvine, 2024*), or Figshare (*Figshare, 2024*). Releasing your dataset openly online would make it available for secondary analyses to anyone in the world, who might discover something new regarding clinical neurosciences. Moreover, it would increase the impact of your study. Remember to anonymize the records prior release.

4. Publish your scientific article in an open access journal. If you have a say in what journal you and your team can submit the article about your EEG study, we strongly suggest to pick an open access one. By publishing your article open access, in fact, you would make it available for free to anyone in the world, including to people in the least developed countries, students of high schools, and taxpayers. Additionally, open access articles obtain more citations (*Tang, Bever & Yu, 2017*). A list of health informatics open access journals where to submit an article on electroencephalographic data can be found on ScimagoJR (*Scimago Journal Ranking, 2024*). While carefully selecting a journal where to submit a scientific article, it is pivotal to avoid predatory journals (*Cobey et al., 2018*).

5. Write open and complete documentation. Documentation is the backbone of scientific research (*Karimzadeh & Hoffman, 2018*; *Witzman et al., 2020*; *Aghajani et al., 2020*): if you write it clearly and release it online openly without restrictions, anyone who will reuse your software scripts will take advantage of it. On the contrary, the absence of well-written documentation would make it impossible for other people to use your scripts and programs. Examples of well-written tutorials for EEG software can be found on the already-mentioned `MNE-Python` package website (*MNE, 2024*). Moreover, keeping a detailed laboratory notebook is essential in this context (*Schnell, 2015*).

These practice for open scientific research, if taken into practice, will not only make your studies more robust and reliable, but will also boost your career in several ways, giving you more visibility (Fig. 1).

CONCLUSIONS

Understanding how the human brain works has always been a fascinating and difficult task for scientific researchers, and electroencephalography (EEG) has been a useful tool for this scope. Even if collecting EEG signal is a non-trivial task, analyzing electroencephalographic data remains one of the most informative way for investigating the dynamic functioning of human brain.

EEG is, indeed, employed in many different clinical neuroscience applications, from diagnosis of epilepsy to motor rehabilitation and EEG-driven wheelchairs. A similar variety is also present in the equipment that can record EEG: some are specifically dedicated to high-level research (with lots of electrodes, implying long preparation times but also high-quality data), others target portability and usability (being embedded with a few electrodes allocated in a more fancy and usable support, but providing much lower data quality). Thus, to face any type of data possibly acquired, a solid acquisition protocol and pre-processing pipeline are needed. Therefore, we presented here our ten quick tips for EEG signal acquisition and processing, by proposing some easy guidelines to avoid common

mistakes in clinical neuroscientific research projects involving electroencephalographic data. Some of these quick recommendations derive from our direct experience, while others come from errors we noticed in other people's studies or in published articles on EEG acquisition and/or analysis. We presented our ten pieces of advice in a simple way, so that they could be understood by anyone, including beginners and unexperienced researchers. Although we originally designed these tips for novices, we believe our ten quick tips should be followed and kept in mind by experienced researchers as well.

List of abbreviations

AEP	auditory evoked potential
BCI	Brain-Computer Interface
CAR	common average
CC	Creative Commons
CRAN	Comprehensive R Archive Network
CT	computed tomography
ECG	electrocardiogram
EEG	electroencephalography
EMG	electromyograph
ERD	event-related desynchronization
fMRI	functional magnetic resonance imaging
Hz	Hertz
ICA	independent component analysis
iEEG	intracranial electroencephalography
IRB	Institutional Review Board
LORETA	low-resolution electromagnetic tomography algorithm
ML	machine learning
MUSIC	multiple signal classification
NIRS	near-infrared spectroscopy
p-**value**	probability value
SSVEP	steady-state visual evoked potentials
WMN	weighted minimum-norm least squares

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests

Both Davide Chicco and Giulia Cisotto are academic editors at PeerJ Computer Science.

