# Quantitative EEG features related to the edge of criticality hypothesis and their predictive value of neurological outcome after coma

author: Sönke van Loh;
primary supervision: Prof.dr.ir. M.J.A.M. (Michel) van Putten;
chair: Dr.ir. T. Heida (Ciska);
external supervision: Prof.dr.ir. L.I. Segerink (Loes);

29.10.2024

## Contents

## Abstract

Humans have a sense of self and our environment as part of our consciousness. In clinical practice, we also deal with this phenomenon. This can be in the context of anesthetizing patients, comatose patients, or other circumstances that influence a patient's consciousness. Consciousness is evaluated based on behavioral metrics in clinical settings. This comes with disadvantages, however, as we know that nonreactive patients can be conscious, as in locked-in syndrome.

From modern theories of consciousness, we assume there is a strong connection between high-level neural dynamics and the conscious state of humans. Recent advances in the calculation of quantitative EEG have allowed for a resurgence of interest in whole-brain mathematical properties that can be used to investigate such neural dynamics. Here, we investigate non-linear features in the EEG related to the edge of criticality hypothesis. It postulates that awake consciousness lies at the border between chaotic and orderly dynamics where signal complexity and information processing are maximized. The two main metrics used are the 0-1 chaos test for the estimation of chaoticity in the signal and the Lempel Ziv complexity to estimate the complexity of the EEG signal—additional metrics of chaos estimation, fractal dimension, and entropy were used alongside during the analysis. Our study investigated three different datasets with states where we assume a change in consciousness: anesthesia, coma, and absence seizures. The aim is to see if prior findings from the literature can be reproduced in the data as, until now, only small sample sizes have been investigated.

We found a significant decrease in complexity and chaoticity during coma with good outcomes, seizures, and anesthesia compared to a waking baseline of similar age. In comas with bad outcomes, we found a large spread of chaoticity and complexity in the value range of altered and awake consciousness states. From the known literature, we expected a decrease in complexity in altered states. This agrees with our findings, except for comas with a poor outcome.

For chaoticity, we found the 0-1 chaos test output to correspond with the literature, with a value of 0.85 for waking baselines. This is close to the edge of criticality, according to Toker. In seizure states, a reduction is found in agreement with the literature. In anesthesia and comas with good outcomes, a decrease is also found. This disagrees with findings from the literature, which show that chaoticity estimates increased in those states. Therefore, our findings partially confirm and partially contradict earlier findings regarding the edge of criticality.

Our nonlinear metrics were also used to predict and investigate the level of cerebral damage after cardiac arrest in our coma dataset. The metrics were compared to a pipeline for predicting the neurological outcome after a coma called the Cerebral Recovery Index (CRI). We could show that permutation entropy is a promising new metric for classifying and evaluating comatose patients after cardiac arrest. It significantly increases the prediction performance of the CRI when features of both methods are combined.



Keywords:

EEG, criticality, complexity, chaos, epilepsy, anesthesia, coma


List of Abbreviations

1. EEG: Electroencephalogram
2. EMG: Electromyography
3. TMS: Transcranial Magnetic Stimulation
4. CPC: cerebral Performance Categories

5. rCRI: revised Cerebral Recovery Index
6. LZC: Lempel-Ziv Complexity
7. LLE: Largest Lyapunov Exponent
8. MSE: Multiscale (Sample) Entropy
9. MSPE: Multiscale Permutation Entropy
10. HE: Hurst Exponent
11. FDH: Fractal Dimension Higuchi
12. FDK: Fractal Dimension Katz

# 1 Introduction

Consciousness is the central phenomenon that allows us to perceive the world and ourselves. Observing and inducing altered states of consciousness are everyday routines in clinical contexts. This can be in the context of anesthetizing patients or pathological states like Seizures or Coma. Consciousness can be described at several different levels. In the clinical context, it is often defined as an awake or easily arousable state in which individuals can perceive themselves and their surroundings [4].

So far, behavioral scales like the Glasgow coma scale or Grady coma scale are the golden standards to assess the level of consciousness in the clinic [4], [15], [33]. However, the current view is that a patient can be conscious without the ability to react to external stimuli, as in locked-in syndrome. This limits the applicability of behavioral scales [15]. It creates the need for non-behavioral metrics of consciousness. Consciousness is so far known to be a mainly subjective and qualitative experience. This makes it hard to have a quantitative measure of it. However, it might not be impossible as modern theories of consciousness assume a strong relationship between our neural dynamics and the phenomenon of consciousness [29].

Theories on how consciousness could arise, like integrated information theory or global workspace theory, indicate that neural dynamics may be a potential measurement for the level of consciousness [25], [29], [30], [33]. They view the brain as a complex dynamical system where consciousness is not based on the individual building blocks but related to the global complex dynamics of the whole system. This study explores EEG signal properties related to these dynamics and their relationship to consciousness.

Using features from the EEG as non-behavioral markers of consciousness has seen a rise in interest over the last few years. One example of a measure of consciousness using EEG measurements of neural complexity is the Perturbational Complexity Index (PCI). It measures the complexity of the signal in response to a Transcranial Magnetic Stimulation (TMS) impulse and showed promising results in sleep, anesthesia, and coma patients for the quantification of consciousness [15]. In another paper, Maschke et al. show that the PCI can be predicted from resting state EEG without the TMS impulse [33]. They show that higher levels of consciousness show increased complexity and information transfer in EEG measurements. Increased complexity in awake consciousness was also demonstrated by Frohlich et al. [25] and Toker et al. [30] in their papers. In systems that can exhibit chaotic dynamics a point of maximum criticality is often associated with a bifurcation point between ordered and chaotic dynamics. It is connected to maximized computational efficiency and information transfer. These properties may also be critical for consciousness. That consciousness lies at this critical border of maximized information transfer between ordered and chaotic dynamics is referred to as the edge of criticality hypothesis [3], [24], [30], [33].

The papers mentioned above either use models or small sample sizes of patients. On the one hand, models are needed for a mechanistic understanding and are favorable due to allowing noise-free measurements of the system while observing its hidden states [22], [30]. On the other hand, models can always describe reality only within a constrained framework, especially when dealing with high-level phenomena in complex systems.

Our work focuses on testing measurements for neural correlates of consciousness in three independent datasets. We test if we can find a point of maximum criticality during fully awake states compared to states of altered consciousness. Additionally, we test if this critical point can be related to a point between order and chaos where seizures represent a deviation in the ordered direction and anesthesia or coma into the chaotic direction as described by Toker et al. [30].

Then, we use our metrics to predict patients' Cerebral Performance Category (CPC) after cardiac arrest. We are interested to see if there might also be a link to the neurological health of a patient. We are comparing our complexity-based metrics to an updated Cerebral Recovery Index (CRI) first published by van Tjepkema et al. [20] in 2017.

Summarizing, we will answer the following two questions: 1. Can we reproduce the findings from Toker in our datasets on the complexity and chaoticity of EEG signals in anesthesized patients and patients during a seizure? 2. Can the computed metrics improve the prediction of neurological outcome after cardiac arrest?

## 2 Background

Our work is based mainly on the paper of Toker et al., ”Consciousness is supported by near-critical slow cortical electrodynamics” [30]. To explore and understand the hypothesis, we elaborate on relevant concepts of criticality and chaoticity and explain their relation to the edge of criticality hypothesis and consciousness.

Our brain comprises around 86 billion neurons with trillions of connections [35]. The neurons and synapses connecting the brain are nonlinear in their behavior, making the brain a highly complex system governed by the interplay of nonlinear subsystems. Systems that fulfill these conditions are known to exhibit chaotic behavior. Chaotic systems are abundant in nature, with well-known examples being the weather or population dynamics [28]. What does it mean for a dynamic system to be chaotic? There are multiple exact definitions of when a system is chaotic, as not all non-linear systems necessarily need to exhibit chaotic dynamics. We will use a definition for deterministic chaos here with three necessary conditions the system must fulfill. First, it needs to be composed of nonlinear subsystems or connected in nonlinear ways, causing the whole system to be nonlinear. Second, the underlying systems must act deterministically and are not driven by stochastic processes. If someone knew the system’s initial conditions perfectly, he could model the system outcome deterministically. Lastly, tiny changes in the system’s initial condition can cause an unpredictable outcome, assuming an outside observer does not know everything about the system in detail. This is due to the exponential divergence of states over time caused by the non-linear dynamics [19]. This makes it seem almost like a stochastic system to an outside observer who can not know all its initial conditions perfectly. Unlike in a linear system, these conditions also entail that the system’s global behavior can not be fully understood by dissecting it into its parts and studying their behavior due to the necessity of understanding all its interactions to understand its behavior [7]. A more precise mathematical description of chaos can be given, which constrains the attractors of the system [19]. We did not test the system at the level of its attractors, and the system fulfills the necessary conditions to exhibit chaotic behavior.

Now, the question arises of how a phenomenon like consciousness may be connected with these high-level properties of our brain as a deterministic chaotic system and how that relates to an edge of criticality. There are many theories on how consciousness may be connected, or even caused, by our neural dynamics. There is no clear consensus on which one is true or false [29]. Our analysis assumes an information-theoretic connection between the epiphenomenon of consciousness and our neural substrate. One theory is that consciousness is associated with maximizing the computational efficiency of our brain. It is the state we are in most of the day, where we must make decisions, recall memories, create new ones, and understand relations between them. There is a multitude of mathematical concepts and measures to classify how well a complex dynamical system can process, store, and transfer information [27], [33], [36]. Maximizing these processes might be connected to a critical point of phase transition between order and chaos. This critical point is well known in other dynamical systems and models in mathematics and physics [21], [36].

The edge-of-chaos criticality, visually depicted in fig 1, connects wakeful brain dynamics with a phase transition between order and chaos. Operating near a critical edge optimizes information processing capabilities while being deterministic and predictable to a sufficient degree. To properly process and utilize information, there must be a balance between storing and integrating information and exploring new possibilities. Therefore, the system needs to be deterministic so far that it can reproduce prior states while also allowing for enough exploration [27], [30], [34], [36]. This balance is often thought to be optimized around the critical point between order and chaos. As we are conscious during most of our lives and must be able to integrate and explore new possibilities continually, it seems plausible that consciousness would coincide with such a point of optimal information processing [27], [30], [33], [34].

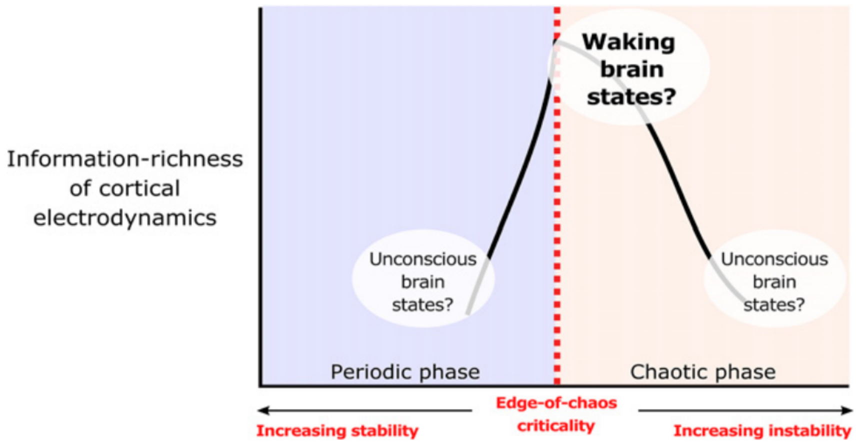



**Figure 1:** Depiction of the edge of criticality. The image is taken from Toker et al.'s paper "Consciousness is supported by near-critical slow cortical electrodynamics" [30]. The x-axis shows an increase in chaoticity from ordered dynamics (left) to entirely chaotic dynamics (right), with the critical point in between. The brain's complexity, or information processing, is depicted by the black line with a peak at the critical point (red dotted line), decreasing towards the outer limits.

# 3 Materials and Methods

Testing chaoticity and complexity in time series data is not trivial, as we can't fully understand the system from its outputs. Adding to that, the output will always be the signal of interest combined with noise from the measurement. Nonetheless, there have been advancements made in recent years in developing metrics to estimate complexity and chaoticity in time-series data [5], [10], [21], [28]. We will use these in our datasets to test the edge of criticality hypothesis and the outcome of coma patients. A summarization of the processing steps can be found in the workflow diagram 2.

## 3.1 Data sets

Three clinical datasets with different pathologies are used. Table 1 shows an overview of patient characteristics.

Set A consists of 21 patients who underwent carotid endarterectomy. Two measurements are done: one before surgery as a baseline while awake with eyes closed and one during propofol anesthesia. We consider the subjects fully conscious during the waking baseline, comparable to a healthy adult.

Set S consists of 23 children with absence epilepsy. Each patient's EEG recording includes seizures and an awake baseline.

Set C consists of 395 patients in a coma 12 hours after cardiac arrest. No awake baseline measurement is present for these patients.

