Sleep Analysis Using Entropy and Fractal Features from Simultaneous Electroencephalography-Functional Magnetic Resonance Imaging Data

Hunor Bartalis-Szélyes

SNR: 2070381

Bachelor Thesis in Psychology

Department of Cognitive Neuropsychology, Tilburg University

Supervisor: Dr. Wouter De Baene

Second Reader: Dr. Gijs Holleman

June 28th, 2024

Abstract

This study investigated the cross-imaging-modality correlations of entropy, as well as fractal features in simultaneous electroencephalography-functional magnetic resonance imaging (EEG-fMRI) data and their effectiveness in distinguishing and classifying sleep stages. Using open-access simultaneous EEG-fMRI data at rest and sleep from healthy individuals (age: 22.1 ± 3.2 years; male/female: 17/16), correlational analyses were conducted on these features from complete and sleep stage-specific EEG and fMRI time series. Furthermore, sleep stage-specific features were used in ANOVAs and machine learning models to assess their ability to distinguish and classify sleep stages. The correlational analyses identified limited and region-specific relationships for the same entropy and fractal features between EEG channels and fMRI brain regions, including the frontoparietal EEG channels and frontoparietal attentional and default mode networks. ANOVAs revealed significant differences in entropy and fractal features across sleep stages, when averaged over all EEG channels or fMRI brain regions. Entropy features were higher in wakefulness relative to other sleep stages, while most fractal features showed the opposite trend. The machine learning models trained on EEG-based features achieved 76.1% accuracy, while models trained on fMRI-based or combined features from both imaging methods reached 75.4% accuracy. Thus, combining features from EEG and fMRI did not improve performance. Despite constraints, including imbalanced and noisy data, this study presented new approaches and insights into entropy and fractal features of simultaneous EEG-fMRI data during sleep, which might have implications for understanding sleep disorders. Future research should examine these features in larger, more balanced datasets, including patient data from sleep disorders.

Keywords: sleep, entropy, fractality, simultaneous EEG-fMRI, machine learning

Sleep Analysis Using Entropy and Fractal Features from Simultaneous Electroencephalography-Functional Magnetic Resonance Imaging Data

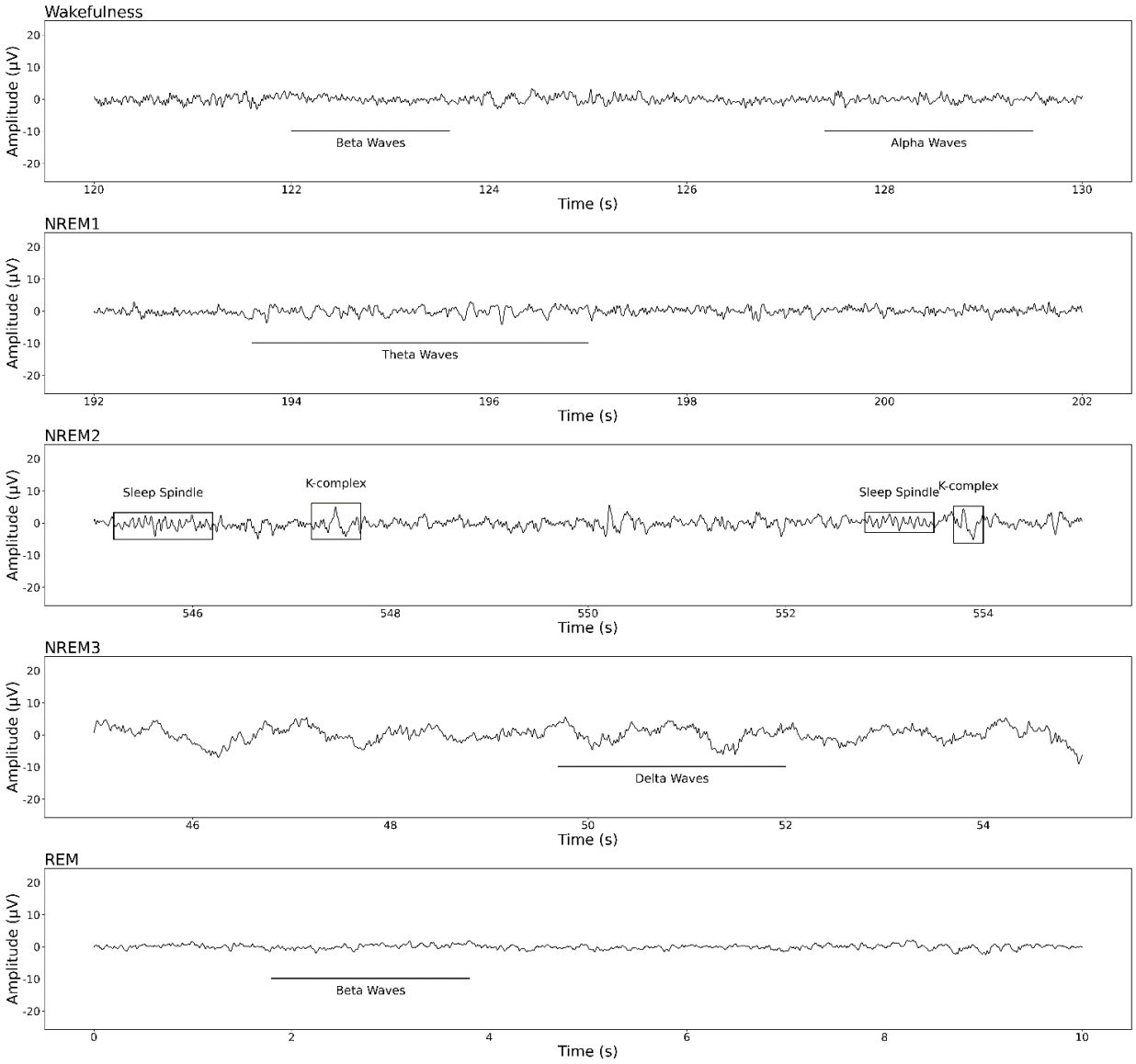
Sleep is a vital physiological process defined as a homeostatically regulated, temporary state, characterized by relative immobility, significantly diminished sensory responsiveness, lowered or lack of conscious awareness, as well as rapid reversibility, distinguishing it from states such as coma or anesthesia (Siegel, 2005; Tagliazucchi, Behrens, et al., 2013; Tononi et al., 2016). Although sleep seems to be vital in mammalian species, including humans, the nature and function of sleep remains enigmatic and largely unmapped (Rechtschaffen, 1998; Siegel, 2008). Besides, this issue becomes even more relevant when we also consider that humans normally devote roughly a third of their lifespan to sleeping, yet there are at least 83 different sleep disorders, some of which are estimated to affect a third of the Netherlands' population and approximately one billion people worldwide (Benjafield et al., 2019; Kerkhof, 2017; Sateia, 2014). Therefore, understanding sleep and its features has been crucial since the emergence of empirical sleep research in the 19th and 20th centuries, marked by the discovery of rapid eye movement (REM) and non-rapid eye movement (N-REM) sleep stages, as well as the advent of brain imaging technologies, such as the electroencephalogram (Aserinsky & Kleitman, 1953; Berger, 1929; Dement & Kleitman, 1957).

Electroencephalography (EEG) is a non-invasive brain imaging technique that uses electrodes placed on the scalp to track voltage fluctuations generated by the flow of ionic currents within neurons, thus enabling researchers to study the brain's electrical activity (Buzsáki et al., 2012). By providing a window into the complex electrical dynamics of the brain, EEG has contributed significantly to several groundbreaking discoveries in sleep research, including the identification of the aforementioned, distinct sleep stages (Aserinsky & Kleitman, 1953; Dement

& Kleitman, 1957; Loomis et al., 1937). This is important because understanding the distinct features of these sleep stages in neurophysiological recordings, such as EEG, may be informative not only about the functions of the processes underlying sleep but also about the presence of sleep disorders. Therefore, in order to reliably classify and interpret these sleep stages based on their characteristics in EEG data, the formulation of standardized criteria was necessary, such as the frameworks developed by Rechtschaffen and Kales (1968) and the American Academy of Sleep Medicine (AASM) (Iber et al., 2007, Moser et al., 2009). The more contemporary guidelines of AASM define sleep as comprising of cycles of five sleep stages, each with their own characteristic EEG features, as can be seen in Figure 1 (Iber et al., 2007, Moser et al., 2009). The sleep cycle starts with wakefulness, which is characterized by mostly beta or alpha waves (Cantero et al., 2002; Cirelli & Tononi, 2015; Stiller & Postolache, 2005). Then, N-REM sleep follows, with light N-REM sleep (N1) marked by theta waves, as well as deeper N-REM sleep (N2), containing distinctive waveforms, such as sleep spindles and K complexes (Gonzalez et al., 2018; Loomis et al., 1935, 1937). After that, the deepest N-REM sleep is reached (N3), which can be inferred based on low-frequency delta waves in the EEG signal (Stiller & Postolache, 2005). Finally, REM is the stage most commonly associated with vivid dreaming, as well as with beta and theta waves (Stiller & Postolache, 2005; Vijayan et al., 2017). Thus, following the criteria of the AASM, sleep researchers should be able to reliably classify sleep stages. However, despite the establishment of these classification standards, the manual assessment of sleep stages based on EEG is a strenuous and time-consuming process, which is also prone to subjectivity (Gaiduk et al., 2023; Rosenberg & Van Hout, 2013). As a result, in recent decades there has been a push towards developing faster, more objective, and automated methods for sleep stage analysis and classification based on EEG signal characteristics to mitigate these issues (Agarwal

Figure 1

Sleep Stages and Their Characteristic EEG Features



Note. Visualization of 10-second segments from the dataset of Gu et al. (2023) used in this study after preprocessing to demonstrate EEG characteristics of the five sleep stages, according to the AASM guidelines. As intervals of REM sleep stage were not available in the selected dataset, a REM segment has been taken from the open-access Sleep-EDF Database (Goldberger et al., 2000; Kemp et al., 2000). Sleep spindles and K-complexes in N-REM 2 are also highlighted.

& Gotman, 2001). Several review articles, including those by Fiorillo et al. (2019), Gaiduk et al. (2023), Phan and Mikkelsen (2022), as well as Lambert and Peter-Derex (2023), evaluated previous studies on EEG feature-based machine learning and deep learning algorithms used for sleep staging, highlighting that such approaches demonstrate considerable promise, some even achieving more than 90% accuracy and a Cohen's kappa of 0.8 or above, which is comparable with the performance and inter-rater reliability of human scorers (Danker-Hopfe et al., 2009). Nevertheless, a persistent challenge that remains is the low spatial resolution of EEG. While EEG excels in temporal resolution, with near-instantaneous view of the brain's electrical activity, its spatial resolution is severely limited, making it difficult for researchers to precisely locate the source of neural activity, specifically considering deeper brain structures (Betta et al., 2021; Murphy et al., 2009). This, in turn, leads to the potential loss of useful spatial information that could otherwise increase performance in the classification process and reveal key aspects of sleep and its stages (Lambert & Peter-Derex, 2023). To overcome this limitation, new brain imaging techniques were necessary to explore neural activity with better spatial accuracy.