Author Contributions

- Giulia Cisotto conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, recommendations design and search in the literature, and approved the final draft.
- Davide Chicco conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the article, recommendations design and search in the literature, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

There is no raw data or code in this study because it is a literature review.

REFERENCES

Aghajani E, Nagy C, Linares-Vásquez M, Moreno L, Bavota G, Lanza M, Shepherd DC. 2020. Software documentation: the practitioners' perspective. In: *Proceedings of ICSE-SEET '20 –the 42nd ACM/IEEE international conference on software engineering.* Piscataway: IEEE, 590–601.

Ang KK, Guan C. 2016. EEG-based strategies to detect motor imagery for control and rehabilitation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **25(4)**:392–401 DOI 10.1109/TNSRE.2016.2646763.

Asadzadeh S, Rezaii TY, Beheshti S, Delpak A, Meshgini S. 2020. A systematic review of EEG source localization techniques and their applications on diagnosis of brain abnormalities. *Journal of Neuroscience Methods* **339**:108740 DOI 10.1016/j.jneumeth.2020.108740.

Babayan A, Erbey M, Kumral D, Reinelt JD, Reiter AMF, Röbbig J, Schaare HL, Uhlig M, Anwander A, Bazin P-L, Horstmann A, Lampe L, Nikulin VV, Okon-Singer H, Preusser S, Pampel A, Rohr CS, Sacher J, Thöne-Otto A, Trapp S, Nierhaus

T, Altmann D, Arelin K, Blöchl M, Bongartz E, Breig P, Cesnaite E, Chen S, Cozatl R, Czerwonatis S, Dambrauskaite G, Dreyer M, Enders J, Engelhardt M, Fischer MM, Forschack N, Golchert J, Golz L, Guran CA, Hedrich S, Hentschel N, Hoffmann DI, Huntenburg JM, Jost R, Kosatschek A, Kunzendorf S, Lammers H, Lauckner ME, Mahjoory K, Kanaan AS, Mendes N, Menger R, Morino E, Näthe K, Neubauer J, Noyan H, Oligschläger S, Panczyszyn-Trzewik P, Poehlchen D, Putzke N, Roski S, Schaller M-C, Schieferbein A, Schlaak B, Schmidt R, Gorgolewski KJ, Schmidt HM, Schrimpf A, Stasch S, Voss M, Wiedemann A, Margulies DS, Gaebler M, Villringer A. 2019. A mind-brain-body dataset of MRI, EEG, cognition, emotion, and peripheral physiology in young and old adults. *Scientific Data* **6**:66 DOI 10.1038/sdata.2018.308.

Bao FS, Liu X, Zhang C. 2011. PyEEG: an open source Python module for EEG/MEG feature extraction. *Computational Intelligence and Neuroscience* **2011**:406391 DOI 10.1155/2011/406391.

Barnes N. 2010. Publish your computer code: it is good enough. *Nature* **467**(**7317**):753–753 DOI 10.1038/467753a.

Bénar C-G, Grova C, Kobayashi E, Bagshaw AP, Aghakhani Y, Dubeau F, Gotman J. 2006. EEG–fMRI of epileptic spikes: concordance with EEG source localization and intracranial EEG. *NeuroImage* **30**(**4**):1161–1170 DOI 10.1016/j.neuroimage.2005.11.008.

Beniczky S, Schomer DL. 2020. Electroencephalography: basic biophysical and technological aspects important for clinical applications. *Epileptic Disorders* **22**(**6**):697–715 DOI 10.1684/epd.2020.1217.