**Table 1:** Overview of patient characteristic. For set A and set S, each patient has both labels. However, the exact absolute time of annotated data can vary per patient. For Set C, each patient only has one label, with the following distribution: good outcome (CPC 1: 137, CPC 2: 52), poor outcome (CPC 3: 10, CPC 5: 196). The CPC scores are assigned six months after cardiac arrest.

| Dataset | age (std) | sex (m/f) | label |
|---|---|---|---|
| Set A | 66 (8.4) | 14/7 | awake, anesthsized |
| Set S | 9 (3.6) | 12/11 | awake, seizure |
| Set C | 62.7 (13.6) | 288/107 | poor outcome, good outcome |

## 3.2 Epoch selection

For set A, a neurologist annotated "Awake eyes closed" states in the baseline and anesthesized parts during anesthesia. The annotations excluded artifacts like eyeblinks or movements. In our analysis, we limit ourselves to these annotated data sections spanning roughly 40-200 seconds in length [14].

In set S, we included patients with at least one seizure of 10 seconds or longer. Besides the seizures, we annotated artifact-free epochs of awake baseline with the help of a neurologist for every child in the same EEG recording. The data was used prior for another paper investigating visual attention during absence episodes [31].

In set C, a 5-minute artifact-free epoch was annotated by a neurologist. Six months after cardiac arrest, a CPC score is assigned for each patient. We split the group into two different outcomes. A good outcome after a coma due to cardiac arrest is a CPC score of 1 or 2. Patients have no to moderate cerebral disability. They are regarded as conscious and can work. Sometimes, they require a protected environment. A poor outcome is a CPC score of 3,4, or 5, indicating severe cerebral disability, coma/vegetative state, or death. With a CPC score of 3, patients are still considered conscious but require assistance in their daily lives. The CPC scores do not directly reflect the EEG during recording due to the long time between EEG acquisition and CPC scoring. We know that there are very few subjects with a CPC score of 5 who did not die due to a neurological reason but of other complications. We did not exclude these patients as a manual revision of the dataset with a neurologist would be needed. This would have been unfeasible in the timeframe of this work.

## 3.3 EEG preprocessing

We used mne 1.6.1 [16] with Python 3.11 and Matlab R2023b for all processing of EEG data. Mne is a Python toolkit specialized for processing MEG and EEG data. It allows easy loading, storing, and handling of the data like annotations, montages, and filter operations. We bandpass filter the annotated EEG segments with 0.5 - 25 Hz. We use a low cutoff frequency, omitting some potential neural signals because muscle artifact noise also increases with higher frequencies. As our measurements can be noise-sensitive, we try to minimize muscle artifacts' influence. Mne automatically adapts the filter to keep it stable, with zero lag and a steep cutoff. The signal is referenced to a bipolar double banana montage and downsampled to 100 Hz. These parameters were determined based on a preliminary analysis of set A. We found that the metrics significantly changed concerning preprocessing parameters. This includes epoch length, sampling frequency, and filter bands. To minimize the bias introduced by these parameters, we chose to use unified preprocessing for all metrics. More extended results of this analysis can be found in appendix A.

It was specifically chosen to keep the data preprocessing minimal. The only goal of the preprocessing was to keep as much of the raw neural signal as possible. Only known noise sources, such as EMG artifacts or eye blinks, are removed by filtering and epoch selection. Second, the literature indicates that minimal preprocessing of EEG can be better than sophisticated preprocessing depending on the question at hand [32].

The signal was downsampled to 100 Hz. Lower sampling frequencies speed up computations. This is relevant as the goal was to analyze more EEGs than other studies for robust findings. Since all signals were initially sampled with more than 100 Hz, downsampling them unified their sampling frequency. All frequencies of interest are contained within the Nyquist frequency (50 Hz), and there is enough room for the filter to converge between the 25 Hz cutoff and the 50 Hz.

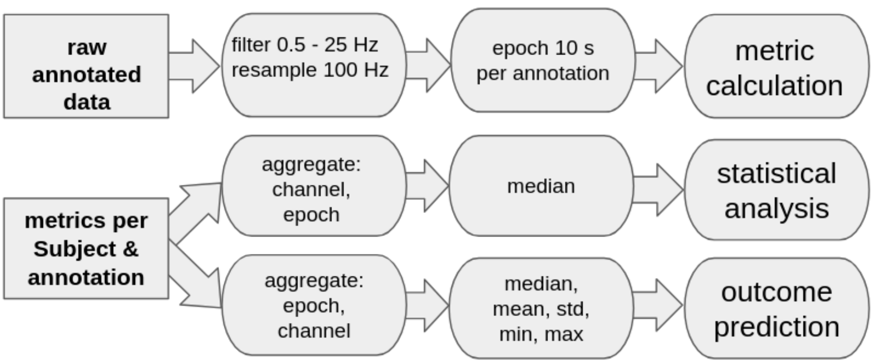



**Figure 2:** Workflow diagram of the preprocessing steps. The upper part shows the central processing of the EEGs for metric calculation. Additional steps like downsampling, binarization, or filtering may be done within the metrics. On the lower level, the metric aggregation and evaluation are displayed. Above for the edge of criticality analysis and below for the outcome prediction after the coma.

## 3.4 Metric calculation in EEG

We split each previously annotated EEG epoch into 10-second epochs. The remainder will be dropped if the annotation length is not dividable by 10. Ten seconds are chosen due to the limited duration of the seizures in the absence dataset. Our preliminary tests showed that some metrics show a time dependence, making it unfeasible to use annotations of different lengths in the metric calculation. Therefore, we used the length of the shortest annotations included in the analysis. They match other papers' choice of epoch length [30]. For each 10-second epoch, all metrics are calculated per channel. Then, we average over all channels per metric to get the channel mean. These means are again averaged over all 10-second epochs per patient and annotation. As the annotations are of very different lengths, the number of averages differs, especially between sets. This was not circumventable due to the nature of the available data, but it needs to be considered in a later analysis. We end up with one value for each state of consciousness in a subject per metric.

In the following, we will briefly introduce the metrics we used to test for the edge of criticality. Explaining the metrics in detail here would go beyond the scope of this paper, but a reference with more information will be added to each metric for the interested reader. We use two Python toolboxes to compute the metrics. The first one is neurokit2 [23]. It focuses on estimating physiological signals' complexity, entropy, and fractality. It is open-source and free to use. The second one is edgeofpy. It is a smaller open-source Python package for chaoticity and avalanche criticality metrics. Both packages are also used for the analysis of chaoticity, entropy, complexity, and fractality by Maschke et al. [33] in their paper.

#### Lempel Ziv Complexity

Lempel Ziv complexity (LZC) measures the complexity of a time series. It is one of the two metrics that Toker bases the observations for the edge of criticality on [30]. It determines how large the vocabulary of a system needs to be to reproduce the time series. A small amount of repeated vocabulary can construct a regular time series. The more complex a time series is, the less repeatability it has. One downside of the method is that a long sequence automatically has a higher chance of being classified as complex than a short sequence. This follows from the property that a complex sequence is built up of more parts than a simple sequence [1]. We use mean symbolization because Lempel Ziv complexity is designed for binarized signals. Every value above the signal mean is assigned a one, and every value below the signal mean is assigned a 0. It is the most common mode of symbolization used.

#### 0-1 chaos test

The 0-1 chaos test is the primary metric that Toker uses to estimate the chaoticity of the EEG signals to test the edge of criticality hypothesis. He applies it in a pipeline that first tests the signal to determine if it has a deterministic basis and is not purely stochastic, then denoises it using Schreiber's denoising algorithm [6], and lastly, downsamples it by only keeping the local minima and maxima of the series. Then, the 0-1 chaos test is applied to the processed signal, which returns a number between 0 (ordered) and 1 (chaotic)[10], [21]. Toker relates the critical point to a value of around 0.85 [30]. In our analysis, we will not use the exact pipeline Toker used, which is implemented in Matlab, but instead, a Python translation of it from edgeofpy, which was also used by Maschke et al. in their paper [33]. Here, we also omit the first step of checking for the stochasticity of the signal as we know its source. One difference to the other metrics is that we have to filter the signal again for this one, as both Toker and Maschke used it for low-frequency dynamics. Toker et al. used a FOOOF algorithm to filter with a cutoff coinciding with the first spectral peak under 6 Hz [30]. Instead, we use a constant filter of 4 Hz, as Maschke et al. did [33], for ease of use and to keep the analysis simple.

#### Multiscale Sample Entropy

Multiscale Sample Entropy (MSE) is the multiscale version of Sample Entropy. Sample Entropy was explicitly designed for physiological signals to be robust to measurement noise [9]. We are interested in testing it as an additional measure for the chaoticity of our signal as it showed an almost perfect significant positive correlation to low-frequency ($< 6$ Hz) chaotic dynamics measured by the 0-1 chaos test in the paper by Maschke et al. [33]. Sample Entropy is based on an embedding approach where different embeddings of the signal are compared. See the following paper for details on the procedure [9]. In multiscale approaches, the entropy is calculated on multiple downsampled signal versions and averaged over all results. This is called different granularities. For the different granularities, we use the 'default' scale, which calculates the various factors used to divide the signal length by the following formula $range(\frac{len(signal)}{dimension+10})$. With a dimension of three, the default value from neurokit2, and a signal length of 1000 samples, that becomes $range(76)$, which means 76 different granularities per signal.

#### Largest Lyapunov Exponent

The Largest Lyapunov exponent (LLE) measures how fast (infinitesimally small) changes in initial conditions lead to a divergence in the systems states. One typical behavior of chaotic systems is the exponential diverge of states due to infinitesimal small changes in initial conditions. The Largest Lyapunov exponent of a system is the largest positive exponent of such an exponential growth.

LLEs are the primary metric in simulations to judge the nonlinear behavior of chaotic dynamics. We are interested in whether their values, when estimated in data, will coincide with our other measures of chaoticity. One known problem of estimating LLEs in data is that noise significantly impacts the estimation of the LLE. In our paper, we use Rosensteins' estimation of the LLE, commonly used in literature. It claims to be robust against noise [5]. It is based on a time delay embedding of the signal. The parameters we use for the time delay embedding are one sample of delay with two embedding dimensions. We decided against optimizing time delay and embedding dimension and stuck to the toolkit's default values. First, this has computational reasons, as the computations are costly and must be redone for every signal. Second, we found in an initial embedding test (appendix A) that different versions of calculating the embeddings and time delays come up with vastly different values. Next, each channel of a single EEG also showed a vast value spread, especially in the embedding delay. These findings make the added value of these computational costs questionable, as the choice of which function to use for approximating the dimensions and delays would again be arbitrary. Therefore, we decided to stick to the standard parameters with low dimensions and time delay.

## 3.5 Testing the edge of criticality

To test the edge of criticality hypothesis, we first evaluate the complexity of the different states of consciousness with the LZC. Then, we evaluate the different states' chaoticity using the 0-1 chaos test. Using these two measures, we try to see if, in our datasets, we can recreate the characteristic inverse U shape that Toker found with chaos on the x and complexity on the y-axis. For detailed information, see the Literature section 2. Additionally, we test chaoticity using the LLE and MSE. MSE showed good representative quality for low-frequency ($< 6$ Hz) chaotic dynamics in Maschke et al.'s papers [33]. Therefore, we expect it to show the same differences as the 0-1 chaos test. LLEs are the primary measure to judge chaoticity in simulations and perfectly match the 0-1 chaos test in simulated noise-free data in Toker et al.'s paper [30]. However, they are noise-sensitive. We want to know if they behave like the 0-1 chaos test in measured physiological signal data or show diverging behavior. We will compare all three different chaoticity measures concerning their relative differences within and between datasets.

### 3.6 Prediction of coma outcome

To test the predictive value for neurological outcomes after coma due to cardiac arrest, we predict the CPC score of a patient from our metrics. For this, the metrics from above are used, and an additional set of metrics, which were also used in Maschke et al.s paper [33]. The metrics are closely connected to chaos and complexity and consist of two fractal dimension estimates, Hurst exponent and permutation entropy. We limit ourselves to the binary problem of predicting a good (CPC 1 or 2) and poor (CPC 3,4 or 5) outcome. Only EEGs 12 hours after the cardiac arrest are included. This time point was optimal for predicting poor outcomes in van Tjepkema et al.'s paper on the Cerebral Recover Index (CRI) [20] that we use in this analysis as well.

We aggregate the metrics. This time, we aggregate per channel all measurements across the 10-second epochs per annotation. Then, we aggregate the metrics over the channels using the median, mean, standard deviation, minimum, and maximum. We use more different aggregation methods here, as we know from our initial hyperparameter tests (appendix A) that there are differences across the channels. This way, there is a tradeoff between using every channel for every metric and losing all information about the variance within the different channels.

We compare our criticality-related metrics with an updated version of the CRI [20], where some additional metrics were added. We calculated the features for the same subjects. The CRI features internally aggregate over the EEG channels and epochs, so we only have one set of metrics per subject.

Lastly, we create a metric set consisting of the combination of the criticality-related and the CRI features. This way, we can test whether combining feature sets performs better than individually.

We predict the coma outcome using random forest classifiers implemented in Python 3.11 with Scikit learn 1.4.1 [13]. The data is divided into a train and test split. Overall, we have data from patients from 6 different hospitals. We used 240 patients from 4 different hospitals in the train set and 133 patients from the other two hospitals in the test set. Twenty-two patients from the 395 patients of set C were excluded due to the inability to calculate all metrics on their EEGs. Ten from the train and twelve from the test split. No preselection or visual investigation of the data was performed prior. Using only data from the train split, we tuned the random forest using 5-fold cross-validation. We did not perform a systematic grid search but instead tuned the forest hyperparameters manually, optimizing the prediction accuracy of all metric sets simultaneously. A change that influenced only one specific metric set was not used. We ended up with the following hyperparameters: criterion=entropy, n_estimators=200, min_impurit_decrease=0.01 and max_features=0.25. Everything else is left as default.