One such advancement was functional magnetic resonance imaging (fMRI), which is a non-invasive technique designed to observe hemodynamic changes associated with brain activity, either during specific tasks or at rest (Bandettini et al., 1992; Kwong et al., 1992; Ogawa et al., 1990). Unlike EEG, fMRI shows the Blood Oxygen Level Dependent (BOLD) signal, which indicates changes in deoxyhemoglobin concentration via magnetic signals from hydrogen nuclei in water molecules, assumed to be induced by shifts in neural metabolism (Glover, 2011). As a result, researchers are now able to explore the neural underpinnings of sleep stages with high spatial fidelity and even develop automatic sleep staging methods using data obtained with fMRI (Logothetis, 2008; Tagliazucchi et al., 2012; Tagliazucchi & Laufs, 2014). However, a

considerable limitation of fMRI is its relatively poor temporal resolution (Boynton et al., 1996; Chen et al., 2023; Glover, 2011; A. Taylor et al., 2018).

Consequently, while EEG directly captures neural electrical activity, offering high temporal resolution, fMRI assesses neural activity indirectly through hemodynamic responses and provides high spatial resolution. Not so surprisingly, the complementary strengths of EEG and fMRI have become increasingly apparent over the past thirty years, leading to a more nuanced understanding of brain function by merging EEG's high temporal resolution and fMRI's high spatial resolution, culminating in what is called simultaneous EEG-fMRI (Sturzbecher & De Araújo, 2012; Warach et al., 1996). While both EEG and fMRI have long been used on their own to explore the features of sleep and its stages, the idea of combining them for this purpose is a relatively new development (Duyn, 2012, Song et al., 2022; Warbrick, 2022). For instance, one automated sleep staging approach has recently been developed by training a deep learning model on EEG data from simultaneous EEG-fMRI (Zou et al., 2022). This approach produced a sleep stage classification accuracy of more than 70% for an independent dataset, which approximates the performance of at least some of the earlier machine-learning models that were trained on unimodal EEG features (Boostani et al., 2017). However, it still falls short of the accuracy of more than 80%, typical for most deep learning models based on data from a single modality (Li et al., 2022; Sri et al., 2022). This lower performance could be due to noise and artifacts induced by the simultaneous acquisition of EEG and fMRI recordings, as well as the possibility that different features may be informative about sleep stages in simultaneously acquired EEG-fMRI compared to unimodal EEG or fMRI data. Thus, one critical question that still needs exploration is identifying the most informative features for efficient machine learning-based sleep stage classification that are also applicable across EEG and fMRI. A number of recent papers suggest

that non-linear measures, such as the entropy and fractal dimension of EEG and BOLD time-series are informative about the complexity in these signals during different sleep stages considering a non-linear and dynamic brain, compared to traditional features, which rely on assumptions of linearity and stationarity (Acharya et al., 2015; Ma et al., 2018; Zhou et al., 2016).

Entropy is one such non-linear measure from thermodynamics, originally used to represent the disorder or randomness in a system (Clausius, 1867; Namdari & Li, 2019). Within physiological time-series analysis, however, entropy has been adopted to calculate the irregularity, or complexity of signals (Pincus & Goldberger, 1994; Shannon, 1948). This is done by analyzing the signal in either the time or frequency domain to see how similar or different segments of the signal are (Gutiérrez-Tobal et al., 2015). Moreover, various types of entropy, such as Shannon entropy, Approximate Entropy (ApEn), Sample Entropy (SampEn) have been applied to study EEG and BOLD time series (Burioka et al., 2005; Da Costa et al., 2002; De Araújo et al., 2003; Ge et al., 2007; Lo et al., 2022; Miskovic et al., 2018; Pincus, 1991; Richman & Moorman, 2000; Shannon, 1948; Wang et al., 2014). In the context of sleep stage classification both ApEn and SampEn seem to be of particular interest. More specifically, in EEG signals, higher ApEn and SampEn, representing increased irregularity, are commonly associated with wakeful states, whereas lower ApEn and SampEn, indicating more regularity, are usually a characteristic of the deeper N-REM sleep stages (Burioka et al., 2005; Ge et al., 2007). Similarly to entropy, fractal dimension (FD) is yet another non-linear measure, that originates from the field of mathematical chaos theory, used to quantify the complexity or roughness of a geometric shape or temporal pattern (Mandelbrot & Van Ness, 1968; Mandelbrot & Wheeler, 1983). Entropy and FD have complementary perspectives on complexity: entropy quantifies

regularity or predictability, while FD captures the complexity of patterns by examining how the details in certain patterns “change with the scale at which they are measured” (Lau et al., 2022; Porcaro et al., 2024). In physiological time-series analysis, FD has been used to evaluate the complexity and self-similarity within biological signals, providing insights into the underlying dynamics of physiological processes (Esteller et al., 2001). To accomplish this, one can examine the signal's structure in the time domain to assess the degree of pattern repetition over various scales. Different forms of FD, such as the Highuchi Fractal Dimension (HFD) and the Katz Fractal Dimension (KFD), have been applied to analyze physiological time series, each method providing unique perspectives on the signal's fractal nature (Higuchi, 1988; Katz, 1988). Similar to the application of entropy measures, fractal dimension analysis in EEG time series reveals higher FD values during wakefulness, and lower FD during the deeper stages of N-REM sleep (Chen, 2017; De Miras et al., 2019). In addition, the Hurst exponent (HE) and Detrended Fluctuation Analysis (DFA) are central in fractal analysis, relating closely to FD by evaluating the persistence or self-similarity over longer timescales, offering further depth in understanding the complexity of physiological states, including sleep dynamics from both EEG and BOLD time-series (Farag & El-Metwally, 2012; Gneiting & Schlather, 2004; Hurst, 1951; Mandelbrot & Wallis, 1969; Peng et al., 1994; Peng et al., 1995; Tagliazucchi, Von Wegner, et al., 2013). Research using EEG has shown that alpha (α) exponents, as measured by DFA, vary across different sleep stages, although the findings are inconsistent (Goshvarpour & Abbasi, 2013; Lee et al., 2001). Some studies report that α from DFA are higher in undifferentiated sleep states than in wakefulness, whereas others find the opposite trend (Goshvarpour & Abbasi, 2013; Lee et al., 2001). Similarly, analyses using both EEG and fMRI data show varied trends for the HE (Acharya et al., 2005; Tagliazucchi, Von Wegner, et al., 2013; Weiss et al., 2009). Typically, the

HE decreases progressively from wakefulness to deep N-REM sleep (Acharya et al., 2005; Tagliazucchi, Von Wegner, et al., 2013). However, some findings in EEG data indicate increases in the HE from higher to lower N-REM stages, with a decrease during REM sleep (Weiss et al., 2009).

That being said, the aforementioned entropy and fractal features have been extensively studied in the context of EEG-based sleep stage classification. Boostani et al. (2017) evaluated automated EEG-based methods for classifying sleep stages, identifying the entropy of wavelet coefficients and the random forest classifier as the most effective feature and classifier, respectively. This approach achieved an accuracy rate of 87% for classifying sleep stages in healthy subjects and 69% in REM behavioral disorder patients. Likewise, according to Rodríguez-Sotelo et al. (2014), EEG-based entropy and fractal features are useful predictors of sleep stages, with ApEn and FD having the best performance. In another study, Chouvarda et al. (2011) also examined, among others, the FD, ApEn, and SampEn of EEG, observing statistically significant differences in these metrics across sleep stages. Finally, Acharya et al. (2015) compared 29 non-linear features, including ApEn, SampEn, Hurst exponent, and DFA, for EEG-based sleep analysis, finding all of these features to be effective in distinguishing between sleep stages. Similar studies are yet to be conducted to gain more insight into how predictive the entropy and fractal dimensions of BOLD time series are regarding sleep stages, especially when using machine learning algorithms. Nonetheless, it is becoming more and more clear that in the context of sleep analysis, EEG-based approaches could benefit immensely from the application of non-linear features, such as ApEn, SampEn, Highuchi's FD, Hurst exponent, and DFA.

Having outlined the historical context leading to the advent of simultaneous EEG-fMRI and the promising results of nonlinear features in previous studies on sleep stages, several areas

that require further exploration should be acknowledged. To date, only a limited number of studies have investigated the correlations of fractal features, such as the Hurst exponent, and, notably, none seem to have examined entropy features, between the signals of various brain regions in fMRI, or between time series of EEG channels and fMRI regions of interest (ROIs) recorded simultaneously during sleep (Tagliazucchi, Von Wegner, et al., 2013). More precisely, the only relevant study by Tagliazucchi, Von Wegner, et al. (2013) investigated the correlation between Hurst exponents calculated for BOLD time series and variance in delta-power across all EEG channels, finding a negative correlation in the frontoparietal regions, including the frontoparietal attentional (FAN) and default mode networks (DMN). These findings underline the need to explore the link between non-linear features of the BOLD time series in brain regions associated with the FAN and DMN and the same features of signals in frontoparietal EEG channels before evaluating their informativeness in sleep stage classification. Moreover, as of now, there seem to be no machine learning models available that have been trained on entropy and fractal features from purely fMRI, or simultaneous EEG-fMRI data, for sleep stage classification and prediction. This research paper will delve into these uncharted territories.

First, this study aims to shed light on how the entropy, as well as fractal features, derived from BOLD time series of brain regions associated with the FAN and DMN relate to the same features calculated for time series in frontoparietal EEG channels in simultaneous EEG-fMRI data during rest and sleep. Since no exact precedent has been found that could serve as a basis for hypotheses regarding this research question, this aspect of the study is rather exploratory in its nature. Nonetheless, inspired by the results of Tagliazucchi, Von Wegner, et al. (2013), the minimal prediction of the current study regarding the first research question was that the Hurst Exponent from time series of fMRI ROIs associated with the FAN and DMN would show a

statistically significant, negative relationship with the Hurst Exponent of time series from the frontal and parietal EEG channels. In addition, considering what the rest of the non-linear measures used in this study quantify, significant correlations between the same EEG channels and fMRI ROIs are expected for SampEn, ApEn, HE, and DFA features, although the direction of the correlations may differ. These relationships will be investigated through correlational analyses of the aforementioned entropy and fractal features from EEG and BOLD time series from complete simultaneous EEG-fMRI recordings.