Benjamin DJ, Berger JO, Johannesson M, Nosek BA, Wagenmakers E-J, Berk R, Bollen KA, Brembs B, Brown L, Camerer C, Cesarini D, Chambers CD, Clyde M, Cook TD, Boeck PD, Dienes Z, Dreber A, Easwaran K, Efferson C, Fehr E, Fidler F, Field AP, Forster M, George EI, Gonzalez R, Goodman S, Green E, Green DP, Greenwald AG, Hadfield JD, Hedges LV, Held L, Ho TH, Hoijtink H, Hruschka DJ, Imai K, Imbens G, Ioannidis JPA, Jeon M, Jones JH, Kirchler M, Laibson D, List J, Little R, Lupia A, Machery E, Maxwell SE, McCarthy M, Moore DA, Morgan SL, Munafó M., Nakagawa S, Nyhan B, Parker TH, Pericchi L, Perugini M, Rouder J, Rousseau J, Savalei V, Schönbrodt FD, Sellke T, Sinclair B, Tingley D, Van Zandt T, Vazire S, Watts DJ, Winship C, Wolpert RL, Xie Y, Young C, Zinman J, Johnson VE. 2018. Redefine statistical significance. *Nature Human Behaviour* **2**(**1**):6–10.

Berger H. 1929. On the EEG in humans. *Archiv für Psychiatrie und Nervenkrankheiten* **87**:527–570 DOI 10.1007/BF01797193.

Beverina F, Palmas G, Silvoni S, Piccione F, Giove S. 2003. User adaptive BCIs: SSVEP and P300 based interfaces. *Psychology Journal* **1**(**4**):331–354.

Biasiucci A, Franceschiello B, Murray MM. 2019. Electroencephalography. *Current Biology* **29**(**3**):R80–R85 DOI 10.1016/j.cub.2018.11.052.

Brain-Computer Interface (BCI) Competition IV Organizers. 2024. BCI Competition IV. *Available at https://www.bbci.de/competition/iv/* (accessed on 25 June 2024).

Cabitza F, Campagner A. 2021. The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies. *International Journal of Medical Informatics* **153**:104510 DOI 10.1016/j.ijmedinf.2021.104510.

Cao Z, Chuang C-H, King J-K, Lin C-T. 2019. Multi-channel EEG recordings during a sustained-attention driving task. *Scientific Data* **6**:19 DOI 10.1038/s41597-019-0027-4.

Cerono G, Melaiu O, Chicco D. 2023. Clinical feature ranking based on ensemble machine learning reveals top survival factors for glioblastoma multiforme. *Journal of Healthcare Informatics Research* **8(1)**:1–18 DOI 10.1007/s41666-023-00138-1.

Chicco D. 2017. Ten quick tips for machine learning in computational biology. *BioData Mining* **10(1)**:1–17 DOI 10.1186/s13040-016-0121-5.

Chicco D, Haupt R, Garaventa A, Uva P, Luksch R, Cangelosi D. 2023. Computational intelligence analysis of high-risk neuroblastoma patient health records reveals time to maximum response as one of the most relevant factors for outcome prediction. *European Journal of Cancer* **193**:113291 DOI 10.1016/j.ejca.2023.113291.

Chicco D, Jurman G. 2023. Ten simple rules for providing bioinformatics support within a hospital. *BioData Mining* **16(1)**:1–12 DOI 10.1186/s13040-022-00318-6.

Cisotto G, Capuzzo M, Guglielmi AV, Zanella A. 2020a. Feature selection for gesture recognition in Internet-of-Things for healthcare. In: *Proceedings of ICC 2020 –the 2020 IEEE international conference on communications*. Piscataway: IEEE, 1–6.

Cisotto G, Capuzzo M, Guglielmi AV, Zanella A. 2022. Feature stability and setup minimization for EEG-EMG-enabled monitoring systems. *EURASIP Journal on Advances in Signal Processing* **2022(1)**:103 DOI 10.1186/s13634-022-00939-3.

Cisotto G, Zancanaro A, Zoppis IF, Manzoni SL. 2023. hvEEGNet: exploiting hierarchical VAEs on EEG data for neuroscience applications. ArXiv arXiv:2312.00799.