Using the tuned random forest, we train on the complete train set and validate on the test set. We compare the performance of the three different metric sets on the final test set using Receiver Operator Curves (ROC). After tuning, we investigate each set's feature importance using the mean entropy decrease in a 5-fold cross-validation of the train set.

## 3.7 Statistical evaluation

Finally, we will conduct statistical tests on the differences within the datasets for the edge of criticality hypothesis and between the classifiers for predicting the neurological outcome.

### Edge of criticality

We use a Wilcoxon signed-rank test to identify significant systematic differences in signal complexity and chaoticity within sets A and S. These sets are too small to assume a normal distribution based on the central limit theorem and have related samples. We use a Mann-Whitney U test between sets for statistical differences in the mean. Sixteen tests are done. Bonferroni correction with factor 16 is applied to the p-values to correct for the increased chance of finding false positives.

### Neurological outcome

To compare the outcomes of the three different classifiers, we use a permutation test with 100 permutations. The original AUC score is calculated and then compared to permuted AUC scores, where the prediction probabilities of the classifiers are randomly shuffled with each other. Bonferroni correction is applied with a factor of 3 for the three comparisons between classifiers.

## 3.8 Data and Code availability

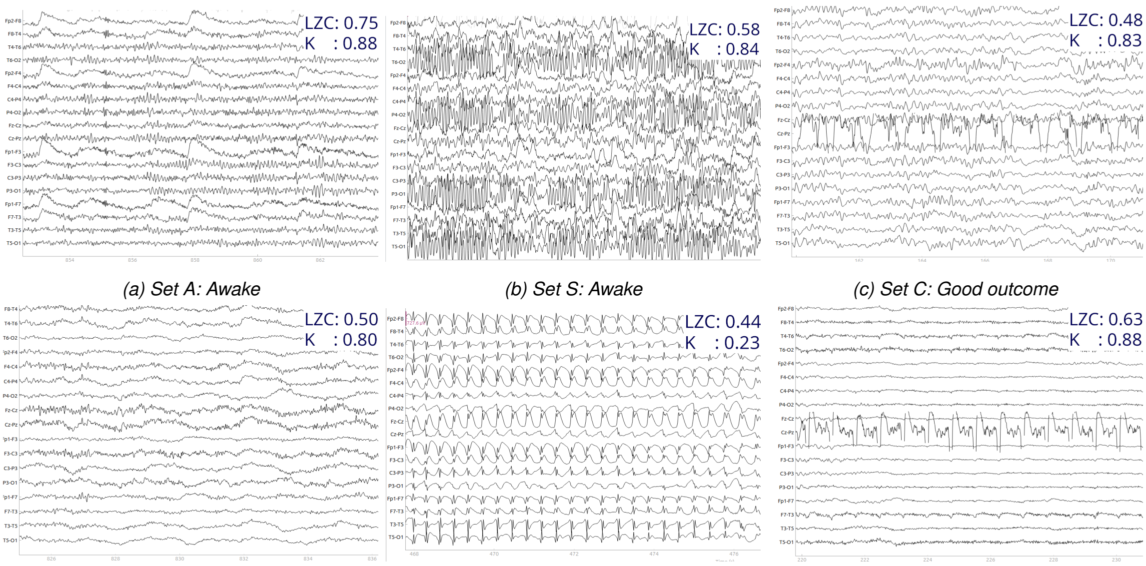
A version of the pipeline used for analysis can be found under EEGAnalyzer on my GitHub. Parts of the analysis were removed for privacy reasons. All study data and the full analysis code are saved and may be made available by contacting the CNPH group of the University of Twente upon valid request.

# 4 Results

First, exploring the edge of criticality, we investigate the relationship of the conscious state with complexity and chaoticity estimated from the EEG signals. Then, in the following subsection, we investigate how our metrics relate to the neurological outcome of patients after cardiac arrest and how they compare to the established CRI metric.

## 4.1 Edge of criticality

Some examples of analyzed EEG signals are shown in figure 3, including their complexity (LZC) and chaoticity (K). These can not reflect all the possible patterns encountered during this analysis, but they show differences between the EEG for the different patient groups and their altered states of consciousness.




*(a) Set A: Awake* *(b) Set S: Awake* *(c) Set C: Good outcome*

*(d) Set A: Anesthetized* *(e) Set S: Absence seizure* *(f) Set C: Poor outcome*

**Figure 3:** EEG signal examples for all conditions analyzed in this paper. The complexity (LZC) and chaoticity (K) of the respective sample are displayed in the top right. All samples included around 10s of the signal used in the analysis. Samples were chosen to reflect the median value of LZC in the analysis per condition. All EEGs have the same scaling for amplitude besides the Seizure signal (e). We downscaled the signal by the factor $\frac{40}{727.6}$ to display the EEG, because it has much higher amplitudes. This factor was determined by the internal steps used in mne to display the EEG. Next to the higher amplitudes, the seizures also show a high degree of inter-channel synchrony with very regular patterns. All the EEGs are visually distinct. This includes the two EEGs in awake states recorded once in an adult population (top left) and once in a pediatric population (top middle).

We compare if the LZC and 0-1 chaos test can reflect the edge of the criticality hypothesis regarding the change in complexity and chaoticity. During that, we will also note differences to the analysis conducted by Toker et al. [30] and Maschke et al. [33]. Then, we compare whether different methods of chaos estimation yield comparable results or if they differ considerably.

### Lempel Ziv complexity

The change in complexity across the different patient conditions is depicted in figure 4a using the LZC. Awake states show significantly higher complexity than seizure, anesthesized, and coma states with good outcomes. The two awake states of adults and children show a significant difference with lower complexity in awake children. The reduction in complexity from awake to altered states of consciousness coincides with the theoretical curve from the edge of criticality hypothesis. Toker and Maschke observed a decrease in complexity during anesthesia in their research. Toker also observed a reduction during seizure states, which is the same as ours. Comas with poor outcomes are vastly spread across the metric in both ranges associated with awake and altered states of consciousness. Coma was neither tested by Toker nor Maschke, so we can not make a direct comparison here.

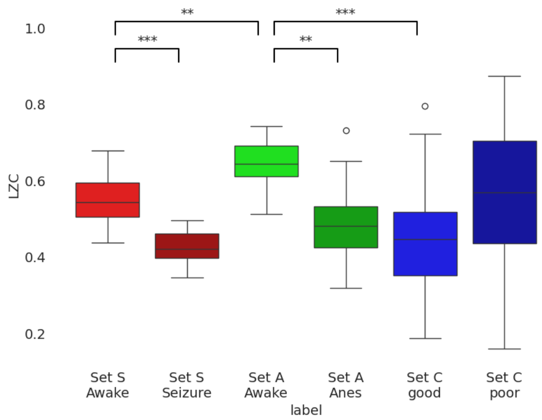
### Chaos estimation with the 0-1 chaos test

The results from the 0-1 chaos test are depicted in figure 4b. Before comparing the chaoticity of the Sets, it was found that the 0-1 chaos test is not calculable on every time series. Our results found that many of the EEGs in the coma patients could not be calculated. This reduced the data in the poor cases by 56.8% and the good cases by 45.9%. For Set A, we lost 2.2% of data in the awake cases. For the anesthetized cases of Set A and the whole Set S, no data was lost during the 0-1 chaos test calculation. This could lead to a systematic bias in the results, especially in Set C.

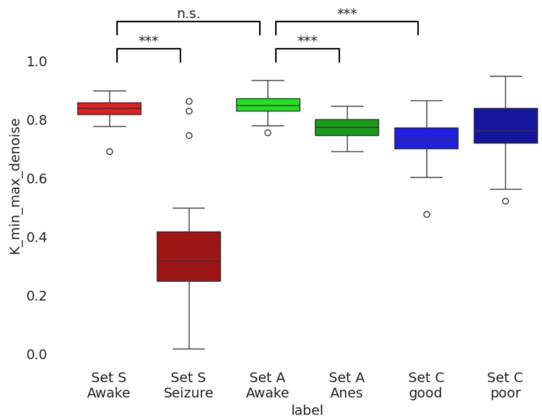
There is a significant decrease in chaoticity in seizures compared to the awake baseline. Toker also found this in their analysis. The awake baselines of children and adults are on par, showing no significant differences. The median values of the awake children lie at 0.84 and the adults at 0.85. This coincided with Toker's hypothesized edge of criticality in their analysis. Comas with good outcomes and anesthesia show a significant decrease in chaoticity. The deviation from baseline is lower than that for the seizure states. This contradicts findings from both Toker and Maschke, which found that chaoticity increased during propofol anesthesia. As mentioned above, we can not be sure about the distribution of the coma results due to the high data loss while calculating them.

### Secondary chaos measures

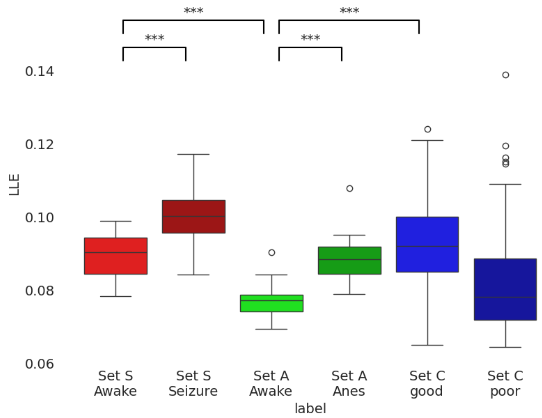
Our secondary measures of chaoticity are depicted in figure 4c using the LLE and figure 4d for the MSE, respectively. Both metrics show a significant increase in estimated chaoticity under anesthesia and coma compared to awake adults. This differs from the 0-1 chaos test in our results but matches the findings from Toker and Maschke. Comas with poor outcomes show lower values of chaoticity on par with healthy adults. This is similar to what we found in our LZC measures, where comas with poor outcomes covered both regions of awake and altered states of consciousness. For the children with absence epilepsy, the metrics differ. LLE increases during the seizures compared to the children's baseline measurements. This contradicts both the 0-1 chaos test and the original hypothesis. MSE decreases like in the 0-1 chaos test. There is a difference in baseline measures of adults and children with lower chaoticity in the children. This also differs from the 0-1 chaos test results.



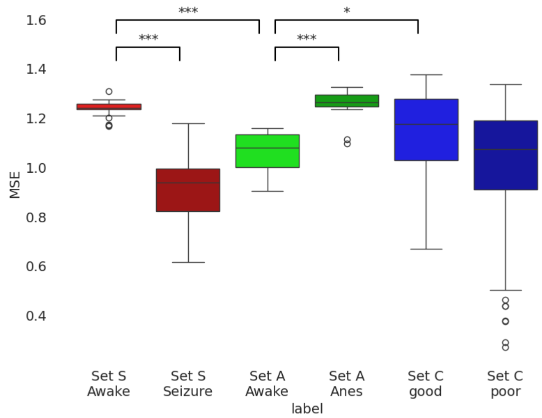

*(a) Lempel Ziv complexity*




*(b) 0-1 chaos test*




*(c) Largest Lyapunov exponent*




*(d) Multiscale sample entropy*

**Figure 4:** Comparison of chaoticity and complexity estimates on the Absence (red), EIRatio (green), and Coma (blue) datasets. Graphs represent the distribution of values across the labels of the datasets for the according metrics. Subsets that are compared to each other are indicated with the bars above. The significance in differences of the population means after Bonferroni correction with a factor of 16 is indicated by the asterisk: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. In the 0-1 chaos test, around half of the EEGs for Set C were not computable in both labels. On top, the two main metrics for the edge of criticality hypothesis are displayed. On the bottom are the two additional chaoticity estimations.

## 4.2 Prediction of neurological outcome after coma

Next to investigating the edge of criticality hypothesis using our calculated metrics, we also use them to predict the neurological outcome after coma due to cardiac arrest to gain more insight into their connection to possible brain damage in a patient. We compare our metrics to the ones from the CRI [20]. We are interested in the importance of our features during the training process and if they can benefit the outcome prediction.

Table 2 shows the feature importance during the training. In the criticality-related features, permutation entropy dominates over the other metrics, but LLE and MSE also show predictive value. Permutation entropy stays the strongest predictor when combining the features of the CRI and the criticality-related features. Besides permutation entropy, the alpha-delta ratio and the brain continuity index from the CRI are strong predictors for the CPC after coma.

**Table 2:** Feature importance based on the average decrease in entropy in the five-fold cross-validation. Feature importance was calculated for each forest. Features are sorted in decreasing order by their average entropy decrease in a forest. 5/5 top 5 features means the feature was under the top 5 in every tree. 4/5 top 10 features mean features were in 4/5 trees in the top 10.

| Feature set | 5/5 top 5 features | 4/5 top 10 features |
|---|---|---|
| complex Features | PE-median, Pe-mean, PE-min | PE-max, LLE-median, MSE-min, LZC-max |
| CRI Features | ADR, BCI, BPM | SE, BSAR, CRI, mean-BC |
| combined Features | PE-mean, PE-median, PE-min, ADR | PE-max, BCI |

To test the performance of the metrics in unseen data, we use the test set with data from different hospitals. The results for the test set are depicted in fig 5. In the sets including the complexity-related metrics, we find better sensitivity at 100% specificity. The only significant difference in AUC scores was between the combined and the CRI features.