Second, another aim of the current research is to explore the cross-modality correlations between each entropy and fractal feature of frontoparietal EEG channels and the same features in BOLD time series from ROIs associated with the FAN and DMN during wakeful- and sleep stage-specific segments of data. In alignment with the previous research question and hypotheses, the same general patterns are expected between features across frontal and parietal EEG channels and fMRI ROIs linked to the FAN and DMN during wakeful and N-REM sleep stages. Furthermore, similarly to the predictions related to the previous research question, these hypotheses will be tested using correlational analyses for entropy and fractal features from EEG and BOLD time series, but this time within sleep stage-specific segments of simultaneous EEG-fMRI recordings.

Third, this study assesses how effective ApEn, SampEn, Highuchi's FD, HE, and DFA features from simultaneous EEG and BOLD signals are in distinguishing and accurately classifying sleep stages. In the context of this final research question, there will be an initial analysis of variance (ANOVA) for the sleep stage-specific entropy and fractal features averaged across all EEG channel and fMRI ROIs to investigate if these features differ across sleep stages. Based on the previously highlighted literature, it is hypothesized that ApEn, SampEn, HFD, as

well as HE and DFA values will decrease statistically significantly from wakefulness to deeper N-REM sleep stages. Finally, this study will evaluate the performance of machine learning models trained with both EEG and BOLD-based ApEn, SampEn, Highuchi's FD, HE, and DFA features from simultaneous EEG and BOLD signals in predicting sleep stages and wakefulness, comparing their performance to models trained only on EEG or BOLD features. The hypothesis in this case is that models trained on features from both EEG and BOLD modalities will achieve higher accuracy in predicting sleep stages, thereby outperforming those based purely on EEG or fMRI data.

To summarize, the objective of this study is to uncover the correlations of several entropy and fractal features between frontoparietal EEG channels and fMRI ROIs related to the FAN and DMN in simultaneous EEG-fMRI data, their effectiveness in distinguishing wakeful and sleep stages, as well as the predictive power of machine learning models trained on these features for accurate sleep stage classification.

Methods

Dataset

The current study used an open-access dataset released by Gu and colleagues (2023) in Brain Imaging Data Structure (BIDS) format, accompanying a previous publication by the authors (Gorgolewski et al., 2016; Gu et al., 2022). This dataset contained raw structural and functional MRI, as well as raw EEG data from 33 healthy participants, all of which were obtained with informed consent at Pennsylvania State University (Gu et al., 2023). The participants were 17 males and 16 females with an average age of 22.1 years, and a standard deviation of 3.2 years (Gu et al., 2023). For each participant, the dataset featured a T1-weighted anatomical scan, two 10-minute sessions of resting-state data, and several 15-minute sleep

recordings. The structural MRI data were collected by a 3 Tesla Prisma Siemens Fit scanner with a Siemens 20-channel receive-array coil and a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (TR/TE = 2300/2.28 ms, TI = 900 ms, flip angle = 8°, FOV = 256mm, matrix size 256 × 256 × 192, voxel size = 1×1×1 mm3, acceleration factor = 2) (Gu et al., 2023). The functional MRI scans were based on an echo-planar imaging (EPI) sequence (TR/TE = 2100/25ms, flip angle = 90°, 35 slices, FOV = 240 mm, voxel size = 3×3×4 mm3) (Gu et al., 2023). As for the EEG, recordings were carried out using a 32-channel MR-compatible equipment, with electrode placements according to the 10-20 international system. Eye movements and heart activity were monitored through electrooculography (EOG) and electrocardiography (ECG), respectively. The sampling rate of the EEG was 5000 Hz and band-pass filtering was applied from 0 to 250 Hz. Besides, the EEG also included R128 markers that represented the fMRI volume triggers to align the simultaneously recorded BOLD and EEG signals. Furthermore, a registered polysomnography expert manually labeled the sleep stages in the EEG data, evaluating each 30-second epoch for wake (‘w’) or sleep stages ‘1’, ‘2’, and ‘3’, which correspond to N-REM 1, N-REM 2, and N-REM 3, in accordance with the sleep scoring guidelines of AASM. The scoring also included annotations for epochs where uncertainty was present, marked as “uncertain”, and epochs where excessive artifacts prevented a reasonable score, labeled as “unscorable” (Gu et al., 2023). The authors noted that this sleep staging process involved certain preprocessing steps, including the removal of MRI-induced artifacts, re-referencing of the EEG data to the contralateral mastoid, and band-pass filtering from 0.3 to 35 Hz.

EEG Data Preprocessing

Guided by the paper associated with the dataset, as well as an initial visual assessment, MRI gradient artifacts and ballistocardiographic (BCG) pulses had to be removed from all raw EEG recordings (Gu et al., 2023). Gu and colleagues (2023) used an algorithm named AMRI for artifact removal (Liu et al., 2012). In addition to this, several other tools have been designed for mitigating MRI-related artifacts in EEG data, including the fMRIb plugin for EEGLAB, as well as more comprehensive pipelines, such as APPEAR (Delorme & Makeig, 2004; Iannetti et al., 2005; Mayeli et al., 2021; Niazy et al., 2005). In order to select an MRI artifact cleaning approach for this study, the outcomes produced by AMRI, fMRIb, and APPEAR were all visually inspected in the same segments of EEG data from randomly selected sessions and participants, as depicted in Figure A1 in Appendix A. Based on the dataset used in this study, AMRI was chosen for developing a custom MATLAB pipeline to clean the EEG data, as it processed sessions 3-7 minutes faster and produced visually cleaner signals with fewer residual rhythmic artifacts from the MRI gradients compared to fMRIb and APPEAR. This process involved eliminating MRI gradient artifacts, identifying QRS complexes from the ECG, integrating potentially missing R128 markers into the EEG, and removing cardiac pulse artifacts (Liu et al., 2012). In addition, the dataset was downsampled to 250 Hz and low-pass filtered at 125 Hz.

Even though AMRI reduced the presence of MRI gradient and BCG artifacts in the EEG data, additional preprocessing was necessary. Correspondingly, a custom Python pipeline was developed using the MNE package (Gramfort, 2013). The first step was importing the AMRI-cleaned data, together with relevant information regarding R128 markers and annotations on the different stages of sleep for each 30-second EEG segment, as provided by Gu et al. (2023). Then, the EOG and ECG channels were specified, after which the standard 10-20 electrode placement

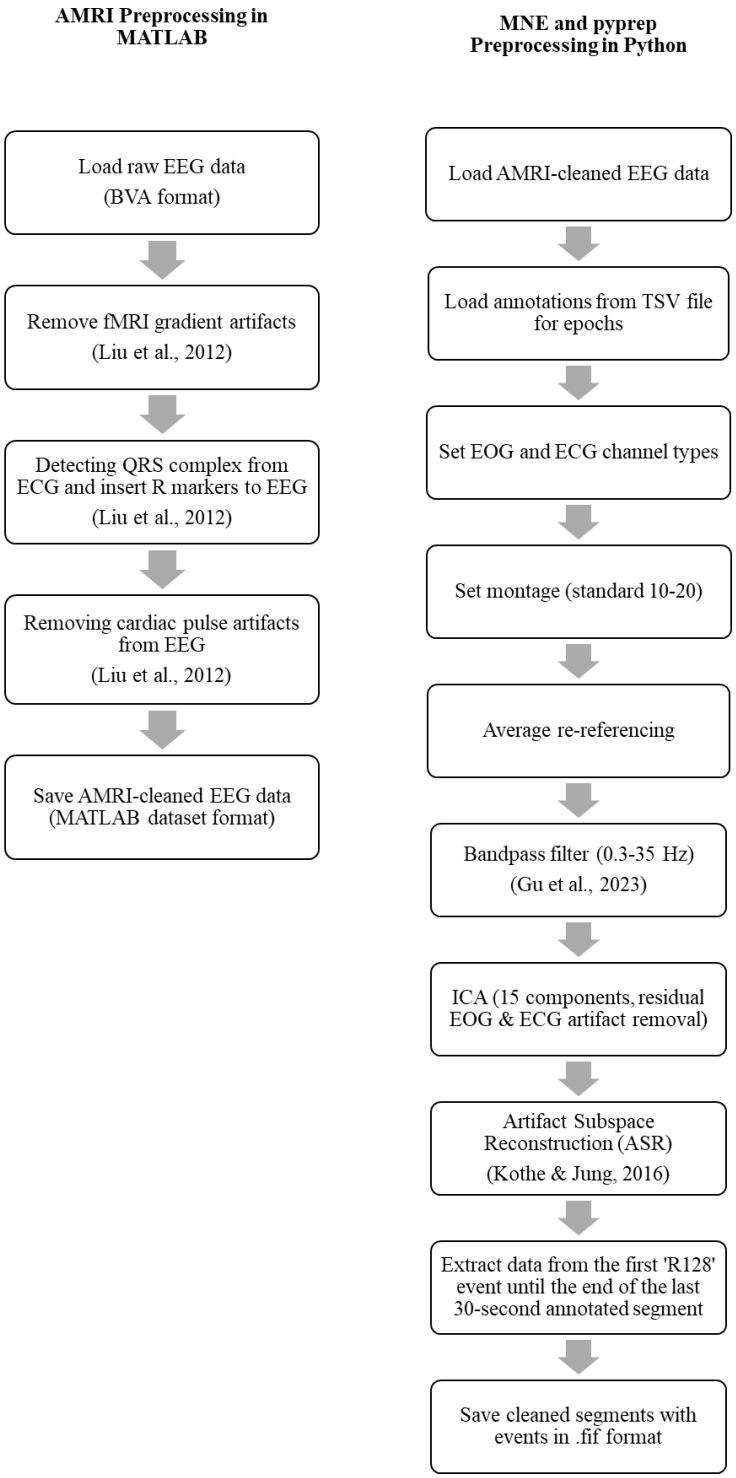
and average re-referencing was applied to the data. This was followed by 0.3 to 35 Herz (Hz) band-pass filtering, in line with the specifications of Gu and colleagues (2023) used for sleep staging. Moreover, independent component analysis (ICA) was conducted to distinguish and remove components related to the remaining EOG and ECG artifacts (Gramfort, 2013). To further ensure that bad portions of the EEG signals are cleaned, artifact subspace reconstruction (ASR) from the asrpy package was used (Kothe & Jung, 2016). Finally, based on the R128 markers and annotations of the 30-second-long segments, the cleaned data was cropped to obtain the segments corresponding to the simultaneous EEG-fMRI recordings. The resulting segments were then saved in fractal image format (FIF) for later analysis. Due to the fact that EEGLAB identified data truncation in three out of 255 sessions during the AMRI cleaning process, and an additional four sessions encountered errors during the preprocessing in MNE, these sessions were marked and subsequently removed from the analysis (Delorme & Makeig, 2004; Gramfort, 2013). The general steps of this preprocessing workflow are outlined in Figure 2.