Cisotto G, Zanga A, Chlebus J, Zoppis I, Manzoni S, Markowska-Kaczmar U. 2020b. Comparison of attention-based deep learning models for EEG classification. ArXiv arXiv:2012.01074.

Cobey KD, Lalu MM, Skidmore B, Ahmadzai N, Grudniewicz A, Moher D. 2018. What is a predatory journal? A scoping review. *F1000Research* **7**:1001 DOI 10.12688/f1000research.15256.2.

Cui S, Duan L, Gong B, Qiao Y, Xu F, Chen J, Wang C. 2019. EEG source localization using spatio-temporal neural network. *China Communications* **16(7)**:131–143 DOI 10.23919/JCC.2019.07.011.

Cuspineda E, Machado C, Virues T, Martínez-Montes E, Ojeda A, Valdés P, Bosch J, Valdes L. 2009. Source analysis of alpha rhythm reactivity using LORETA imaging with 64-channel EEG and individual MRI. *Clinical EEG and Neuroscience* **40(3)**:150–156 DOI 10.1177/155005940904000306.

Del Felice A, Arcaro C, Storti SF, Fiaschi A, Manganotti P. 2014. Electrical source imaging of sleep spindles. *Clinical EEG and Neuroscience* **45(3)**:184–192 DOI 10.1177/1550059413497716.

Delorme A, Majumdar A, Sivagnanam S, Martinez-Cancino R, Yoshimoto K, Makeig S. 2019. The open EEGLAB portal. In: *Proceedings of NER 2019 –the 9th international IEEE/EMBS conference on neural engineering*. Piscataway: IEEE, 1142–1145.

Delorme A, Mullen T, Kothe C, Acar ZA, Bigdely-Shamlo N, Vankov A, Makeig S. 2011. EEGLAB, SIFT, NFT, BCILAB, and ERICA: new tools for advanced EEG processing. *Computational Intelligence and Neuroscience* **2011**:130714 DOI 10.1155/2011/130714.

de Peralta Menendez RG, Murray MM, Michel CM, Martuzzi R, Andino SLG. 2004. Electrical neuroimaging based on biophysical constraints. *NeuroImage* **21(2)**:527–539 DOI 10.1016/j.neuroimage.2003.09.051.

Ding J, Sperling G, Srinivasan R. 2006. Attentional modulation of SSVEP power depends on the network tagged by the flicker frequency. *Cerebral Cortex* **16(7)**:1016–1029 DOI 10.1093/cercor/bhj044.

Dugué L, Merriam EP, Heeger DJ, Carrasco M. 2020. Differential impact of endogenous and exogenous attention on activity in human visual cortex. *Scientific Reports* **10**:21274.

EEGkit. 2024. CRAN EEGkit: toolkit for Electroencephalography Data. *Available at https://doi.org/10.32614/cran.package.eegkit* (accessed on 25 June 2024).

EEGUtils. 2024. GitHub EEGUtils. *Available at https://craddm.github.io/eegUtils/* (accessed on 25 June 2024).

Ein Shoka AA, Dessouky MM, El-Sayed A, Hemdan EE-D. 2023. EEG seizure detection: concepts, techniques, challenges, and future trends. *Multimedia Tools and Applications* **82(27)**:42021–42051.

Fabris A, Messina S, Silvello G, Susto GA. 2022. Tackling documentation debt: a survey on algorithmic fairness datasets. In: *Proceedings of EAAMO '22 –the 2nd ACM conference on equity and access in algorithms, mechanisms, and optimization*. New York: ACM, 1–13.

Figshare. 2024. Store, share, discover research. *Available at https://www.figshare.com* (accessed on 25 June 2024).

Google. 2024. Dataset search. *Available at https://datasetsearch.research.google.com* (accessed on 25 June 2024).

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Goj R, Jas M, Brooks T, Parkkonen L, Hamalainen M. 2013. MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience* **7**:267 DOI 10.3389/fnins.2013.00267.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS. 2014. MNE software for processing MEG and EEG data. *NeuroImage* **86**:446–460 DOI 10.1016/j.neuroimage.2013.10.027.