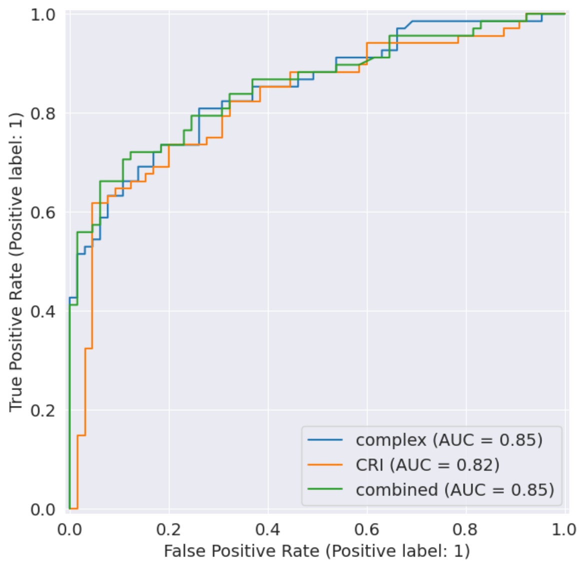



**Figure 5:** Comparison of the three feature sets on the left out external test set. A statistically significant difference in AUC score is found between the combined and the CRI approach with $p = 0.03$ after correction.

# 5 Discussion

First, we discuss our criticality-related metrics, focusing on the paper of Toker et al. [30], which mainly postulated the edge of criticality hypothesis, and the paper of Maschke et al. [33] who did an extensive testing of criticality related metrics in their analysis. Then, we will discuss the predictive value of our metrics in coma after cardiac arrest cases.

## 5.1 Edge of critcality

We tested the hypothesis postulated in Toker et al.'s paper on consciousness lying on the edge of criticality. Entailing that awake consciousness lies at a maximum complexity between chaotic and regular dynamics [30]. Both Toker and Maschke showed this to be testable in data. They used small sample populations of mainly anesthetized patients and some patients with seizures [30], [33]. To retest that hypothesis, we used three different datasets. We computed several criticality-related metrics and will compare them to the original results.

#### Complexity in EEG signals

Starting with LZC, the main metric in our and Toker et al.'s paper to estimate the complexity of the EEG [30]. We found a correlation between the level of consciousness and complexity. Anesthesia and Seizures showed the anticipated significant decrease in complexity compared to awake baselines. Additionally, in our study, we also found this in patients with a good neurological outcome in a coma after cardiac arrest. This maps the inverse U shape of the Toker hypothesis, where awake consciousness lies at the top, and states of reduced consciousness like seizures or anesthesia would show a lower complexity due to less active information processing [30]. Maschke et al. also made the same finding for awake and anesthetized patients[33]. However, we also found high values of complexity in a coma with a poor neurological outcome. This could contradict the one-to-one correlation between the level of consciousness and LZC. It seems unlikely, from our current understanding of coma, that persons with a worse neurological outcome would show a higher degree of consciousness during coma comparable to awake states. Next, we also found significant changes between different age groups, with higher values in adults than in children. According to the hypothesis, this could mean adults are more conscious, but from interacting with children and adults, one knows both tend to have a vivid inner life and are conscious about their surroundings. Therefore, it is more likely that factors other than the level of consciousness also play a role in the complexity of the EEG signal. However, it is not impossible that there might be differences in information processing between children and adults or that at least some patients in a coma with a bad neurological outcome have a vivid inner life. A lot of the neurological bad outcomes had a CPC of 5, meaning the patients died. Factors influencing the measurements may be different noise components in the EEG acquisitions. We also observed differences in EEG amplitude between groups, potentially impacting the metrics and signal-to-noise ratios. Due to time constraints, no further testing was done on this.

#### Chaoticity in EEG signals

The second part of Toker's thesis regarding the edge of criticality is that the chaoticity of the signal needs to be at a critical point between order and chaos to achieve this maximized point of information processing indicated by the signal complexity [30]. We used the same metric as toker with the 0-1 chaos test. However, we simplified our measurements by filtering with a standard lowpass of 4 Hz instead of using the FOOOF algorithm to find a particular peak under 6 Hz to determine the filter cutoff. However, Maschke et al. did show that both methods should yield only minimal deviations in results [33]. First, we found that our awake subjects, children and adults, had chaoticity estimates that matched the critical edge Toker found with a K value of 0.85 [30]. That waking baselines confine to values associated with the critical edge while also showing maximized complexity values strengthens the claim towards optimal information processing being related to it. We also saw a significant decrease in chaoticity in seizure states. This supports the hypothesis that seizures reduce consciousness due to moving it toward more ordered dynamics away from the critical edge. This makes sense as seizures are known to be caused by synchronization, where neurons fire in synchronized and regular patterns [17]. In anesthesia patients, we also found a significant decrease in chaoticity. This contradicts the results Toker and Mascke found in their papers, where the chaoticity of slow cortical dynamics increased during exposure to propofol [30], [33]. We do not have a conclusive explanation of where this difference comes from. Patients in a coma with good neurological outcomes also show the same behavior of significantly decreased chaoticity in our experiments. However, we also experienced technical difficulties during this calculation, so we can not exclude some systematic bias in the coma data introduced by left-out EEGs.

However, our secondary metrics for the chaoticity of the signals, the MSE and the LLE, did find the expected increase of chaoticity during anesthesia and coma with good outcomes. On the one hand, this could show that maybe something went wrong in our experiment's estimation of chaoticity using the 0-1 chaos test. There were slight deviations in our data preprocessing compared to Toker and Maschke's papers. On the other hand, we also found varying results for the awake baselines and seizure states using these secondary metrics. In MSE and LLE, a significant difference exists between the awake states of children and adults with children showing higher chaoticity in LLE and lower in MSE. We even found an increased chaoticity during seizure states in LLE measurements directly contradicting the 0-1 chaos test results in both our experiments as well as Toker and Maschke's experiments [30], [33].

This widespread and contradictory behavior of results in chaos estimation shows that estimating a signal's chaoticity seems complicated and could be a limiting factor in this kind of analysis. The metrics are often abstract, and it is hard to predict how they will react to specific patterns in data. Further research would be needed to understand where the contradiction in results between our analysis and Toker's analysis comes from. These unknown differences and the complexity of chaos estimation metrics also reduce their interpretability on physiological signals, an essential factor for acceptance into clinical routine [26].

#### Ground truth problem

Next to the unknowns mentioned above, there is also a general ground truth problem regarding the interpretability of results. We are interested in finding a nonbehavioural metric for consciousness but can only use behavioral outcomes to investigate possible methods. This limits the certainty we can have in making claims on the results. Nevertheless, showing apparent differences between the groups can still indicate the potential for finding a nonbehavioural metric.

## 5.2 Prediction of neurological outcome

Besides investigating the edge of criticality, we used our complexity-based metrics to predict the cerebral damage people suffer after a coma due to cardiac arrest. In that process, we were interested in connecting our metrics to patients' neurological health. Showing a clear link between the predictive quality of neurological outcomes and our metrics could indicate that they reflect more than the patient's conscious state.

We found that our metrics could predict the neurological outcome on par with the metrics from the CRI. On an external test set with data from other hospitals, we found a significant increase in prediction accuracy when combining our complexity-based metrics with the CRI metrics versus just using the CRI. Permutation entropy was the strongest predictor in the trees, independent of using only the complexity-based or if combined with the CRI metrics. However, LZC, LLE, and MSE were among the strongest predictors when using only complexity-based metrics. Regarding the absolute performance of our classifiers, only very minimal tuning and feature selection were made. A dedicated analysis of hyperparameters in the machine learning model and training data could further improve prediction results. As we tuned the forests by hand, this may have introduced a bias towards specific metrics, even though the tuning goal was to reach the highest prediction accuracy among all metric sets.

We conclude from the results that our metrics, especially permutation entropy, seem valuable in predicting neurological outcomes after a coma. The metrics seem to, in different degrees, also reflect the neurological health of the cerebral tissue in a patient. If this is due to a link between a patient's consciousness and his level of brain damage or due to other reasons, can not be concluded based on this analysis. So far, the literature has focused mainly on patient groups under anesthesia or epilepsy patients in an investigation of consciousness [15], [30], [33]. This limits a possible comparison to other findings in the literature.

Permutation entropy was the main predictor for brain damage and was not one of the metrics used for edge of criticality assessment. Therefore, we might be able to disentangle the two phenomena by carefully investigating and choosing metrics and correcting terms. However, this is very speculative and requires a lot of further research!

## 5.3 Limitations

Our study has several limitations that need to be discussed in light of the current findings. This concerns the analysis's technical aspects and the interpretability of the found results.

#### Metric calculation in EEG

Regarding calculating metrics in the EEG, we encountered very different lengths of EEG segments for the various datasets and labels. To counteract this, as we knew that the length of an analyzed EEG segment could influence the metrics, we split each into 10-second epochs and averaged the results per segment over these epochs. However, the number of epochs we average over differs per segment. We are not sure what influence this could have precisely on the analysis, but it can not be excluded that statistical differences could arise due to this. We applied Bonferroni correction to the results to make our findings as rigid as possible. We reported significant levels with three different p-value thresholds. Both these measures aim to keep the statistical analysis rigid despite possible influences of unknown nature.

Concerning the prior, there is an assumption of stationarity underlying this method of epoching and averaging where we assume it makes sense to average the EEG over timeframes of 10-240 seconds. Stationarity in EEG is a widely discussed topic [11], and most EEG analyses assume some stationarity. However, this does not need to be accurate, especially for the longer analysis segments, such as in the coma cases.

Furthermore, different metric calculations may fail on a physiological signal because they encounter division by zero errors or other calculation problems, making the results unusable. We explicitly mentioned and discussed these cases throughout our analysis if they happened, but it still limits the comparability of the metrics and results.

Unexpected results or uncalculatable EEG segments may, in part, be caused by leftover artifacts in the data, especially in the coma data, where no manual revision of all EEGs was done due to time limitations and the amount of EEG data available. Other neurologists made a preselection to exclude artifacts in the coma data, but sometimes, not all artifacts could be removed, as the best possible epoch had to be chosen.

Last, regarding metrics calculation, a certain level of educated guessing is needed to determine an optimal set of preprocessing and metrics parameters for calculating the metrics. Even though we tried to find valid reasons for each parameter and conducted prior research on how these metrics could react to differences in acquisition parameters, there are no fixed golden standards established in the literature, and every paper has slight differences in how exactly data should be preprocessed and which parameters should be used within metrics. In our analysis, we tried to choose a simplistic and computationally efficient preprocessing and metric calculation method. This makes our results easily reproducible on the same or other datasets. It could, however, limit the comparability to other datasets like the ones from Toker [30] or Maschke [33] where additional steps like ICA were used in preprocessing.

**Interpretation of results**

Finally, to interpret results, one needs to be cautious in what the findings from our analysis mean. We can confidently say that we used similar patient populations regarding the altered states of consciousness as Toker et al. [30] did but could not fully replicate their results. However, saying that this would disprove their thesis or make the metrics invalid as a potential nonbehavioral metric would be speculative. We did not have the same patients or the same acquisition parameters. We showed clearly that many things need to be considered in how these metrics are calculated, and some things, like the 0-1 chaos test chain multiple analysis steps together with refiltering, downsampling, denoising, and then calculating the metric, which makes them very complex and can cause failure in the calculation.

## 5.4 Conclusion

Our research aimed to answer two questions: 1. Can we reproduce Toker's finding in EEG regarding the edge of criticality hypothesis? 2. Can our metrics aid in the prediction of outcomes after coma due to cardiac arrest? Additionally, we tested how our metrics react to acquisition and EEG preprocessing to evaluate these questions. Reporting these as accurately as possible is vital due to their influence on the metrics.

In an Ockhams's razor manner, we tried to design as simple as possible preprocessing and metric acquisition parameters to test the edge of criticality hypothesis in new and prior untested data with patients undergoing the same conditions or procedures as in the original paper, namely seizures and anesthesia with propofol. We extended these experiments with a large coma cohort with good and poor neurological outcomes. Regarding question 1, we could not reproduce all findings from Toker. In our analysis, we could find parallels to the results of Toker in seizure patients, which showed decreases in chaoticity and complexity values based on their EEG. We found contradictions in the case of anesthesized patients, which showed the anticipated decrease in complexity but no increase in chaoticity, but a decrease. This would falsify the hypothesis, at least on a data-driven basis. Coma patients with a good outcome showed similar behavior to anesthesized patients during the analysis.

In contrast, coma patients with a bad outcome showed large variability within metrics both spanning awake and altered states of consciousness of the other two sets. This could have multiple possible sources. The chaoticity and complexity metrics also capture dynamics unrelated to the conscious level. Complexity and chaoticity are also influenced by brain damage in ways that were not anticipated beforehand, showing dynamics similar to those of the awake state. Alternatively, we may miss insights into the conscious state, especially in a coma with bad outcomes. Which of these is the case can not be verified from EEG data alone.