FMRI Data Pre- and Postprocessing

Although some of the conventional approaches to fMRI data pre- and postprocessing have typically relied on manual techniques with software like SPM, AFNI, FSL, and AIR, these methods can be tedious and impractical in large datasets, while also introducing variability and subjectivity in method and parameter selection (Cox, 1996; Fischl et al., 2002; Friston et al., 1994; Smith et al., 2004). To address these challenges and enhance the reliability and replicability of research findings, this study applied standardized, robust pipelines which integrate a number of established tools and criteria for processing large resting-state fMRI (rs-fMRI) datasets. This strategy involved a first preprocessing stage with fMRIPrep and then postprocessing with the eXtensible Connectivity Pipeline-DCAN (XCP-D) (Esteban et al., 2018;

Figure 2

Overview of Preprocessing Steps for EEG in This Study.



Mehta et al., 2023). The anatomical and functional MRI data underwent preprocessing using fMRIPrep version 23.1.4 with default settings, ensuring artifact-free data that was suitable for subsequent analyses (Esteban et al., 2018). As part of the preprocessing, the T1-weighted anatomical data of each participant went through several steps, such as intensity normalization, skull stripping, tissue segmentation, and also spatial normalization. As for the functional MRI, these were corrected for head motion and slice-timing, aligned with anatomical data, and underwent smoothing, noise reduction and spatial normalization to a standard space. To the request of the creators of fMRIPrep, an automatically generated, detailed report of the preprocessing procedures in this dataset can be found in Appendix B.

Following the preprocessing of fMRI data with fMRIPrep, the data were further refined using the eXtensible Connectivity Pipeline-DCAN (XCP-D) (Mehta et al., 2023). The postprocessing involved transforming native-space T1-weighted images to MNI space, and identifying non-steady-state volumes. Besides, nuisance regressors, including six filtered motion parameters, mean global signals (global, white matter, cerebrospinal fluid), quadratic expansion of motion parameters and tissue signals, as well as their temporal derivatives were selected and regressed from the BOLD data, with additional steps including despiking for motion outliers, band-pass filtering, and smoothing (Mehta et al., 2023). Framewise displacement was disabled to avoid censoring segments of the data containing motion artifacts. Even though, in general, this is considered bad practice, the decision was made to preserve the temporal order and integrity of the time series, which was important for accurately interpreting the relationship between BOLD and EEG features across sleep stages. As a final step, the BOLD time-series data were extracted using various brain atlases. While XCP-D offered cortical, subcortical, and combined atlases, such as the 4S156 atlas, the time series extracted from the latter contained a large number of non-

numeric values with up to 27,742 instances in some sessions, and affecting at least 22 out of 156 ROIs for some participants (Schaefer et al., 2018). Consequently, the BOLD time series in this study were derived from the 48 ROIs defined by the cortical Harvard Oxford Atlas in FSL (Jenkinson et al., 2012). In a manner similar to fMRIprep, the specific methods and parameters of the XCP-D post-processing pipeline are detailed in an automatic report, which is included in Appendix C at the authors' request.

Feature Extraction

Approximate Entropy (ApEn)

ApEn, as highlighted beforehand, is a complexity measure introduced by Pincus (1991), capturing the irregularity or unpredictability of a time series. In simple words, ApEn quantifies complexity by measuring how certain sequences in a signal remain similar or dissimilar over consecutive observations. The general steps of estimating ApEn in time series were explained in Appendix D.

To calculate ApEn in physiological time series, the Python package NeuroKit2 was used (Makowski et al., 2021). Regarding the parameter selection, the calculation of ApEn depends on the time series length (N), as well as the specific combination of parameters *m* and *r*, or, in other words, segment length and tolerance (Pincus, 1991). Since no universal and well-established method exists for optimizing these parameters, different values had to be considered for the separate stages of analysis and brain imaging modalities based on previous literature. Consequently, as the BOLD time-series used in this dataset ranged between 286 and 429 data points, m=2 and r=0.2 * SD were chosen for calculating ApEn for both for the entire time series, as well as the short, sleep stage-specific segments, which will be discussed later in this method section (Pincus, 1995; Pincus & Huang, 1992; Yentes et al., 2012). As for the EEG data, some of

the most relevant studies have found that m=2 and r=0.2 * SD yielded consistent ApEn values that could be used to distinguish between sleep stages (Burioka et al., 2005). Thus, in this research, m=2 and r=0.2 * SD were also adopted for analyzing EEG data from both specific sleep stages and the complete signal.

Sample Entropy (SampEn)

As mentioned previously, SampEn is an adaptation of ApEn, that is highly relevant in physiological time-series analysis (Richman & Moorman, 2000). This is because SampEn, in contrast to ApEn, avoids the issues of inconsistent results across different data series lengths by not including identical, matching segments in its calculation, thereby ensuring more reliable and consistent quantification of irregularity or complexity (Delgado-Bonal & Marshak, 2019; Richman & Moorman, 2000; Xie et al., 2008). The main steps involved in calculating SampEn were described in Appendix E.

Similar to the calculation of ApEn, NeuroKit2 was used to compute SampEn (Makowski et al., 2021). Regarding the parameter selection for SampEn, a study by Yang et al. (2018) provided a strategy for selecting these parameters and presented combinations of *m* and *r* depending on the length of various BOLD time series, that seemed to produce reliable output and that were consistent with the previously discussed findings. Based on this strategy, m=2 and r=0.35 * SD were chosen for SampEn calculation in complete BOLD time-series with 286 and 429 data points in the dataset used in this study, as they fell into the acceptable range highlighted by Yang et al. (2018). Similarly, following the guidelines of Yang et al. (2018), for obtaining SampEn in BOLD signals of sleep stage-specific segments, m=1 with r=0.35 * SD were selected. For the EEG data, while some studies used m=2 in combination with r=0.1 and r=0.25, the majority of papers in sleep analysis supported values of r=0.2 * SD and m=2 (Bruce et al., 2009;

Ge et al., 2007; Jiang et al., 2015; Li et al., 2015). Correspondingly, this study applied a combination of m=2 and r=0.2 for SampEn calculation in both sleep stage-specific and full EEG time series.

Higuchi's Fractal Dimension (HFD)

As pointed out in the introduction, HFD is another non-linear measure that has been used for measuring complexity in time series, including in neurophysiological signals during sleep stages (Higuchi, 1988; Kesić & Spasić, 2016). To demonstrate the process of calculating HFD, a detailed description can be found for this purpose in Appendix F. HFD values can theoretically take values in the range of 1 and 2, depending on the time series length and k_{max} value, which is the single, most important parameter for HFD.

One implementation of HFD calculation in Python programming language is available in the NeuroKit2 package (Makowski et al., 2021). Moreover, besides providing a way to obtain HFD values based on the paper of Higuchi (1988), NeuroKit2 also incorporates a function that can optimize the k_{max} parameter for each individual time series (Makowski et al., 2021; Vega & Noel, 2015). This is especially important because there was no consensus in the literature on what is an adequate time series length and associated k_{max} parameter for HFD estimation. Therefore, this study applied a customized, multi-step approach for determining adequate k_{max} values for the specific dataset used in this research. First, the parameter optimization tool of NeuroKit2 was applied across all complete or sleep stage-specific segments from both EEG channels and fMRI ROIs, resulting in numerous k_{max} values for every channel or ROI, as well as series length and modality (Makowski et al., 2021). This optimization tool assessed a range of k_{max} values and selected the best fitting one for a specific time series based on the point at which HFD values started to reach a plateau (Makowski et al., 2021; Vega & Noel, 2015). For both

complete and sleep stage-specific EEG series, as well as for full BOLD series, the maximum k_{max} value was set at 64. This limit was chosen to prevent excessive computational demands and was consistent with some of the highest k_{max} values found in the relevant literature (Varley et al., 2020). Following this, the mean k_{max} values were computed for each EEG channel and fMRI ROI, as well as series length and modality. The third and final step involved taking these means for each EEG channel and fMRI ROI, as well as series length and modality, and rounding them up to the nearest integer, ensuring a better probability of the HFD values stabilizing at that k_{max}. Using this method, optimal k_{max} values were selected for each EEG channel and fMRI ROI for both complete time series and sleep stage-specific segments, as presented Table F1 and Table F2 of Appendix F, respectively.

Hurst Exponent (HE) and Detrended Fluctuation Analysis (DFA)

Another non-linear measure that relates closely to fractal dimension and quantifies the long-term complexity, persistence, or self-similarity of signals is the Hurst Exponent (HE) (Hurst, 1951). When it comes to time series that is self-similar, fractal dimension (D) relates directly to HE, as can be seen in Equation 1:

$$D = 2 - HE \,, \tag{1}$$

where D can have a value between 1 and 2, whereas the HE value ranges from 0 to 1 (Nazarychev et al., 2019). Moreover, the HE of time series can be estimated through different approaches, including rescaled-range analysis (R/S) and DFA (Hurst, 1951; Peng et al., 1994). This study used the R/S implementation of NeuroKit2 with default settings to estimate the HE directly within a range of 0 to 1, in which case values precisely at 0.5 suggest non-correlated behavior, while values above and below 0.5 are signs of persistence and anti-persistence, respectively (Makowski et al., 2021).

Furthermore, this study applied monofractal DFA analysis using default settings in NeuroKit2 to obtain the alpha exponent (α), which is a robust measure of the overall scaling behavior and persistence of the features in the EEG and BOLD data and important for understanding long-term correlations across imaging modalities (Hardstone et al., 2012; Makowski et al., 2021; Peng et al., 1994). The α in monofractal DFA ranges from 0 to 2, where values less than 0.5 indicate anti-persistence, values exactly at 0.5 suggest non-correlated behavior, and values above 0.5 up to 1 indicate persistence (Hardstone et al., 2012). Values between 1 and 2 generally imply non-stationarity (Hardstone et al., 2012). The exact procedures of calculating HE with R/S and the alpha exponent with monofractal DFA can be found in Appendix G.