Grootswagers T, Zhou I, Robinson AK, Hebart MN, Carlson TA. 2022. Human EEG recordings for 1,854 concepts presented in rapid serial visual presentation streams. *Scientific Data* **9**:3 DOI 10.1038/s41597-021-01102-7.

He B, Yang L, Wilke C, Yuan H. 2011. Electrophysiological imaging of brain activity and connectivity—challenges and opportunities. *IEEE Transactions on Biomedical Engineering* **58(7)**:1918–1931 DOI 10.1109/TBME.2011.2139210.

Higgins JP, Green S. 2008. *Cochrane Handbook for Systematic Reviews of Interventions.* Hoboken, New Jersey: Wiley Online Library.

Hollenstein N, Rotsztejn J, Troendle M, Pedroni A, Zhang C, Langer N. 2018. ZuCo, a simultaneous EEG and eye-tracking resource for natural sentence reading. *Scientific Data* **5**:180291 DOI 10.1038/sdata.2018.291.

Homan RW, Herman J, Purdy P. 1987. Cerebral location of international 10–20 system electrode placement. *Electroencephalography and Clinical Neurophysiology* **66(4)**:376–382 DOI 10.1016/0013-4694(87)90206-9.

Kaggle. 2024a. Kaggle datasets–Explore, analyze, and share quality data. *Available at* *https://www.kaggle.com/datasets* (accessed on 25 June 2024).

Kaggle. 2024b. State of data science and machine learning 2022. *Available at* *https://www. kaggle.com/kaggle-survey-2022* (accessed on 25 June 2024).

Karimzadeh M, Hoffman MM. 2018. Top considerations for creating bioinformatics software documentation. *Briefings in Bioinformatics* **19(4)**:693–699 DOI 10.1093/bib/bbw134.

Kaur C, Singh P, Bisht A, Joshi G, Agrawal S. 2022. Recent developments in spatio-temporal EEG source reconstruction techniques. *Wireless Personal Communications* **122(2)**:1531–1558 DOI 10.1007/s11277-021-08960-9.

Khademi Z, Ebrahimi F, Kordy HM. 2023. A review of critical challenges in MI-BCI: from conventional to deep learning methods. *Journal of Neuroscience Methods* **383**:109736 DOI 10.1016/j.jneumeth.2022.109736.

Krol LR. 2024. EEG time series 62 channels. *Available at* *https://commons.wikimedia.org/ wiki/File:EEG_time_series_62_channels.svg* (accessed on 25 June 2024).

Liang J, Yu ZL, Gu Z, Li Y. 2023. Electromagnetic source imaging with a combination of sparse Bayesian learning and deep neural network. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* **31**:2338–2348 DOI 10.1109/TNSRE.2023.3251420.

Luciw MD, Jarocka E, Edin BB. 2014. Multi-channel EEG recordings during 3,936 grasp and lift trials with varying weight and friction. *Scientific Data* **1**:10047 DOI 10.1038/sdata.2014.47.

Ma J, Yang B, Qiu W, Li Y, Gao S, Xia X. 2022. A large EEG dataset for studying cross-session variability in motor imagery brain-computer interface. *Scientific Data* **9(1)** DOI 10.1038/s41597-022-01647-1.

MacFarland TW, Yates JM, MacFarland TW, Yates JM. 2016. Mann–Whitney *U* test. In: *Introduction to Nonparametric Statistics for the Biological Sciences using R.* Cham: Springer, 103–132.

Makeig S, Bell A, Jung T-P, Sejnowski TJ. 1995. Independent component analysis of electroencephalographic data. *Advances in neural information processing systems 8 (NIPS 1995).*

Makowski D, Pham T, Lau ZJ, Brammer JC, Lespinasse F, Pham H, Schölzel C, Chen SA. 2021. NeuroKit2: a Python toolbox for neurophysiological signal processing. *Behavior Research Methods* **53**:1689–1696 DOI 10.3758/s13428-020-01516-y.