Regarding question 2, we showed that our set of nonlinear metrics related to the entropy, fractality, chaoticity, and complexity of the data performed on par with the Cerebral Recover Index (CRI) in predicting the outcome of coma patients after cardiac arrest. A significant increase in the CRI was found when combining the CRI and our metrics. This indicates a potential connection of our metrics beyond the conscious state and to patients' neurological health. Permutation entropy was the strongest predictor of neurological outcome in our tested metrics.

We concluded that further research regarding the validity of the edge of criticality hypothesis will be needed. We believe that the field of understanding nonlinear dynamics will only increase in importance over the coming years, especially in understanding complex and high-level phenomena like consciousness, which can not be explained on a level of individual building blocks of the complex system of our brain.

## 5.5 Future outlook

Based on the findings and the experience gained in this study, some steps should be carried out in future research. These steps will help to resolve possible ambiguities, create a more rigid and fair comparison to the original papers we compared to, and deepen our general understanding of these nonlinear metrics.

#### Recreate experiments of Toker using his pipeline on new data

First, use the same data and try to recreate the exact experimental conditions used by Toker and Maschke and recalculate, especially the 0-1 chaos test, to see if this will yield the same results or if the differences between our findings persist. Similarly, using their data and applying our analysis would deepen our understanding of the metrics and their relation to EEG processing. This can give insight into the different outcomes in our study compared to existing literature and what caused them.

#### Extend research with more states of consciousness

Analyzing more different and diverse datasets will be needed in general. This is the first study that applied EEG analysis regarding the edge of criticality hypothesis in this scale. To make findings rigid, they should be expanded to include more different states of consciousness, including meditation or psychedelic substances, which we know can change the state of consciousness and be communicated by the individual. Sleep might also be an exciting direction in which to carry out research. These findings need to be replicated by individual research groups to create a clear and consensus-based picture regarding the edge of the criticality hypothesis. During that, a consensus on preprocessing steps and metric calculation should be made.

#### Investigation of non-linear EEG metrics in cardiac arrest

Second, one can further investigate our complexity-based markers' predictive qualities in coma after cardiac arrest. Even though we already showed increased prediction accuracy compared to CRI features alone, we did not optimize the process or do an in-depth comparison to the original CRI paper. There will be differences in signal processing and maybe even slight differences caused by our Python translation of the original code. Further investigation could shed more light on the potential of our metrics and how they could increase the predictive value of EEG metrics in cardiac arrest patients.

# References


[1] A. Lempel and J. Ziv, "On the Complexity of Finite Sequences," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 75–81, Jan. 1976, Conference Name: IEEE Transactions on Information Theory, ISSN: 1557-9654. DOI: 10.1109/TIT.1976.1055501. [Online]. Available: https://ieeexplore.ieee.org/document/1055501 (visited on 01/31/2024).

[2] B. B. Mandelbrot, *The Fractal Geometry of Nature*, en. Henry Holt and Company, 1983, ISBN: 978-0-7167-1186-5.

[3] C. G. Langton, "Computation at the edge of chaos: Phase transitions and emergent computation," *Physica D: Nonlinear Phenomena*, vol. 42, no. 1, pp. 12–37, Jun. 1990, ISSN: 0167-2789. DOI: 10.1016/0167-2789(90)90064-V. [Online]. Available: https://www.sciencedirect.com/science/article/pii/016727899090064V (visited on 02/13/2024).

[4] S. C. Tindall, "Level of Consciousness," eng, in *Clinical Methods: The History, Physical, and Laboratory Examinations*, H. K. Walker, W. D. Hall, and J. W. Hurst, Eds., 3rd, Boston: Butterworths, 1990, ISBN: 978-0-409-90077-4. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK380/ (visited on 02/22/2024).

[5] M. T. Rosenstein, J. J. Collins, and C. J. De Luca, "A practical method for calculating largest Lyapunov exponents from small data sets," en, *Physica D: Nonlinear Phenomena*, vol. 65, no. 1-2, pp. 117–134, May 1993, ISSN: 01672789. DOI: 10.1016/0167-2789(93)90009-P. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/016727899390009P (visited on 02/06/2024).

[6] T. Schreiber, "Extremely simple nonlinear noise-reduction method," *Physical Review E*, vol. 47, no. 4, pp. 2401–2404, Apr. 1993, Publisher: American Physical Society. DOI: 10.1103/PhysRevE.47.2401. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevE.47.2401 (visited on 09/17/2024).

[7] A. Goldberger, "Non-linear dynamics for clinicians: Chaos theory, fractals, and complexity at the bedside," en, *The Lancet*, vol. 347, no. 9011, pp. 1312–1314, May 1996, ISSN: 01406736. DOI: 10.1016/S0140-6736(96)90948-4. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0140673696909484 (visited on 01/31/2024).

[8] C. Bandt and B. Pompe, "Permutation Entropy: A Natural Complexity Measure for Time Series," *Physical Review Letters*, vol. 88, no. 17, p. 174 102, Apr. 2002, Publisher: American Physical Society. DOI: 10.1103/PhysRevLett.88.174102. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.88.174102 (visited on 02/06/2024).

[9] J. S. Richman, D. E. Lake, and J. R. Moorman, "Sample Entropy," in *Methods in Enzymology*, ser. Numerical Computer Methods, Part E, vol. 384, Academic Press, Jan. 2004, pp. 172–184. DOI: 10.1016/S0076-6879(04)84011-4. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0076687904840114 (visited on 09/09/2024).

[10] G. A. Gottwald and I. Melbourne, "Testing for chaos in deterministic systems with noise," *Physica D: Nonlinear Phenomena*, vol. 212, no. 1, pp. 100–110, Dec. 2005, ISSN: 0167-2789. DOI: 10.1016/j.physd.2005.09.011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167278905004045 (visited on 02/08/2024).

[11] W. Klonowski, "Everything you wanted to ask about EEG but were afraid to get the right answer," *Nonlinear Biomedical Physics*, vol. 3, p. 2, May 2009, ISSN: 1753-4631. DOI: 10.1186/1753-4631-3-2. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2698918/ (visited on 09/16/2024).

[12] B. S. Raghavendra and D. Narayana Dutt, “A note on fractal dimensions of biomedical waveforms,” *Computers in Biology and Medicine*, vol. 39, no. 11, pp. 1006–1012, Nov. 2009, ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2009.08.001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0010482509001486 (visited on 04/17/2024).

[13] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, no. 85, pp. 2825–2830, 2011, ISSN: 1533-7928. [Online]. Available: http://jmlr.org/papers/v12/pedregosa11a.html (visited on 08/26/2024).

[14] R. Hindriks and M. J. A. M. van Putten, “Meanfield modeling of propofol-induced changes in spontaneous EEG rhythms,” *NeuroImage*, vol. 60, no. 4, pp. 2323–2334, May 2012, ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2012.02.042. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1053811912002200 (visited on 02/15/2024).

[15] A. Casali, O. Gosseries, M. Rosanova, *et al.*, “A theoretically based index of consciousness independent of sensory processing and behavior,” English, *Science Translational Medicine*, vol. 5, no. 198, 2013, ISSN: 1946-6242. DOI: 10.1126/scitranslmed.3006294.

[16] A. Gramfort, M. Luessi, E. Larson, *et al.*, “MEG and EEG data analysis with MNE-Python,” English, *Frontiers in Neuroscience*, vol. 7, Dec. 2013, Publisher: Frontiers, ISSN: 1662-453X. DOI: 10.3389/fnins.2013.00267. [Online]. Available: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2013.00267/full (visited on 08/01/2024).

[17] V. R. Rao and D. H. Lowenstein, “Epilepsy,” *Current Biology*, vol. 25, no. 17, R742–R746, Aug. 2015, ISSN: 0960-9822. DOI: 10.1016/j.cub.2015.07.072. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0960982215009409 (visited on 09/11/2024).

[18] S. Berger, G. Schneider, E. F. Kochs, and D. Jordan, “Permutation Entropy: Too Complex a Measure for EEG Time Series?” en, *Entropy*, vol. 19, no. 12, p. 692, Dec. 2017, Number: 12 Publisher: Multidisciplinary Digital Publishing Institute, ISSN: 1099-4300. DOI: 10.3390/e19120692. [Online]. Available: https://www.mdpi.com/1099-4300/19/12/692 (visited on 04/15/2024).

[19] R. Bishop, “Chaos,” in *The Stanford Encyclopedia of Philosophy*, E. N. Zalta, Ed., Spring 2017, Metaphysics Research Lab, Stanford University, 2017. [Online]. Available: https://plato.stanford.edu/archives/spr2017/entries/chaos/ (visited on 02/20/2024).

[20] M. C. Tjepkema-Cloostermans, J. Hofmeijer, A. Beishuizen, *et al.*, “Cerebral Recovery Index: Reliable Help for Prediction of Neurologic Outcome After Cardiac Arrest,” en-US, *Critical Care Medicine*, vol. 45, no. 8, e789, Aug. 2017, ISSN: 0090-3493. DOI: 10.1097/CCM.0000000000002412. [Online]. Available: https://journals.lww.com/ccmjournal/fulltext/2017/08000/cerebral_recovery_index__reliable_help_for.34.aspx (visited on 08/26/2024).

[21] D. Toker, F. T. Sommer, and M. D’Esposito, “A simple method for detecting chaos in nature,” en, *Communications Biology*, vol. 3, no. 1, pp. 1–13, Jan. 2020, Number: 1 Publisher: Nature Publishing Group, ISSN: 2399-3642. DOI: 10.1038/s42003-019-0715-9. [Online]. Available: https://www.nature.com/articles/s42003-019-0715-9 (visited on 01/31/2024).

[22] D. S. Glass, X. Jin, and I. H. Riedel-Kruse, “Nonlinear delay differential equations and their application to modeling biological network motifs,” en, *Nature Communications*, vol. 12, no. 1, p. 1788, Mar. 2021, Number: 1 Publisher: Nature Publishing Group, ISSN: 2041-1723. DOI: 10.1038/s41467-021-21700-8. [Online]. Available: https://www.nature.com/articles/s41467-021-21700-8 (visited on 02/05/2024).

[23] D. Makowski, T. Pham, Z. J. Lau, *et al.*, "NeuroKit2: A Python toolbox for neurophysiological signal processing," en, *Behavior Research Methods*, vol. 53, no. 4, pp. 1689–1696, Aug. 2021, ISSN: 1554-3528. DOI: 10.3758/s13428-020-01516-y. [Online]. Available: https://doi.org/10.3758/s13428-020-01516-y (visited on 02/08/2024).

[24] J. M. Beggs, "Addressing skepticism of the critical brain hypothesis," eng, *Frontiers in Computational Neuroscience*, vol. 16, p. 703 865, 2022, ISSN: 1662-5188. DOI: 10.3389/fncom.2022.703865.

[25] J. Frohlich, J. N. Chiang, P. A. M. Mediano, *et al.*, "Neural complexity is a common denominator of human consciousness across diverse regimes of cortical dynamics," en, *Communications Biology*, vol. 5, no. 1, pp. 1–17, Dec. 2022, Publisher: Nature Publishing Group, ISSN: 2399-3642. DOI: 10.1038/s42003-022-04331-7. [Online]. Available: https://www.nature.com/articles/s42003-022-04331-7 (visited on 04/03/2024).

[26] H. Hakkoum, I. Abnane, and A. Idri, "Interpretability in the medical field: A systematic mapping and review study," *Applied Soft Computing*, vol. 117, p. 108 391, Mar. 2022, ISSN: 1568-4946. DOI: 10.1016/j.asoc.2021.108391. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1568494621011522 (visited on 09/11/2024).

[27] J. O'Byrne and K. Jerbi, "How critical is brain criticality?" English, *Trends in Neurosciences*, vol. 45, no. 11, pp. 820–837, Nov. 2022, Publisher: Elsevier, ISSN: 0166-2236, 1878-108X. DOI: 10.1016/j.tins.2022.08.007. [Online]. Available: https://www.cell.com/trends/neurosciences/abstract/S0166-2236(22)00164-3 (visited on 02/08/2024).

[28] T. L. Rogers, B. J. Johnson, and S. B. Munch, "Chaos is not rare in natural ecosystems," en, *Nature Ecology & Evolution*, vol. 6, no. 8, pp. 1105–1111, Aug. 2022, Number: 8 Publisher: Nature Publishing Group, ISSN: 2397-334X. DOI: 10.1038/s41559-022-01787-y. [Online]. Available: https://www.nature.com/articles/s41559-022-01787-y (visited on 02/05/2024).

[29] A. K. Seth and T. Bayne, "Theories of consciousness," en, *Nature Reviews Neuroscience*, vol. 23, no. 7, pp. 439–452, Jul. 2022, Number: 7 Publisher: Nature Publishing Group, ISSN: 1471-0048. DOI: 10.1038/s41583-022-00587-4. [Online]. Available: https://www.nature.com/articles/s41583-022-00587-4 (visited on 02/22/2024).

[30] D. Toker, I. Pappas, J. D. Lendner, *et al.*, "Consciousness is supported by near-critical slow cortical electrodynamics," *Proceedings of the National Academy of Sciences*, vol. 119, no. 7, e2024455119, Feb. 2022, Publisher: Proceedings of the National Academy of Sciences. DOI: 10.1073/pnas.2024455119. [Online]. Available: https://www.pnas.org/doi/10.1073/pnas.2024455119 (visited on 01/31/2024).