Correlational Analysis

Between-Imaging Modality Correlational Analysis of Entropy and Fractal Features in Complete EEG and BOLD Time Series

To address the first research question, a correlational analysis was conducted using entropy and fractal features extracted from the full EEG and BOLD time-series. Before the analysis, the distribution and variance of features from every EEG channel and fMRI ROI across all 248 sessions were plotted to visually assess both transient interactions across sessions, as well as cross-individual patterns. The resulting plots indicated similar mean estimates and variance across separate scanning sessions. Thus, features from each EEG channel and fMRI ROI were averaged across all sessions for each of the 33 participants, resulting in 30 EEG channel and 48 fMRI ROI feature values per participant. Importantly, correlational analysis with Pearson's r is based on several key assumptions including normal distribution of the data, absence of outliers, data derived from a random or representative sample, both variables being measured at an

interval or ratio level, as well as an expected linear relationship between the variables (Schober et al., 2018). The last three assumptions were inherently met. However, the normality of data distributions and absence of outliers were rigorously checked using histograms, Q-Q plots, Shapiro-Wilk tests for normality, and boxplots for outlier detection (Shapiro & Wilk, 1965). The Shapiro-Wilk test was chosen due to its effectiveness for datasets containing fewer than 50 samples, given the presence of 30 EEG and 48 fMRI ROI values per feature and participant in the dataset (Mishra et al., 2019). Outliers were quantified as data points more than one and a half times the interquartile range (IQR) below and above the first and third quartiles for each feature (Tukey, 1977). Based on visual assessment of the histograms and Q-Q plots, combined with the quantitative analysis from the Shapiro-Wilk test, violations of the normality of distributions were detected in 36 out of 150 EEG channel and 69 out of 240 fMRI ROI features. Likewise, several outliers had to be removed to maintain the integrity of the correlational analysis. Due to these deviations, Spearman's rank correlation coefficients were computed instead of Pearson's r to evaluate the relationship between frontoparietal EEG channels and fMRI ROIs associated with the FAN and DMN for each entropy and fractal feature (Schober et al., 2018). The EEG electrodes of interest were the channels from the frontal (Fp1, Fp2, F3, F4, F7, F8, FC1, FC2, FC5, FC6, Fz) and parietal (P3, P4, P7, P8, Pz) clusters. These clusters were partially based on the findings of Rojas et al. (2018), who demonstrated that frontal and parietal EEG channels are indicative of the functional connectivity in both the frontoparietal attention network (FAN) and the default mode network (DMN). Furthermore, fMRI ROIs in the FAN were set using the references provided by Markett et al. (2013), namely intracalcarine cortex (ICc) for a close approximation of intraparietal sulcus; superior parietal lobule (SPL), supramarginal gyrus, anterior division (SMGa), and supramarginal gyrus, posterior division (SMGp) for the inferior

parietal lobule; middle frontal gyrus (MFG) for dorsolateral prefrontal cortex; the juxtapositional lobule cortex (JLC), formerly known as supplementary motor cortex, for the supplementary motor area; and insular cortex (IC) for the anterior insula. As for the DMN, it included regions like the medial prefrontal cortex (MPFC), posterior cingulate cortex (PCC), the left and right temporal parietal junctions (TPJ), and the precuneus, which correspond to frontal medial cortex (FMC), cingulate gyrus, posterior division (CGp), angular gyrus (AG), and precuneus cortex (PC) respectively, in this study (Utevsky et al., 2014; Wang et al., 2017). Significance tests on the resulting coefficients were performed using an alpha threshold of 0.05. To mitigate alpha-error inflation, the Benjamini-Hochberg procedure was implemented (Benjamini & Hochberg, 1995). Additionally, non-significant coefficients were masked, and hierarchical clustering analysis was applied to the correlational matrices to improve the visualization of statistically significant correlations.

Between-Imaging Modality Correlational Analysis of Entropy and Fractal Features in Sleep Stage-Specific EEG and BOLD Time Series

For the second research question, a similar correlational analysis was conducted for sleep stage-specific segments. As indicated earlier, the dataset included annotations for sleep stages in 30-second-long epochs, providing a substantial amount of data for wakeful, N-REM 1, and N-REM 2 sleep stages with 3344, 2148, and 1325 epochs, respectively. However, due to the limited number of 42 N-REM 3 segments and the absence of REM sleep stages, these stages were omitted from the study. Given the lower sampling rate of the BOLD time series, it served as the reference for determining the minimum segment size for entropy and fractal feature calculations. Previous research suggested that physiological time series containing at least 200 data points, approximately corresponding to 14 epochs in the BOLD time series in this dataset, is sufficient

for reliably calculating the features used in this study (Delignières et al., 2006; Gomolka et al., 2018; Mayer et al., 2014; Pincus, 1995; Yentes et al., 2012). Accordingly, continuous BOLD segments of at least 14 epochs from the same sleep stage were extracted. Additionally, as segment lengths varied from 200 to 429 data points, a sliding window approach was applied to achieve time series with consistent lengths of 200 data points for feature calculation. For segments exceeding 200 data points, a 200 data point-long window was slid across the segment with a step-size of 100, creating a 50% overlap with the previous window, to extract sleep stage-specific feature values from each window. This technique produced feature values from 218 windows, with distributions of 167 from wakeful stages, 21 from N-REM 1, and 30 from N-REM 2. This method was similarly adapted for EEG data, which had a higher sampling rate of 250 Hz. The EEG window durations were precisely matched to those used for BOLD, making sure that the segment sizes and sliding window operations were aligned with the BOLD windows. This alignment facilitated precise and consistent feature value calculations across different sleep stages for both EEG and BOLD data. For the resulting sleep stage-specific features, the same steps for verifying normality of distributions and the absence of outliers, alongside correlational analyses, were implemented. Visual assessments of histograms, Q-Q plots, and the Shapiro-Wilk tests revealed major deviations from normality in the distributions of 48 EEG channel features from the N-REM 2 stage, 38 from the N-REM 1 stage, and 97 from the wakeful stage. Moreover, the normality assumption was also violated in 16 ROI features from the N-REM 2 sleep stage, 21 ROI features from N-REM 1, and 58 ROI features from wakeful stages. Similarly, several outliers, defined as values exceeding one and a half times the interquartile range (IQR) either below or above the first and third quartiles for each feature, were also identified and needed to be excluded. Considering these violations of the assumptions,

Spearman coefficients were calculated to evaluate the relationships between EEG channels and fMRI ROIs for each entropy and fractal feature. This was followed up with significance testing using an alpha threshold of 0.05, as well as the Benjamini-Hochberg procedure to correct for alpha-error inflation. Ultimately, non-significant coefficient values were masked, and hierarchical clustering analysis was performed on the correlational matrices to further reveal the patterns of relationships in the data.

Analysis of Variance (ANOVA) for Entropy and Fractal Features Across Sleep Stages in Sleep Stage-Specific EEG and BOLD Time Series Segments

As the first part of the third research question, this study explored whether entropy and fractal features from sleep stage-specific segments showed statistically significant differences across sleep stages. Taking into account that multiple entropy and fractal features (dependent variables) were collected for the various sleep stages (independent variable) within the same subjects, repeated measures ANOVA (rANOVA) was used on the features averaged across all EEG channel and fMRI ROI features. This was necessary to account for the correlation between these repeated measurements, reduce the error variance associated with individual differences, and increase the statistical power to detect differences between conditions compared to a one-way ANOVA with independent groups. Importantly, rANOVA requires several assumptions. More specifically, the dependent variable should be on an interval or ratio scale; the independent variable must include two or more categorical groups; the data should be collected through random sampling and normally distributed within each group; variances among the groups need to be homogenous; and the assumption about sphericity should not be violated (Weaver et al., 2017). The criteria for the types of data required, along with random sampling, were already met. Additionally, the normality of distributions was evaluated during the correlational analysis of

sleep stage-specific features. Despite deviations from normality in certain features, rANOVA is generally considered robust to normality violations in case the sphericity assumption is not violated (Blanca et al., 2023). Therefore, sphericity was tested using Mauchly's test, which showed that the sphericity assumption was met in all features (Mauchly, 1940). Homogeneity of variances was verified using Levene's test, showing violations of this assumption in the EEG-based HE and DFA features (Levene, 1960). In these instances, the non-parametric Friedman test was used instead of the standard rANOVA, and the Kendall's W for a measure of effect size instead of the eta squared (η^2) (Friedman, 1937; Kendall & Smith, 1939). The threshold for statistical significance was set at $p < 0.05$ for the F and chi-square (χ^2) values of the rANOVAs and Friedman tests, respectively (Weaver et al., 2017). Given the multiple tests conducted, the Benjamin-Hochberg procedure was applied to correct for alpha error inflation. To determine how feature means varied across sleep stages, Tukey's Honestly-Significant Difference (Tukey's HSD) post-hoc test was used for rANOVA results (Tukey, 1949). Meanwhile the outputs of the Friedman tests were analyzed through the Nemenyi test, which was a standard post-hoc analysis for the Friedman test (Nemenyi, 1963). The Benjamin-Hochberg procedure was applied again due to the multiple comparisons. Significant pairs at the $p < .05$ chance level following the Tukey's HSD and Nemenyi post-hoc analyses were plotted in boxplots.

Machine Learning Classification of Sleep Stages Trained on Entropy and Fractal Features From Sleep Stage-Specific Segments of EEG, BOLD, and Combined EEG and BOLD.

Ultimately, the second part of the third research question and the main focus of the current study was to evaluate the effectiveness of entropy and fractal features from EEG channels and fMRI ROIs, collected during simultaneous EEG-fMRI, in predicting sleep stages through supervised machine learning models. To be more specific, the performance of three

different model training approaches was examined: using features exclusively from EEG, exclusively from BOLD, and a combination of both. This approach allowed for assessing the performance of features from each imaging modality independently, as well as the potential benefit of integrating features from both modalities. To train models, the sleep stage-specific feature data from EEG and BOLD time series were loaded and randomly split into 80% training and 20% test sets, making sure that some subjects' sleep-stage specific features were entirely reserved for the test set. These steps were carried out using the scikit-learn (sklearn) package (Pedregosa et al., 2012). Importantly, EEG channel features were then clustered according to frontal (Fp1, Fp2, F3, F4, F7, F8, FC1, FC2, FC5, FC6, Fz), central (C3, C4, Cz, CP1, CP2, CP5, CP6), parietal (P3, P4, P7, P8, Pz), occipital (O1, O2, Oz), and temporal (T7, T8, TP9, TP10) regions by averaging the features in each cluster. This was done to reduce the total number of features so that these were more proportional to the sample size available for machine learning models. Following this, Group K-Fold cross-validation was implemented to further split the training data into five folds, while still maintaining the integrity of subject groups. For each fold, a SimpleImputer with mean strategy, as well as standard scaling were implemented (Pedregosa et al., 2012). Moreover, since more than 80% of the data constituted features from wakeful states, correction needed to be applied. To do so, synthetic minority over-sampling technique (SMOTE) was used from the imbalanced-learn (imblearn) package, which generates synthetic samples from the minority class to balance the dataset more effectively (Chawla et al., 2002; Lemaitre et al., 2016). This process made sure that each fold had a balanced representation of sleep stages, facilitating robust model training. Finally, the selection of appropriate models and parameters was decisive for such a classification task, especially considering the non-linear nature of the features involved. Given these aspects, models that can handle such types of data were

prioritized. The model performances were evaluated using metrics such as accuracy, precision, recall, and F1-score, which is a "harmonic mean of precision and recall" (Hicks et al., 2022). The definitions and formulas of these metrics can be found in Appendix H. In addition, confusion matrices were plotted to conduct a more detailed model evaluation. In a confusion matrix, true positives (TP) are positive cases identified correctly and true negatives (TN) are correctly identified negative cases (Hicks et al., 2022). Besides, false positives (FP) are positive cases identified incorrectly, while false negatives (FN) are negative instances classified incorrectly (Hicks et al., 2022). Moreover, the models were further evaluated based on the receiver operating characteristic (ROC) curve that compares the true and false positive rates (Bradley, 1997). More specifically, the area under the ROC Curve (AUC) was used as a metric, which ranges from 0.5 to 1 (Bradley, 1997). An AUC value of 0.5 shows random chance, whereas 1 demonstrates perfect ability of the model to distinguish between classes (Bradley, 1997).