Markiewicz CJ, Gorgolewski KJ, Feingold F, Blair R, Halchenko YO, Miller E, Hardcastle N, Wexler J, Esteban O, Goncavles M, Jwa A, Poldrack R. 2021. The OpenNeuro resource for sharing of neuroscience data. *eLife* **10**:e71774 DOI 10.7554/eLife.71774.

Martínez-Cancino R, Delorme A, Truong D, Artoni F, Kreutz-Delgado K, Sivagnanam S, Yoshimoto K, Majumdar A, Makeig S. 2021. The open EEGLAB portal interface: high-performance computing with EEGLAB. *NeuroImage* **224**:116778 DOI 10.1016/j.neuroimage.2020.116778.

McHugh ML. 2013. The chi-square test of independence. *Biochemia Medica* **23(2)**:143–149.

MedicalImaging. 2024. CRAN task view: medical image analysis. *Available at https://cran.r-project.org/web/views/MedicalImaging.html* (accessed on 25 June 2024).

Michel CM, Brunet D. 2019. EEG source imaging: a practical review of the analysis steps. *Frontiers in Neurology* **10**:325 DOI 10.3389/fneur.2019.00325.

Michel CM, He B. 2019. EEG source localization. *Handbook of Clinical Neurology* **160**:85–101 DOI 10.1016/B978-0-444-64032-1.00006-0.

Michel CM, Murray MM, Lantz G, Gonzalez S, Spinelli L, De Peralta RG. 2004. EEG source imaging. *Clinical Neurophysiology* **115(10)**:2195–2222 DOI 10.1016/j.clinph.2004.06.001.

Mikulan E, Russo S, Parmigiani S, Sarasso S, Zauli FM, Rubino A, Avanzini P, Cattani A, Sorrentino A, Gibbs S, Cardinale F, Sartori I, Nobili L, Massimini M, Pigorini A. 2020. Simultaneous human intracerebral stimulation and HD-EEG, ground-truth for source localization methods. *Scientific Data* **7**:127 DOI 10.1038/s41597-020-0467-x.

Mishra P, Singh U, Pandey CM, Mishra P, Pandey G. 2019. Application of Student's t-test, analysis of variance, and covariance. *Annals of Cardiac Anaesthesia* **22(4)**:407 DOI 10.4103/aca.ACA_94_19.

MNE. 2024. Tutorials. *Available at https://mne.tools/stable/auto_tutorials/index.html* (accessed on 25 June 2024).

Mosher JC, Leahy RM. 1999. Source localization using recursively applied and projected (RAP) MUSIC. *IEEE Transactions on Signal Processing* **47(2)**:332–340 DOI 10.1109/78.740118.

Nieto N, Peterson V, Rufiner HL, Kamienkowski JE, Spies R. 2022. Thinking out loud, an open-access EEG-based BCI dataset for inner speech recognition. *Scientific Data* **9(1)** DOI 10.1038/s41597-022-01147-2.

Oostenveld R, Fries P, Maris E, Schoffelen J-M. 2011. FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience* **2011**:156869 DOI 10.1155/2011/156869.

Ostertagova E, Ostertag O, Kováč J. 2014. Methodology and application of the Kruskal-Wallis test. *Applied Mechanics and Materials* **611**:115–120 DOI 10.4028/www.scientific.net/AMM.611.115.

Pantazis D, Adler A. 2021. MEG source localization via deep learning. *Sensors* **21(13)**:4278 DOI 10.3390/s21134278.

Pascual-Marqui RD, Michel CM, Lehmann D. 1994. Low resolution electromagnetic tomography: a new method for localizing electrical activity in the brain. *International Journal of Psychophysiology* **18(1)**:49–65 DOI 10.1016/0167-8760(84)90014-X.