[31] V. Barone, M. C. Piastra, J. P. Van Dijk, G. H. Visser, M. H. Debeij-van Hall, and M. J. Van Putten, "Neurophysiological signatures reflect differences in visual attention during absence seizures," en, *Clinical Neurophysiology*, vol. 152, pp. 34–42, Aug. 2023, ISSN: 13882457. DOI: 10.1016/j.clinph.2023.05.007. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1388245723006260 (visited on 01/26/2024).

[32] A. Delorme, "EEG is better left alone," en, *Scientific Reports*, vol. 13, no. 1, p. 2372, Feb. 2023, Number: 1 Publisher: Nature Publishing Group, ISSN: 2045-2322. DOI: 10.1038/s41598-023-27528-0. [Online]. Available: https://www.nature.com/articles/s41598-023-27528-0 (visited on 02/18/2024).

[33] C. Maschke, J. O'Byrne, M. A. Colombo, *et al.*, "Criticality of resting-state EEG predicts perturbational complexity and level of consciousness during anesthesia," eng, *bioRxiv: The Preprint Server for Biology*, p. 2023.10.26.564247, Oct. 2023. DOI: 10.1101/2023.10.26.564247.

[34] D. Toker, E. Müller, H. Miyamoto, *et al.*, *Criticality supports cross-frequency cortical-thalamic information transfer during conscious states*, en, Pages: 2023.02.22.529544 Section: New Results, Feb. 2023. DOI: 10.1101/2023.02.22.529544. [Online]. Available: https://www.biorxiv.org/content/10.1101/2023.02.22.529544v1 (visited on 01/29/2024).

[35] *Are There Really as Many Neurons in the Human Brain as Stars in the Milky Way? — Brain Metrics — Learn Science at Scitable*, en, Cg_cat: Brain Metrics Cg_topic: Are There Really as Many Neurons in the Human Brain as Stars in the Milky Way? [Online]. Available: https://www.nature.com/scitable/blog/brain-metrics/are_there_really_as_many/ (visited on 09/16/2024).

[36] J. P. Crutchfield and K. Young, "Computation at the Onset of Chaos," en,

[37] M. Kale and F. B. Butar, "Fractal Analysis of Time Series and Distribution Properties of Hurst Exponent," en,

# A Hyperparameter exploration

At the beginning of our research, we investigated the sensitivity of the acquisition parameters using set A. This is not done systematically, as its purpose lies in getting familiar with the metrics before a structure analysis. It was done to understand the sensitivity of parameters towards epoch length, sampling frequency, different filter cutoffs, and whether there are regional differences within the brain. Lastly, we also tested various methods of estimating optimal time delays and embedding dimensions, as this can be done for some features, especially time embedding-dependent ones like the LLE. The different tests and results are described below.

Starting with the montage, we found significant differences in absolute values between a common average and bipolar double banana montage.

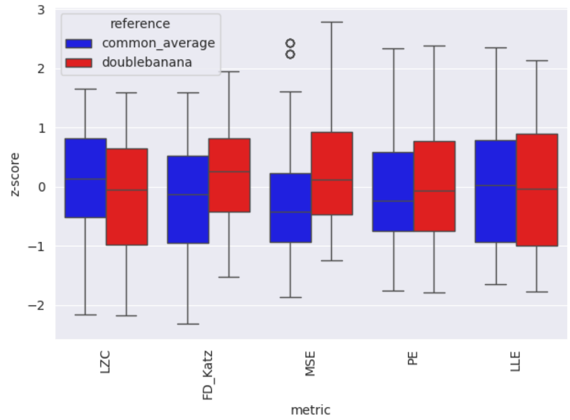



**Figure 6:** Comparison of montages using differences in z-scores. Montages show differences in absolute values between metrics.

For epoch length in seconds, we found a significant negative correlation of time and metric value in multiscale sample entropy and Hurst exponent using person correlations. Only in anesthesized states is a significant positive correlation between Multiscale permutation entropy and epoch length. Additionally, we found indications of a positive correlation with time for the fractal dimension using Katz's estimation, but it was not significant. For this analysis, we used the data's natural differences in annotation length. In our final analysis, we decided to unify the length to the shortest annotations used.

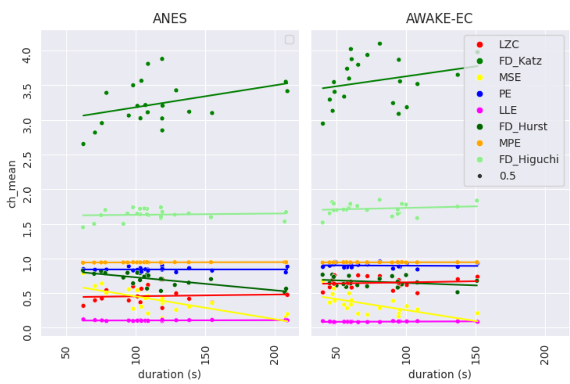



**Figure 7:** Influence of sample length on metrics. Significant decrease in metric for MSE and Hurst exponent. Significant increase in anesthesia for Multiscale permutation entropy.

We always used a 0.5 Hz highpass filter for the different frequency bands with a 4, 25, and 35 Hz lowpass filter, respectively. Lower cutoff frequencies reduce the variance in the metrics. At 4 Hz, we found a decrease in absolute values, except for the chaoticity estimation using the Largest Lyapunov exponents. In addition to lower variance and absolute values, the separability of classes also decreased, as can be seen by the minor separation in the boxplots for the 4 Hz lowpass filter. We also analyzed the typical EEG bands, including delta, theta, alpha, and low beta, but we will omit the results here. They all showed differences in absolute values and the separability of classes, but no advantage was found in using the individual bands for further analysis.

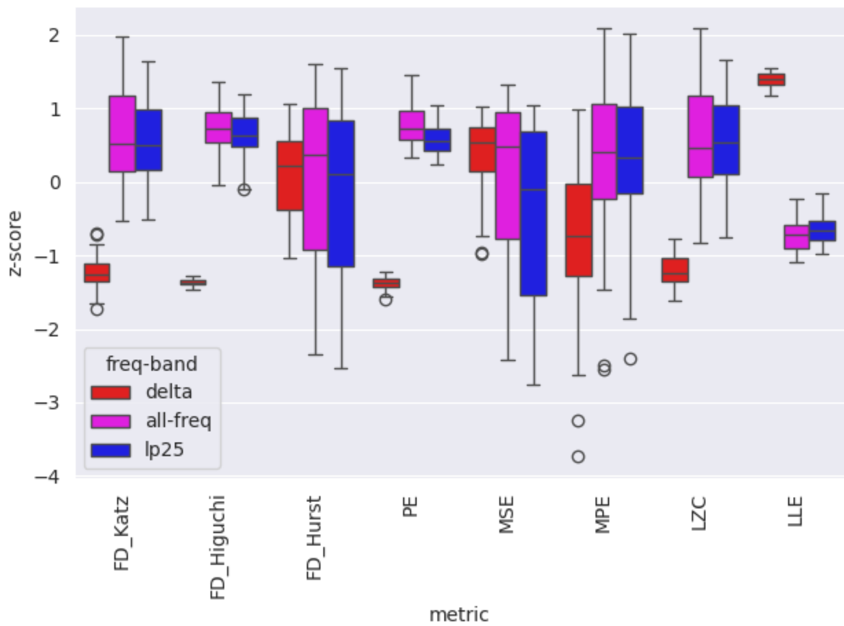



**Figure 8:** Different z-scores between metrics for a direct comparison of filter bands influence on metrics.

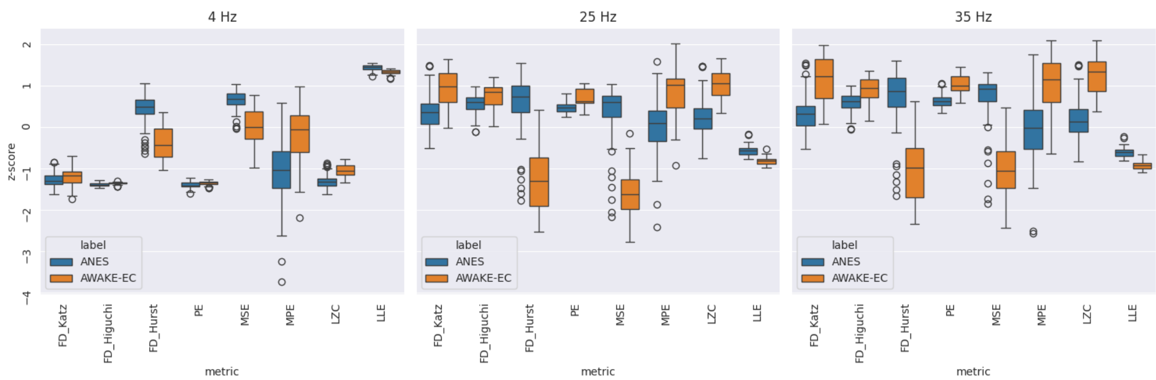



**Figure 9:** Comparison of the Anesthesized and Awake state based on lowpass frequency to show differences in the separability of states.

We compared 50 and 100 Hz sampling frequencies. At 50 Hz, we found better separability of classes only in the fractal dimension using Higuchi's method. At 100 Hz, we found better separability in classes for LZC and LLE. It should be mentioned here that, in hindsight, 50Hz resampling was not a good choice as the Nyquist frequency and the filter cutoff frequency now overlap, which can cause aliasing.

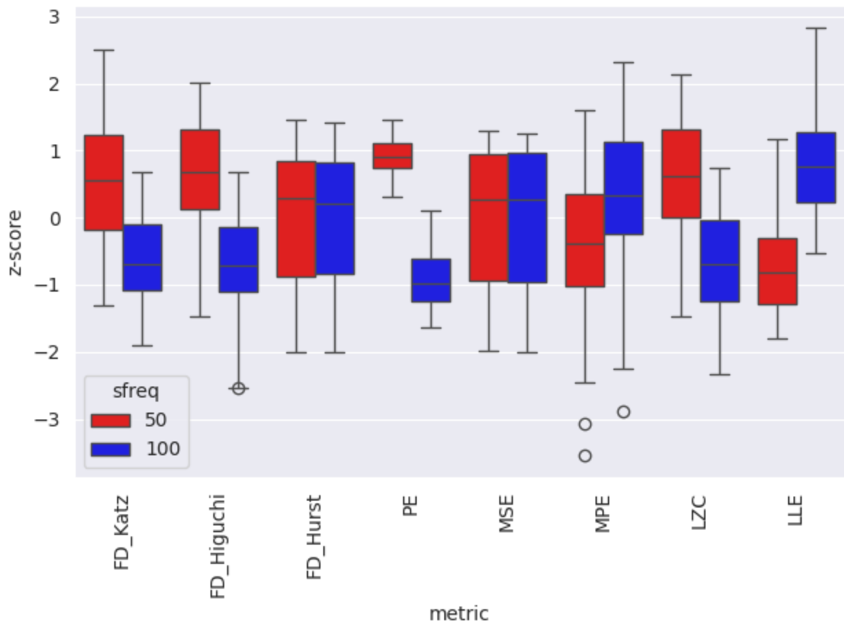



**Figure 10:** Comparison of z-scores between different sampling frequencies

Regarding the different channels, we found apparent local differences in metrics. For a selection of metrics, we display the change in local difference and change between labels in fig 11.

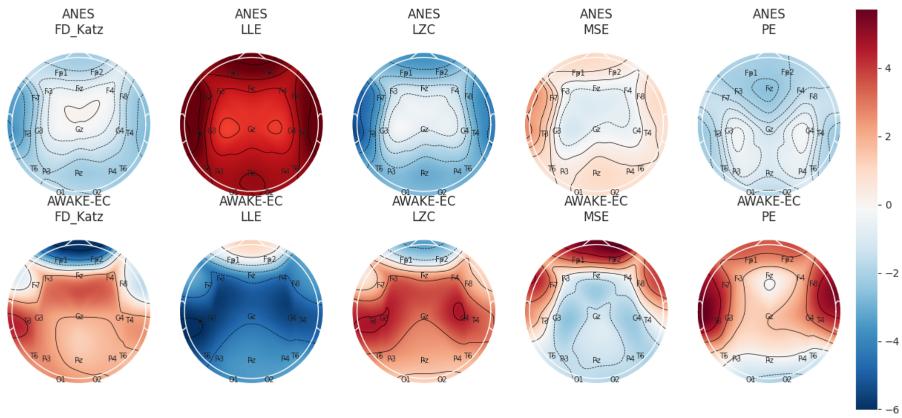



**Figure 11:** Visualization of the z-score for a selection of metrics for awake baseline (bottom) and anesthesized patients (top). Means and variances were calculated per metric across both labels.

To calculate metrics dependent on time delay embeddings or using time delays in general to estimate signal characteristics, such as LLEs or PE, neurokit2 provides different algorithms for calculating the 'optimal' parameter for the given time series. To check how ambiguous this parameter is, we tested the agreement of various methods in choosing this optimal embedding dimension and time delay. We found vastly different results across both time delays and embedding dimensions, with a clear minority of time series in which they agreed. We define an agreement by the difference in metrics being 0. In figure 12, one can see the difference in embedding dimensions between the three methods used. In figure 13, one can see the difference in time delay (in samples) for the three estimation methods. The names used for labeling correspond to the names in neurokit2.