Random Forest

Random forest (RF) is a form of ensemble methods that are extensions of the decision trees (Breiman, 2001). These models build on several decision trees, where each of the trees is trained on random samples of the feature data, thereby improving robustness and accuracy in comparison to a single decision tree (Breiman, 2001; Louppe, 2014). By having multiple trees, RF is less prone to overfitting and also reduces variability in predictions attributable to noise (Breiman, 2001). Some of the most important hyperparameters of RF models are the number of trees, the maximum number of features each tree should consider, the maximum depth of the trees, and the minimum number of samples per leaf (Scornet, 2017). To optimize these parameters, five-fold GridSearch CV was used on each of the five folds from Group K-Fold (Pedregosa et al., 2012). The parameter grid and the best model parameters can be found in

Table I1 in Appendix I. The best combination of these hyperparameters was then implemented into a RF model using sklearn to evaluate the performance on the test sets (Pedregosa et al., 2012).

Support Vector Machines

Another class of strong supervised learning algorithms that are useful for complex regression and classification problems are support vector machines (SVMs) (Cortes & Vapnik, 1995). SVMs are ideal for the non-linear features in this study because they can handle complex and high-dimensional data while lowering the danger of overfitting. Hyperparameter adjustment is necessary, though, for SVMs to operate optimally. Thus, some of the hyperparameters are the kernel coefficient (γ), which influences the model's decision boundary, and the regularization value (C), which balances low training error with model simplicity (Wainer & Fonseca, 2020). In this study, the best SVM parameters were selected using a five-fold GridSearchCV within each of the five groups from a Group K-Fold (Pedregosa et al., 2012). This process tested different combinations of the aforementioned hyperparameters to find the best ones, as displayed in Table I2 in Appendix I. This model was then evaluated on imputed and scaled test data.

Results

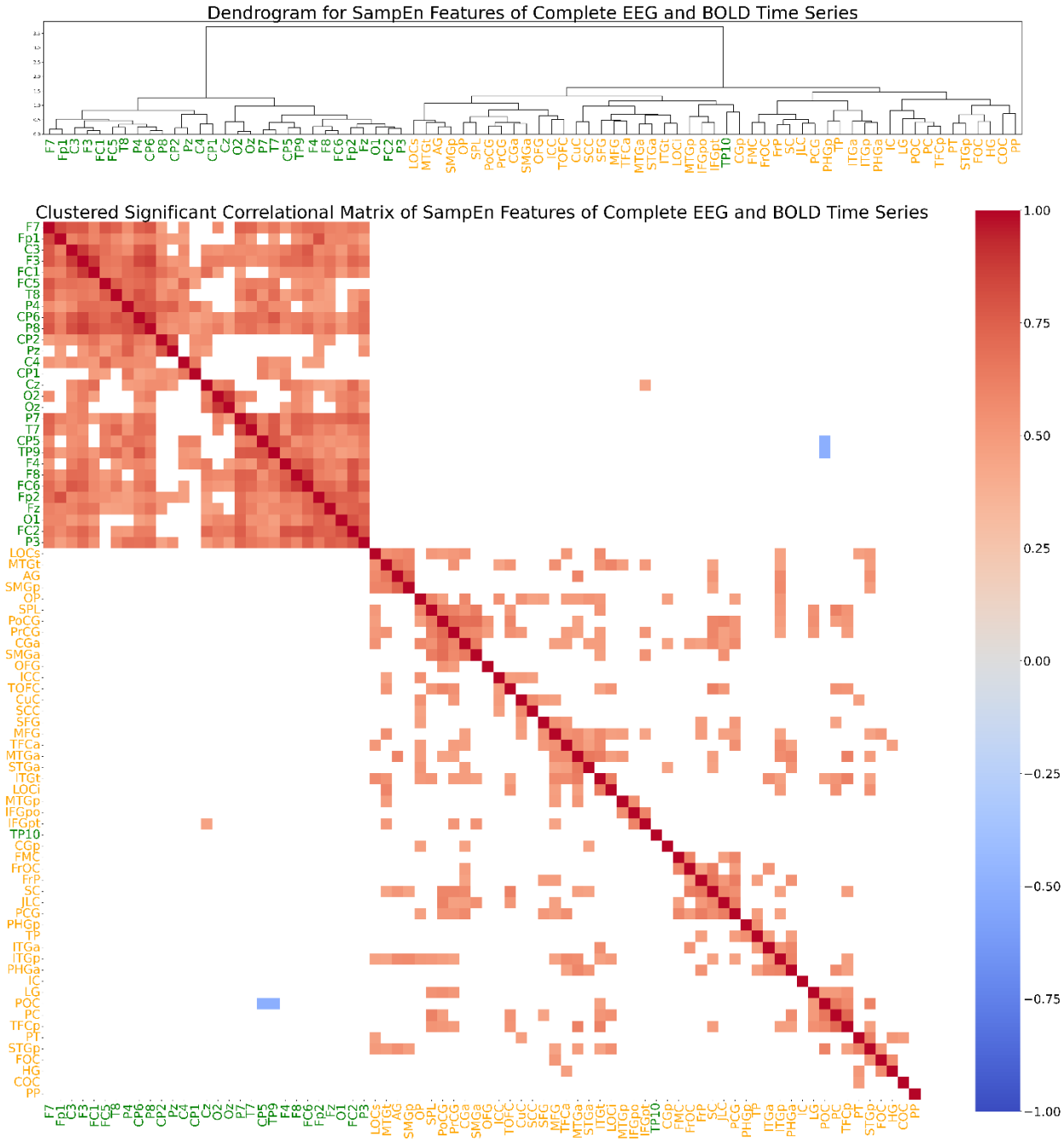
Correlational Analysis

Between-Imaging Modality Correlational Analysis of Entropy and Fractal Features in Complete EEG and BOLD Time Series

The correlations of features from the complete EEG and BOLD time series, together with hierarchical clustering analysis and clustered correlational matrices, were investigated between each EEG channel and fMRI ROI. As illustrated in the clustered correlational matrix and dendrogram in Figure 3, between-imaging modality analysis of SampEn features revealed three

Figure 3

Dendrogram and Clustered Correlational Matrix After Hierarchical Clustering Analysis for SampEn Features of Complete EEG and BOLD Time Series



Note. Non-significant correlations were masked. EEG channels were marked with green and fMRI ROIs in orange. The full ROI names were shared in the notes of Table F2 in Appendix F.

significant correlations, out of which one was positive and two negative. However, no significant correlations were found between SampEn features from EEG electrodes in frontal or parietal clusters and ROIs linked to the FAN or DMN.

For ApEn, the clustered correlational matrix and dendrogram from hierarchical clustering analysis are presented in Figure 4. Here, 83 between-modality correlations met the significance threshold of corrected $p < .05$. In particular, considering frontoparietal EEG channels and fMRI ROIs linked to the FAN, significant negative correlations were observed involving the Middle Frontal Gyrus (MFG) and F8, $r(31) = -.554$, $p = .005$, FC6 $r(31) = -.512$, $p = .011$, as well as P7, $r(31) = -.492$, $p = .016$. Additionally, when it comes to frontoparietal electrodes and fMRI ROIs associated with the DMN, significant negative correlations were identified between the Angular Gyrus (AG) and F8, $r(31) = -.479$, $p = .020$, as well as FC6, $r(31) = -.420$, $p = .046$.