Pernet CR, Appelhoff S, Gorgolewski KJ, Flandin G, Phillips C, Delorme A, Oostenveld R. 2019. EEG-BIDS, an extension to the brain imaging data structure for electroencephalography. *Scientific Data* **6(1)** DOI 10.1038/s41597-019-0104-8.

Pfurtscheller G, Da Silva FL. 1999. Event-related EEG/MEG synchronization and desynchronization: basic principles. *Clinical Neurophysiology* **110(11)**:1842–1857 DOI 10.1016/S1388-2457(99)00141-8.

PhysioNet. 2024. The research resource for complex physiologic signals. *Available at https://www.physionet.org/* (accessed on 25 June 2024).

Pion-Tonachini L, Makeig S, Kreutz-Delgado K. 2017. Crowd labeling latent Dirichlet allocation. *Knowledge and Information Systems* **53**:749–765 DOI 10.1007/s10115-017-1053-1.

Pion-Tonachini L, Makeig S, Kreutz-Delgado K. 2024. ICLabel tutorial: EEG independent component labeling. *Available at https://labeling.ucsd.edu/tutorial/practice* (accessed on 22 March 2024).

Plummer C, Harvey AS, Cook M. 2008. EEG source localization in focal epilepsy: where are we now? *Epilepsia* **49(2)**:201–218 DOI 10.1111/j.1528-1167.2007.01381.x.

Polich J. 1993. Cognitive brain potentials. *Current Directions in Psychological Science* **2(6)**:175–179 DOI 10.1111/1467-8721.ep10769728.

pypi. 2024. EEG packages. *Available at https://pypi.org/search/?q=eeg{&}o={&}c= Development+Status+%3A%3A+5+-+Production%2FStable{&}c=Development+ Status+%3A%3A+6+-+Mature* (accessed on 25 June 2024).

PYPL. 2024. PopularitY of Programming Language. *Available at https://pypl.github.io/ PYPL.html* (accessed on 25 June 2024).

Rashid M, Sulaiman N, Abdul Majeed APP, Musa RM, Ab. Nasir AF, Bari BS, Khatun S. 2020. Current status, challenges, and possible solutions of EEG-based brain-computer interface: a comprehensive review. *Frontiers in Neurorobotics* **14**:25 DOI 10.3389/fnbot.2020.00025.

Rasmussen KB, Blank G. 2007. The data documentation initiative: a preservation standard for research. *Archival Science* **7**:55–71 DOI 10.1007/s10502-006-9036-0.

re3data. 2024. Registry of research data repositories. *Available at https://www.re3data.org/* (accessed on 25 June 2024).

Schnell S. 2015. Ten simple rules for a computational biologist's laboratory notebook. *PLOS Computational Biology* **11(9)**:e1004385 DOI 10.1371/journal.pcbi.1004385.

Scimago Journal Ranking. 2024. Health informatics open access journals. *Available at https://www.scimagojr.com/journalrank.php?openaccess=true&type=j&category=2718* (accessed on 25 June 2024).

Sharma N, Sharma M, Singhal A, Vyas R, Malik H, Afthanorhan A, Hossaini MA. 2023. Recent trends in EEG-based motor imagery signal analysis and recognition: a comprehensive review. *IEEE Access* **11**:80518–80542 DOI 10.1109/ACCESS.2023.3299497.

Shen M, Yang F, Wen P, Song B, Li Y. 2024. A real-time epilepsy seizure detection approach based on EEG using short-time Fourier transform and Google-Net convolutional neural network. *Heliyon* **10(11)**:e31827 DOI 10.1016/j.heliyon.2024.e31827.

Shin J, von Lühmann A, Kim D-W, Mehnert J, Hwang H-J, Müller K-R. 2018. Simultaneous acquisition of EEG and NIRS during cognitive tasks for an open access dataset. *Scientific Data* **5(1)** DOI 10.1038/sdata.2018.3.