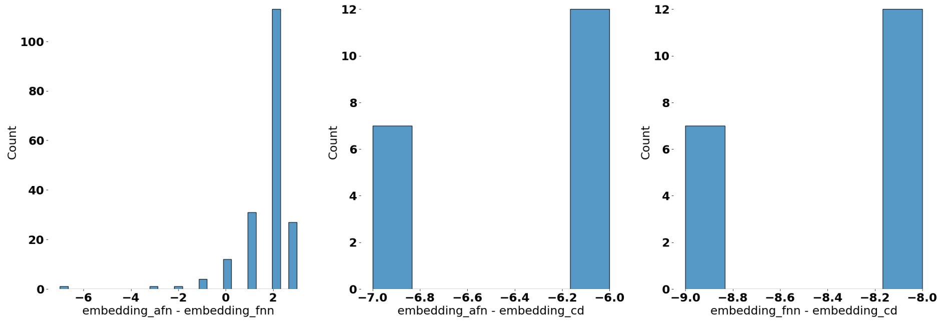



**Figure 12:** Differences in embedding dimension estimations using three different methods provided by neurokit2. Not all signals were processable for this purpose, which explains the large difference in absolute counts between the first and the second two graphs. There is very little agreement between methods on the optimal embedding dimension for a signal. Differences were only calculated on the same signals, individual EEG channels.

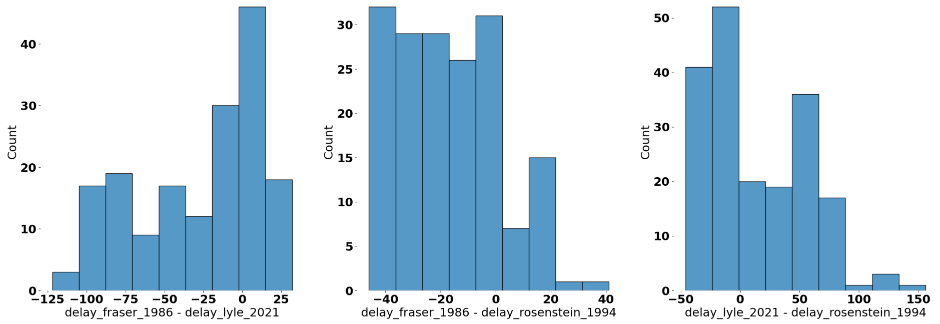



**Figure 13:** Differences in embedding delay estimations using three different methods provided by neurokit2. There is little agreement between methods on the optimal embedding delay for a signal. Differences were only calculated on the same signals, individual EEG channels.

Summarizing, one can say that these nonlinear metrics depend on the parameters used to process the signal. Our analysis is not structured and encompassing enough to decide what optimal parameters would be, but further investigation of these differences should be encouraged. Understanding the relations between acquisition parameters and metrics can aid in understanding what the metrics represent, especially concerning our physiological signals. It also is a clear argument for why it is important to accurately report on all parameters used in signal processing to make studies and results comparable. Understanding regional differences in metrics can also help connect them to current ideas from the literature on how consciousness and neural signals are linked. They often include assumptions that focus on specific brain regions over others [29].

## B Correlation between metrics

We check how different fractality, complexity, entropy, and chaoticity metrics correlate. We compare our findings to the paper of Maschke et al. [33] to see if their findings on the correlation of complexity, fractality, and chaoticity can be replicated. They found the fractal dimension to be significantly positively correlated to complexity. There is a significant negative correlation between chaoticity and complexity. A significant but weak positive correlation of Multiscale Sample Entropy and Largest Lyapunov exponents.

We calculate the Pearson correlation between all metrics to test how the metrics correlate and if the findings from Maschke et al. [33] can be reproduced.

Figure 14 depicts the correlation between the different metrics. We find a significant positive correlation between the fractal dimension estimation in both Katz and Higuchi's methods with LZC and each other. A strong negative correlation exists between LZC and the LLE. We found no correlation between MSE and the LLE. We also found no correlation between the single scale and multiscale permutation entropy.

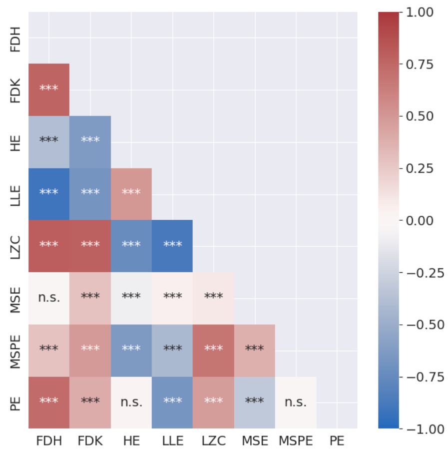



**Figure 14:** Correlation results of all the different metrics used in the analysis. Significance values of the person correlation after Bonferroni correction with a factor of 28: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, n.s. means no significance was found.

Lemepl Ziv's complexity and the fractal dimension using Higuchis and Katz's methods showed a significant positive correlation. Fractality is one of the primary mechanisms by which complex structures can emerge from simple instructions and are found widely in nature [2]. For a more thorough understanding of how complexity and fractality are linked, it is not sufficient to show correlations in data. Modeling or in vitro experiments would be needed to investigate what might cause these changes and how fractal properties could be included further in the investigation of brain signal emergence.

## C Fractal Dimension

In this section, the focus is on metrics related to the fractal dimension of the signals. First, two fractal dimension estimation methods, the Higuchis and Katz methods, will be compared. Then, the estimated fractal dimension is related to the Hurst exponent, which is mathematically linearly related to the fractal dimension of a signal.

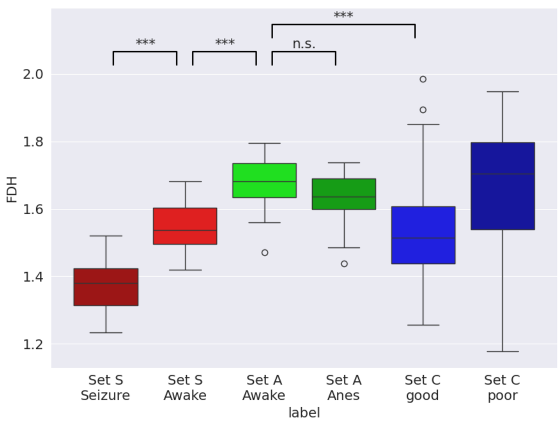
## C.1 Katz vs Higuchi

We test two different methods for estimating the fractal dimension, previously compared by Raghavendra et al. [12]. They found differences between the two methods in simulated data and EEG measurements during sleep. Katz's method is biased towards higher dimensions, which surpass the theoretical maximum dimension of 2 for a one-dimensional time series. Both had averages within the theoretical limit. Interclass differences regarding the investigated sleep stages in their EEG were more prominent using the Higuchis method.

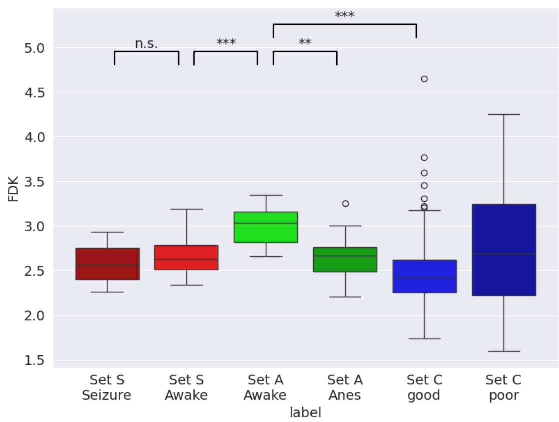
The fractal dimension is an alternative way of describing the symmetries and properties of geometrical structures and is built around the principles of symmetries and roughness or jaggedness [37]. Initially, the term fractal was coined by Benoit Mandelbrot to describe self-similar, infinitely complex structures often observed in nature. Opposite to Euclidian geometry, fractal dimensions are not restricted to the Natural numbers but give dimensionality as a fraction [2]. The fractal dimension is between one and two for a one-dimensional time series. One would refer to a perfectly continuous and smooth signal. Two would describe a wholly random and unstructured signal. 1.5 is the dimension of a random walk. An alternative description often used is to define a higher fractal dimension with a more room-filling quality where a fractal dimension closer to 2 becomes almost as room-filling as a plane [37]. Additionally, the fractal dimension is not restricted to perfectly self-similar structures. There are stochastic fractals where there is only a certain level of self-similarity across different scales of the signal [7].

We use two different methods to estimate the fractal dimension of our time series. Similar to complexity and chaoticity, we can only estimate the dimensionality of a measured signal. Both methods are implemented in neurokit2 [23]. The first one is Katz's method for estimation due to its supposed robustness towards noise [33]. It is, however, known to be better at estimating relative differences in fractal dimension than the absolute value of it [12]. The other method we use is Higuchis method with a k of 10. It is sometimes said to be less robust towards noise than Katz's method but better at estimating the absolute fractal dimension and more computationally efficient, especially for longer time series [12].

Figure 15 compares the two fractal dimension estimators. It shows that the relative differences, as shown in the correlation analysis, are very similar, with differences in significance level between classes. The absolute values differ with higher values in the Katz method.




*(a) Fractal Dimension Higuchi*




*(b) Fractal Dimension Katz*

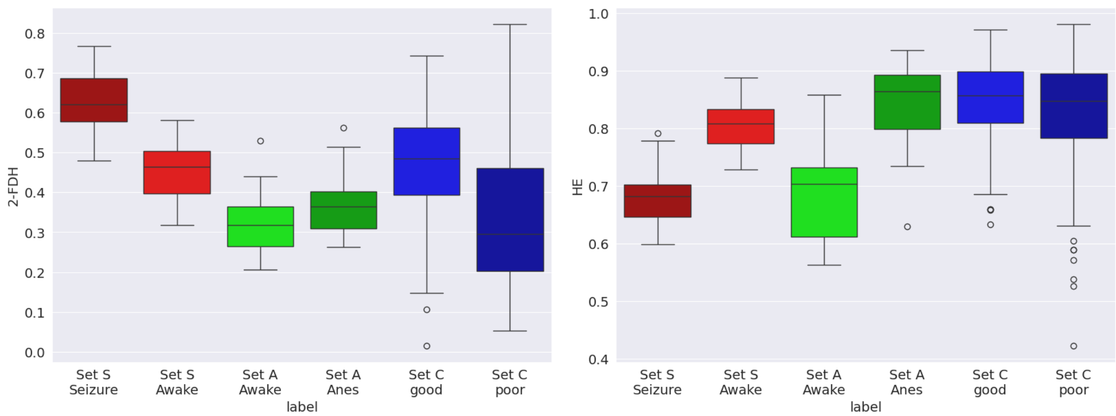
**Figure 15:** Comparison of Katz and Higuchi's methods to estimate fractal dimensions in the EEG data. Relative differences within Katz and Higuchi's estimation method show the same trends at different absolute values.

In our analysis, we found that, in relative terms, both Katz's and Higuchi's methods seem to capture similar differences between the different states. Katz's method shows higher absolute values, surpassing the theoretical maximum for the fractal dimension of a time series. Higuchi's method stays bounded between the theoretically feasible values of one and two. In their investigation of EEG data, Raghavendra et al., [12] found bounded average values in both methods. They also found EEG segments using Katz method with dimensionality surpassing the theoretical bound of two. Their maximums were more bounded than ours. One possible reason for the differences is that we downsampled our signals to 100 Hz. This is lower than in their analysis. They also showed a significant negative correlation between sampling frequency and fractal dimension estimation in the Katz method. This is a possible cause of the overestimation of fractal dimension in our EEG data as Raghavendra also showed an increase in dimensionality with lower sampling frequencies. It may not be the only difference that causes the discrepancy in dimensionality in the Katz method. The Higuchi method seems more stable in this regard and shows good class separation.

## C.2 Hurst exponent

Also belonging to the family of fractal dimension estimators is the Hurst exponent. It measures long-range autocorrelation in the data and is a metric commonly used to predict how complex time series like stock markets will develop knowing their history. It is theoretically related to the fractal dimension by the formula H = 2-D, where H is the Hurst exponent, and D is the fractal dimension of the time series. A Hurst exponent of $< 0.5$ means the series is anti-persistent: A current trend is unlikely to continue. A value of 0.5 refers to a random walk process. A Hurst exponent of $> 0.5$ indicates trend stability where a signal is more likely to follow its current trend [37]. It should be noted that these properties do not describe fast fluctuations of the signal but the global trajectories around which the signal fluctuates. Here again, the implementation of neurokit2 [23] is used to estimate the Hurst exponent in the data.

We used the Higuchi fractal dimension estimation and calculated the Hurst exponent based on the above formula. We will compare it to the Hurst exponent estimated directly from the data. We omitted the Katz method from this analysis as the fractal dimension was not in the limit between one and two. The comparison is depicted in figure 16.