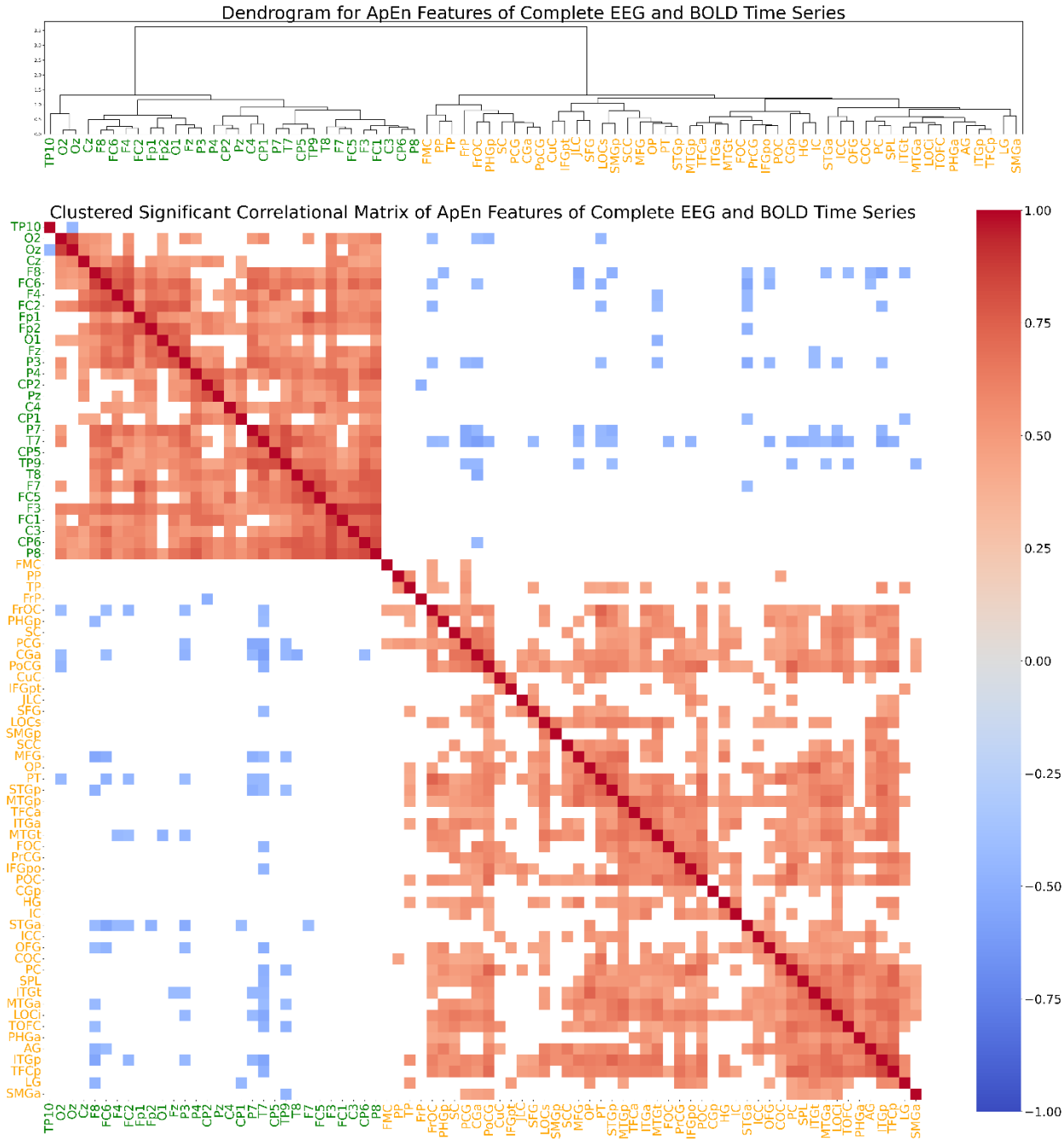
The analysis of HE features from R/S revealed only one significant correlation between EEG and BOLD, as shown in Figure 5. Nevertheless, no significant correlations were found between HE features of fMRI ROIs linked to the FAN or DMN and EEG channels from the frontal or parietal regions.

As for the alpha exponents from DFA, only one significant correlation was identified between EEG channels and fMRI ROIs, as depicted in Figure 6. However, the analysis did not reveal any significant correlations between alpha exponents from DFA in frontal or parietal EEG electrodes and ROIs associated with the FAN or DMN.

Finally, the HFD analysis showed 26 significant correlations between EEG and BOLD, as detailed in Figure 7. Among these, significant negative correlations were identified between FC5 and Middle Frontal Gyrus, $r(31) = -.499$, $p = .023$, and between Fz and Middle Frontal Gyrus, $r(31) = -.555$, $p = .009$, which are linked to the FAN. No significant correlations were

Figure 4

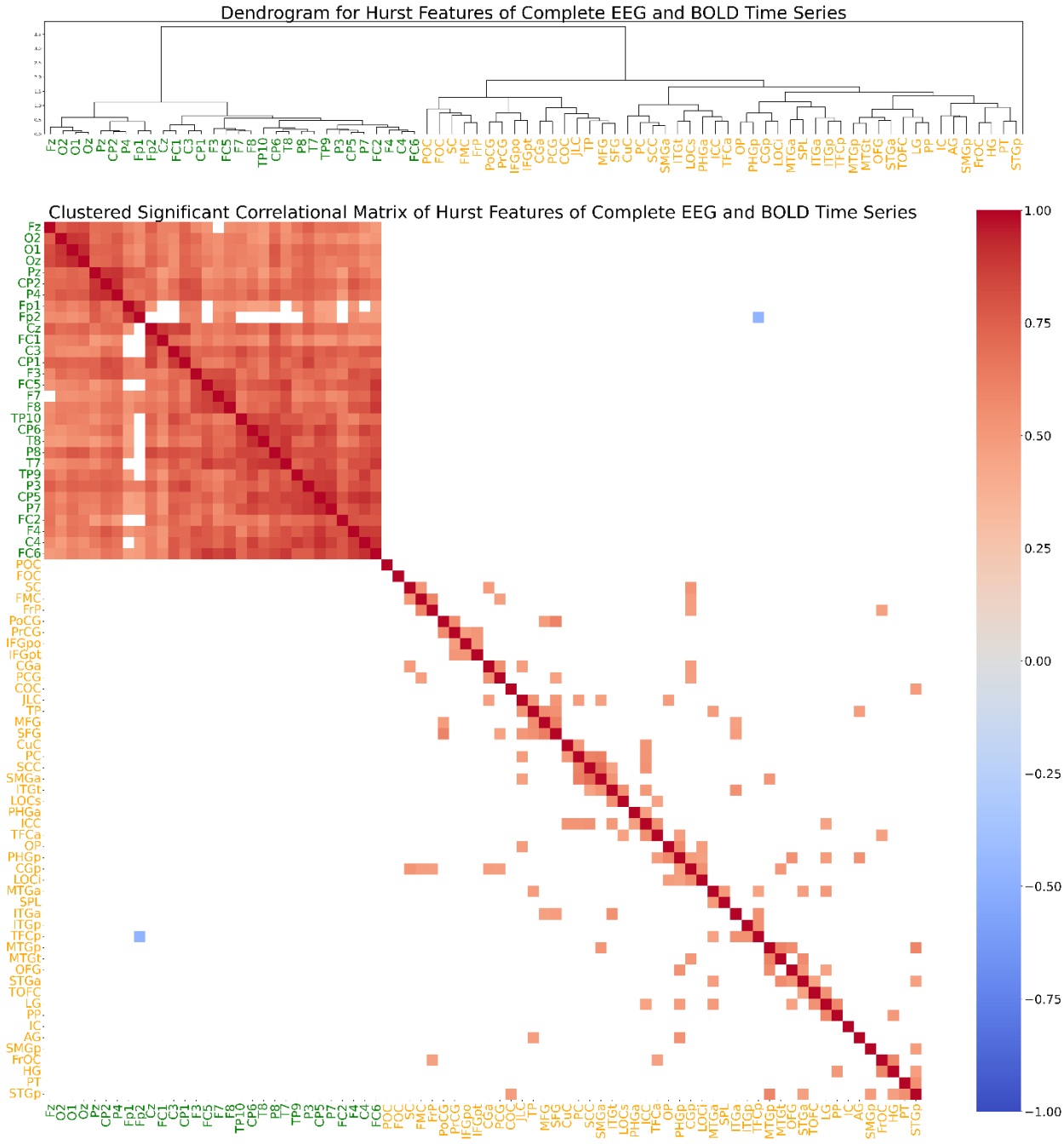
Dendrogram and Clustered Correlational Matrix After Hierarchical Clustering Analysis for ApEn Features of Complete EEG and BOLD Time Series



Note. Non-significant correlations were masked. EEG channels were marked with green and fMRI ROIs in orange. The full ROI names were shared in the notes of Table F2 in Appendix F.

Figure 5

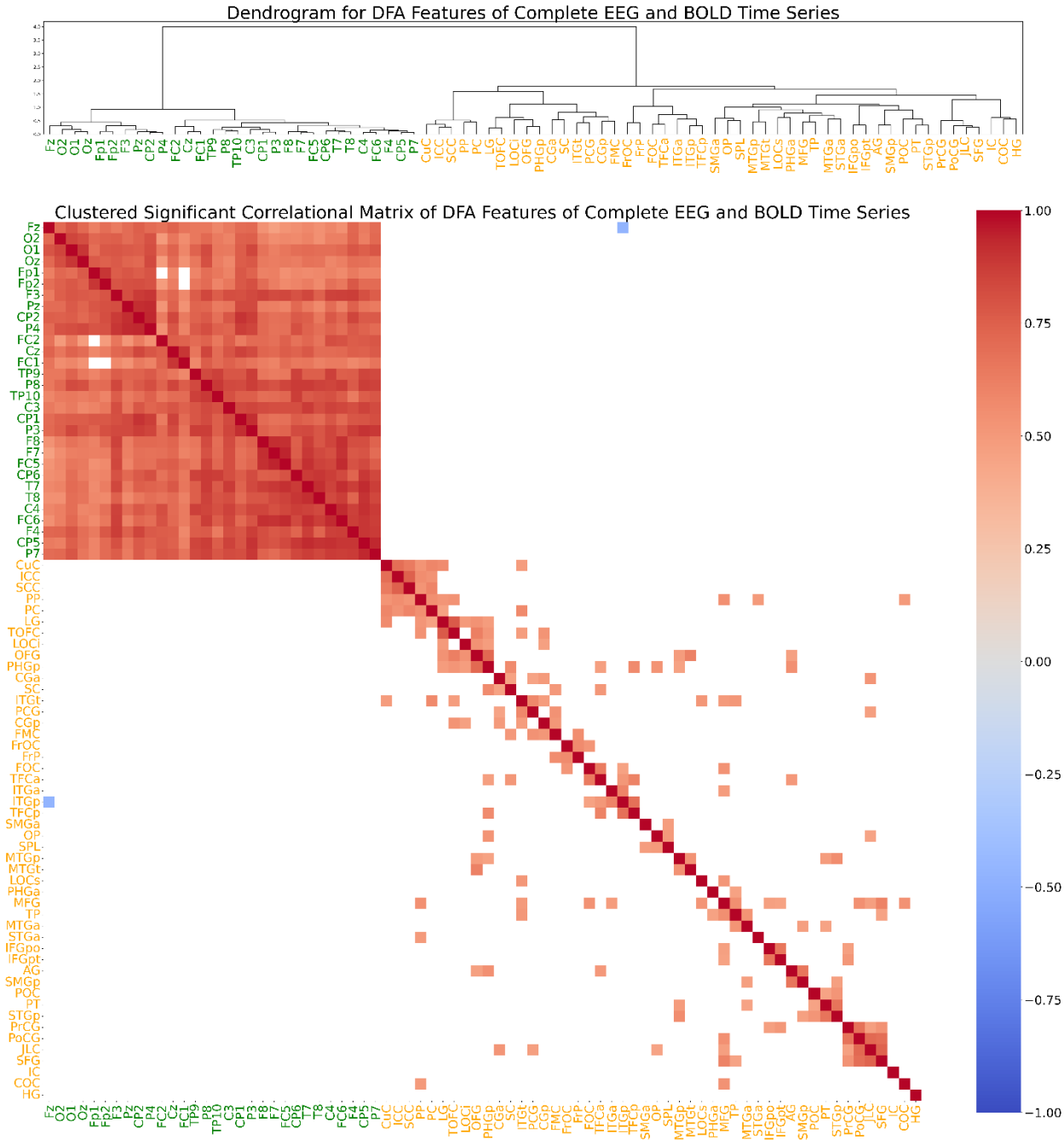
Dendrogram and Clustered Correlational Matrix After Hierarchical Clustering Analysis for Hurst Exponent Features of Complete EEG and BOLD Time Series



Note. Non-significant correlations were masked. EEG channels were marked with green and fMRI ROIs in orange. The full ROI names were shared in the notes of Table F2 in Appendix F.