Stevenson NJ, Tapani K, Lauronen L, Vanhatalo S. 2019. A dataset of neonatal EEG recordings with seizure annotations. *Scientific Data* **6(1)** DOI 10.1038/sdata.2019.39.

Sun C, Mou C. 2023. Survey on the research direction of EEG-based signal processing. *Frontiers in Neuroscience* **17**.

Tang M, Bever JD, Yu F-H. 2017. Open access increases citations of papers in ecology. *Ecosphere* **8(7)**:e01887 DOI 10.1002/ecs2.1887.

Teplan M. 2002. Fundamentals of EEG measurement. *Measurement Science Review* **2(2)**:1–11.

The Cochrane Collaboration. 2024. The University of Camberra's guidelines for systematic reviews in health. *Available at https://endoc.cochrane.org/sites/endoc.cochrane.org/files/public/uploads/CMED_protocol_template.pdf* (accessed on 25 June 2024).

TIOBE. 2024. TIOBE Index for November 2023. *Available at https://www.tiobe.com/tiobe-index/* (accessed on 25 June 2024).

University of California Irvine. 2024. Machine Learning Repository. *Available at https://archive.ics.uci.edu/* (accessed on 25 June 2024).

Valdes-Sosa PA, Galan-Garcia L, Bosch-Bayard J, Bringas-Vega ML, Aubert-Vazquez E, Rodriguez-Gil I, Das S, Madjar C, Virues-Alba T, Mohades Z, MacIntyre LC, Rogers C, Brown S, Valdes-Urrutia L, Evans AC, Valdes-Sosa MJ. 2021. The Cuban Human Brain Mapping Project, a young and middle age population-based EEG, MRI, and cognition dataset. *Scientific Data* **8(1)** DOI 10.1038/s41597-021-00829-7.

Vardigan M, Heus P, Thomas W. 2008. Data documentation initiative: toward a standard for the social sciences. *International Journal of Digital Curation* **3(1)**:107–113 DOI 10.2218/ijdc.v3i1.45.

Witzman S, Massey SL, Kessler S, Gonzalez-Giraldo E, Fridinger SE, Worden L, Lewin N, Dlugos D, Melamed S, Fitzgerald M, Fung FW, Ferruzi M, McNamee N, LaFalce D, Donnelly M, Haywood A, Allen-Napoli L, Banwell B, Abend NS. 2020. Acceptability of standardized EEG reporting in an electronic health record. *Journal of Clinical Neurophysiology* **37(5)**:455 DOI 10.1097/WNP.0000000000000639.

Won K, Kwon M, Ahn M, Jun SC. 2022. EEG dataset for RSVP and P300 speller brain-computer interfaces. *Scientific Data* **9**:388 DOI 10.1038/s41597-022-01509-w.

Xu G, Guo W, Wang Y. 2022. Subject-independent EEG emotion recognition with hybrid spatio-temporal GRU-Conv architecture. *Medical & Biological Engineering & Computing* **61(1)**:61–73 DOI 10.1007/s11517-022-02686-x.

Zaman A, Kumar S, Shatabda S, Dehzangi I, Sharma A. 2024. SleepBoost: a multi-level tree-based ensemble model for automatic sleep stage classification. *Medical & Biological Engineering & Computing* **62(9)**:2769–2783 DOI 10.1007/s11517-024-03096-x.

Zancanaro A, Cisotto G, Paulo JR, Pires G, Nunes UJ. 2021. CNN-based approaches for cross-subject classification in motor imagery: from the state-of-the-art to DynamicNet. In: *Proceedings of IEEE CIBCB 2021–the IEEE conference on computational intelligence in bioinformatics and computational biology*. Piscataway: IEEE, 1–7.

Zenodo. 2024. Research, shared. *Available at* https://www.zenodo.org (accessed on 25 June 2024).