*(a) Hurst exponent calculated from Higuchis fractal dimension*

*(b) Hurst exponent estimated from data directly*

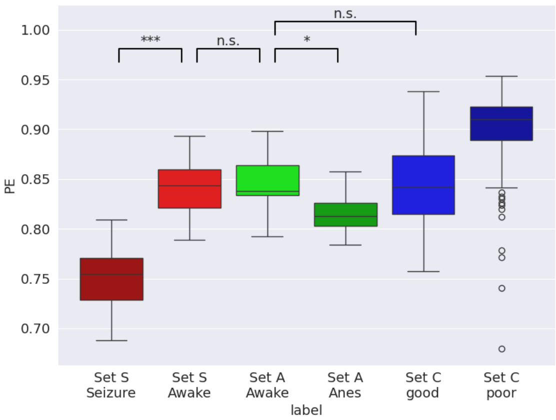
**Figure 16:** Comparison of the estimated Hurst exponent from data directly and by using the theoretical connection between fractal dimension and Hurst exponent. There is a significant difference between the two estimations

We did not find a connection between the two. Neither the absolute values nor the relative differences within the datasets match. This could be because both metrics are influenced by noise. It would be interesting to see if the relationship could be found in a system with a known fractal dimension.

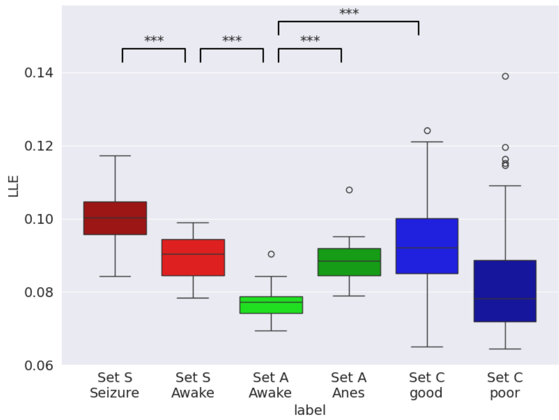
# D Permutation entropy as chaoticity estimator

Permutation entropy is a simple entropy measure. It is not used for our edge of criticality measure but was used prior in EEG studies [18]. It is straightforward and fast to compute. We were primarily interested in its role as a possible secondary chaoticity estimator similar to MSE. In their original paper on PE, Bandt et al. [8] show similar behavior of permutation entropy and positive LLEs on an audio recording.

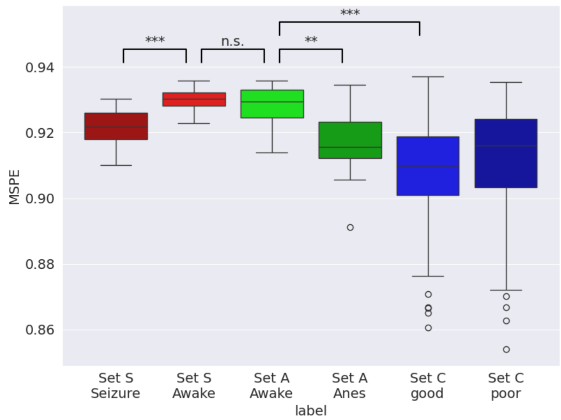
Like LZC, permutation entropy operates based on a vocabulary of repeating patterns in the data. However, computation is different as the vocabulary consists of a set of limited substrings, permutations of a fixed series. The six permutations of a time series of 3 subsequent points are used by default. It calculates the rate of occurrence for each substring and uses an average of the surprise for each as its final metric. It claims to be robust to noise and signal length. We used the implementation from neurokit2 [23] with an embedding delay of 1 and a series length of 3. We used the normalized version of permutation entropy to confine the output to 0-1 and make it independent of the size of the sampled signal.




(a) *Permutation entropy*




(b) *Largest Lyapunov exponent*




(c) *Multiscale Permutation entropy*

**Figure 17:** Comparison of permutation entropy (PE), multiscale permutation entropy (MSPE), and Largest Lyapunov exponents (LLE). Testing if the LLE and PE behave similarly as they did in Bandt et al.'s paper [8]. No similar behavior of LLE and PE could be found, neither in the single nor the multiscale variant.

As depicted in figure 17, PE and LLE did not show similar behavior in the EEG data. Our correlation analysis also confirmed this, where they showed a significant negative correlation B. Next to comparing the PE to LLE, we also compared the single and multiscale version of it. Our correlation analysis showed no correlation between the two metrics despite using the same entropy estimation method. This fact, along with the results from the hyperparameter tests, shows that permutation entropy is a sampling frequency-sensitive metric. This might have to do with the number of monotonous segments increasing and decreasing between local maxima and minima, changing with different sampling frequencies. The overall shape of the signal should not be influenced as long as the sampling frequency is higher than twice the filter cutoff containing all relevant frequencies of the signal. This would lead to an increase in monotonous sections compared to local maxima and minima. This hypothesis would need rigorous mathematical testing in artificial and real data.

The most striking feature of permutation entropy in our analysis was its strong predictive value in the neurological outcome of coma patients after cardiac arrest. Permutation entropy showed in our study of features the best predictive quality, even surpassing metrics from the Cerebral Recover Index [20].

# E Supplementary material for neurological outcome prediction

The following figures were used to create table 2 on the importance of features in the results section. They show the average entropy decrease in a tree per forest and the standard deviation of that feature's importance. Each subsection of the figures represents one forest trained per 5-fold cross-validation split.

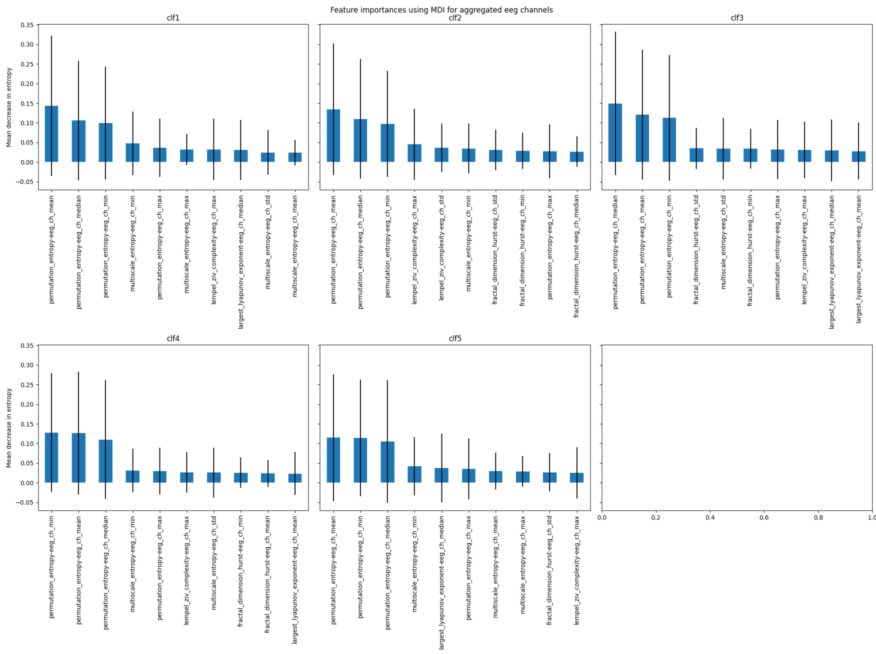



**Figure 18:** Feature importance for the five forests using complexity related features. Blue graphs show the average decrease of entropy in a tree. The white line indicates the standard deviation across the 100 trees of a forest. The forests are individually represented in the five figures.

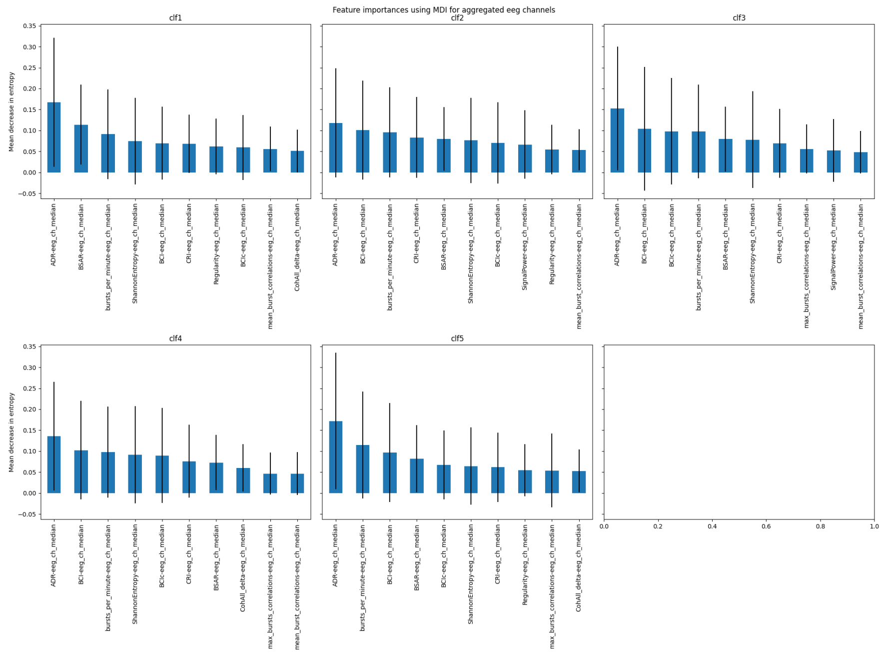



**Figure 19:** Feature importance for the five forests using the CRI features. Blue graphs show the average decrease of entropy in a tree. The white line indicates the standard deviation across the 100 trees of a forest. The forests are individually represented in the five figures.

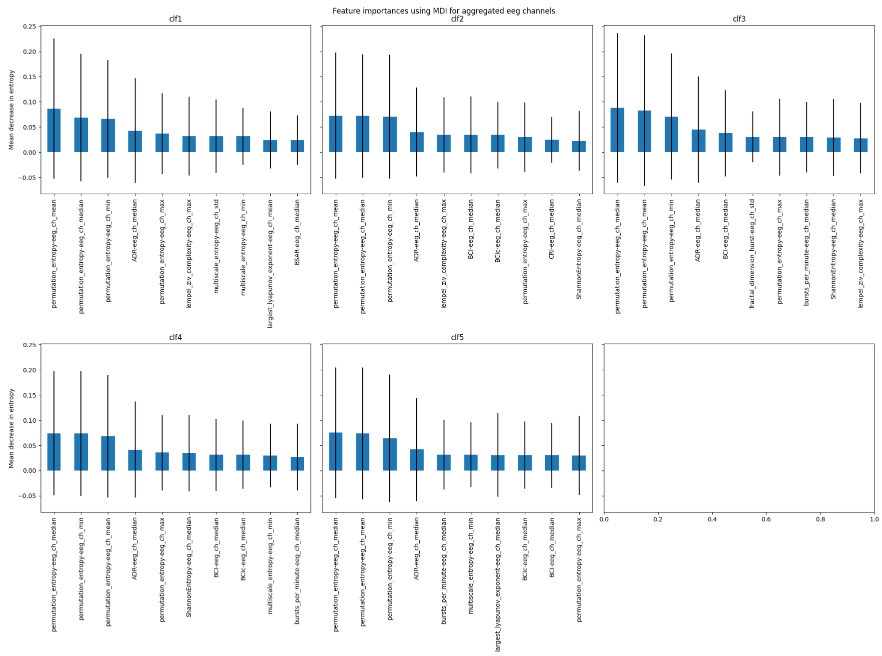



**Figure 20:** Feature importance for the five forests using combined features. Blue graphs show the average decrease of entropy in a tree. The white line indicates the standard deviation across the 100 trees of a forest. The forests are individually represented in the five figures.

# F EEG analysis

## EEG processing pipeline

I developed my processing pipeline to process many EEGs from different sources. The main parts of the code are in my Github repo. The pipeline easily adapts to almost any metrics computed on a one-dimensional time series. Every channel of an EEG is seen as an individual 1d time series in that regard. It consists of three main components. A config file in the form of a yaml document. A Python file where the metrics that should be used are listed. Lastly, two pipeline files combine the config file, the metric file, and the EEGs. One version uses EEG files on local storage in BIDS format. The other one uses a CSV file from which the file paths of the EEGs can be inferred.

In big data projects, a clearly defined standard must be used for storing the EEGs and the results computed from them. For this, we use an internationally common standard called BIDS. It allows us to store the EEGs and metrics calculated from them in an organized manner based on a hierarchy of datasets and subjects.

As many of the experiments are similar but differ in the EEG processing parameters, I use a config file for most settings. A sample config file with an explanation of the individual parameters is included in the complementary codebase. Some settings include where the EEGs are stored, filter frequencies, sampling frequencies, multiprocessing parameters, which metric set should be used, and where the results are stored. This allows maximum reusability between different experiments and implicitly keeps track of essential hyperparameters when combined with a sensible naming scheme.

An extra Python file contains lists of metrics. Three lists are combined. The first has the functions, the second is a dictionary of optional parameters, and the third is the names for the functions as they are saved in the metrics CSV files computed for later analysis.

The process's main file is the pipeline file. It uses parameters from the config and metric files to load and process the EEG, compute all metrics per channel, and save the results to the desired path in a CSV.

## Metrics analysis

After computation from EEG, the metrics are stored as small CSV files in a folder specific to the subject. The file path is assigned in the config file. I load the individual CSVs into a big data frame to analyze the metrics. From there, I can do plots, statistical analysis, and aggregation as I want. For convenience, I save the data frames I frequently need into a SQLite database so I can reuse them without recreating them every time.