Figure 6

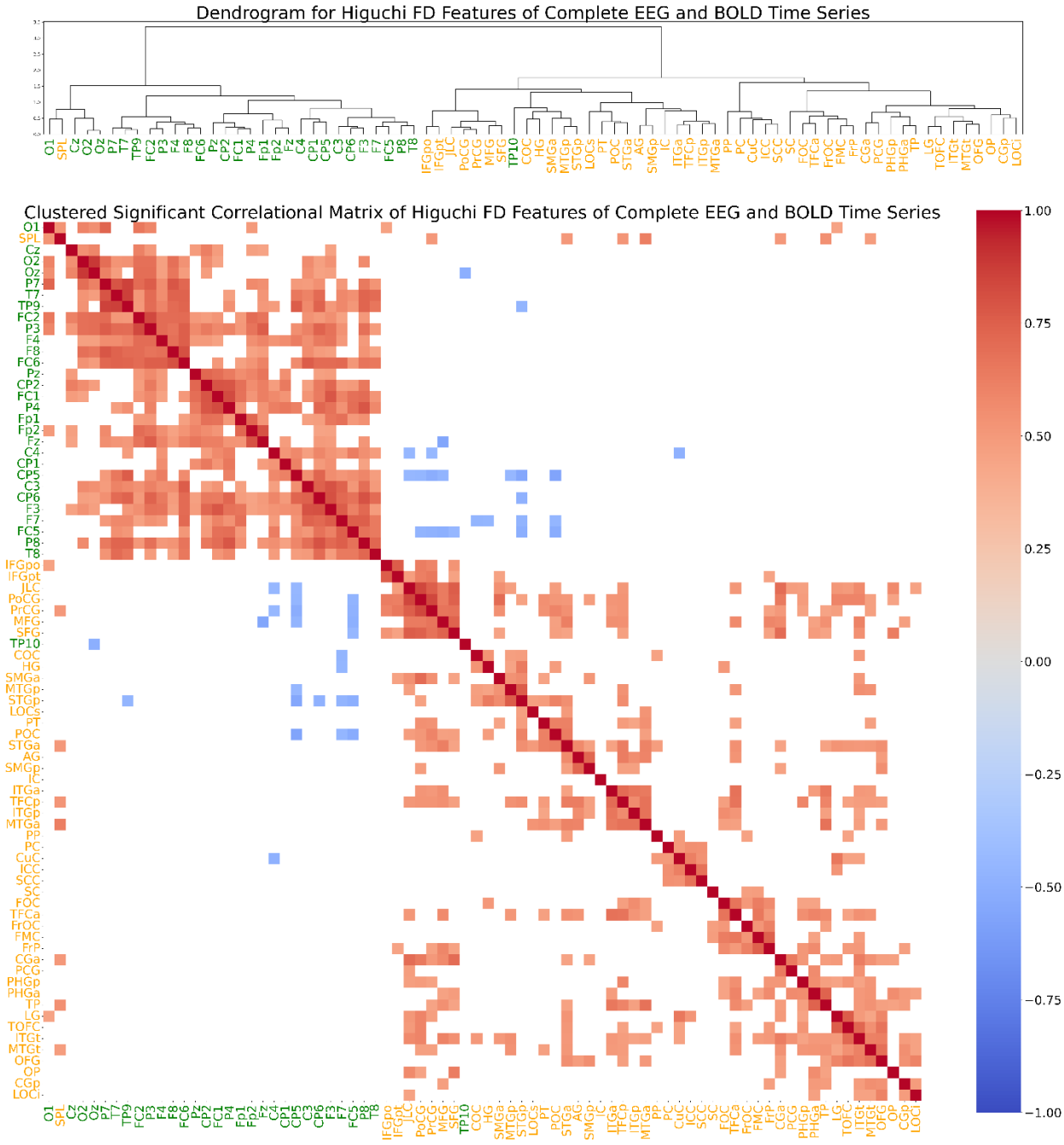
Dendrogram and Clustered Correlational Matrix After Hierarchical Clustering Analysis for DFA Features of Complete EEG and BOLD Time Series



Note. Non-significant correlations were masked. EEG channels were marked with green and fMRI ROIs in orange. The full ROI names were shared in the notes of Table F2 in Appendix F.

Figure 7

Dendrogram and Clustered Correlational Matrix After Hierarchical Clustering Analysis for Higuchi Fractal Dimension Features of Complete EEG and BOLD Time Series



Note. Non-significant correlations were masked. EEG channels were marked with green and fMRI ROIs in orange. The full ROI names were shared in the notes of Table F2 in Appendix F.

detected between Higuchi FD features from either the frontal or parietal clusters of EEG electrodes and ROIs associated with the DMN.

Between-Imaging Modality Correlational Analysis of Entropy and Fractal Features in Sleep Stage-Specific EEG and BOLD Time Series Segments

For the sleep stage-specific EEG and BOLD features the correlations were also assessed. Throughout the SampEn values from the wakeful stage 79 significant correlations were found between EEG channels and fMRI ROIs. Notably, within the frontoparietal attentional network (FAN), significant correlations included F7 with Middle Frontal Gyrus (MFG), $r(165) = -.198$, $p = .040$), and Supramarginal Gyrus, anterior division (SMGa), $r(165) = -.201$, $p = .036$, FC5 with the MFG, $r(165) = -.200$, $p = .037$, and SMGa, $r(165) = -.196$, $p = .041$, as well as P8 with SMGa, $r(165) = -.209$, $p = .029$. For the default mode network (DMN), the positive correlation of the Angular Gyrus (AG) with FC2 stood out, $r(165)$ $p = .011$, along with the negative correlation between the Cingulate Gyrus and channels P4, $r(165) = -.197$, $p = .040$, and Pz $r(165) = -.194$, $p = .044$. During the N-REM 1 and N-REM 2 stage for SampEn, two and one significant cross-modality correlations were obtained, respectively. Nevertheless, none of these correlations were between the frontal or parietal clusters of EEG electrodes and fMRI ROIs associated with the FAN or DMN.

The analysis of ApEn during the wakeful stage revealed 30 significant between-modality correlations. Considering the FAN, correlations included F7 with the MFG, $r(165) = -.215$, $p = .034$, FC2 with the Insular Cortex (IC), $r(165) = -.247$, $p = .010$, and P8 with SMGa, $r(165) = -.256$, $p = .007$. In the N-REM 1 stage, only one significant correlation was identified for ApEn between EEG channels and fMRI ROIs, which was not between frontal or parietal EEG

electrodes and fMRI ROIs associated with the FAN or DMN. Likewise, regarding the ApEn features in the N-REM 2 stage, there was a single significant between-modality correlation, posterior division's (CGp) = .236, which did not involve frontal or parietal EEG channels, nor fMRI ROIs associated with the FAN or DMN.

For HE in the wakeful stage, the number of significant between EEG and BOLD feature correlations was 44. Within the context of the FAN, significant correlations between HE features of EEG channels and fMRI ROIs included F8 with the Supramarginal Gyrus, posterior division (SMGp), $r(165) = -.220$, $p = .027$, FC6 with the SMGa, $r(165) = -.229$, $p = .019$, and IC with both P7, $r(165) = .265$, $p = .004$, and P8, $r(165) = .211$, $p = .037$. In addition, a significant correlation was observed between MFG and P7, $r(165) = .250$, $p = .008$. Despite these significant findings within the FAN, no significant correlations were found between HE features of the frontal or parietal clusters of EEG electrodes and ROIs associated with the default mode network (DMN). Furthermore, HE features in the N-REM 1 stage had six significant between EEG and fMRI correlations. Importantly, none of these significant correlations were between frontal or parietal clusters of EEG electrodes and ROIs linked to the FAN or DMN. Similarly, for the HE features extracted from N-REM 2 time series, significant correlations were found in six instances between EEG channels and fMRI ROIs. Nonetheless, similar to the findings in N-REM 1, there were no significant correlations between the frontal or parietal clusters of EEG electrodes and ROIs linked to the DMN or FAN.

In wakeful stage-based alpha exponent features from DFA, 55 significant correlations were shown between EEG channel and fMRI ROI features. Among these correlations, FAN-related findings were especially prominent in SMGp. In particular, this was observed in the case of channels F3, $r(165) = -.202$, $p = .041$, F4, $r(165) = -.205$, $p = .036$, F8, $r(165) = -.225$, $p =$

.018, FC5, r(165) = -.215, p = .026, and FC6, r(165) = -.239, p = .010, all showing significant negative correlations with SMGp. Besides, other significant correlations included FC1 with the Juxtapositional Lobule Cortex (JLC), which was formerly the Supplementary Motor Cortex, r(165) = .204, p = .038, JLC with P4, r(165) = .204, p = .037, and F8 with the Superior Parietal Lobule (SPL), r(165) = -.227, p = .016, which are also linked to the FAN. Two significant correlations that can be linked to the DMN were also demonstrated, with AG showing negative correlations with F8, r(165) = -.241, p = .009, and FC6 r(165) = -.216, p = .025. In N-REM 1 stage, only one significant correlation was found between the two types of sources. It is important to note that this correlation was not found between the frontal or parietal clusters of EEG electrodes and ROIs associated with the FAN or DMN. However, from the seven significant correlations between DFA features from EEG and BOLD during N-REM 2, one relationship related to the FAN that stood out was between P8 and SMGa, displaying a strong positive correlation r(28) = .584, p = .023.

Ultimately, in the correlational analysis of HFD features during the wakeful stage, 78 significant correlations were identified between EEG channels and fMRI ROIs. A series of significant cross-modality correlations were observed within the FAN, particularly concerning the SMGa. These included negative correlations of the SMGa with channels, such as F7, r(165) = -.227, p = .016, FC1, r(165) = -.212, p = .027, FC5, r(165) = -.256, p = .005), P4, r(165) = -.230, p = .014, and P8, r_s(165) = -.197, p = .043. However, in the N-REM 1 stages, although 11 correlations between EEG channels and fMRI ROIs reached statistical significance, none demonstrated significant relationships between EEG channels from the frontal or parietal regions and ROIs commonly found in FAN or DMN. Lastly, the HFD features showed only two significant between EEG channel and fMRI ROI correlations in N-REM 2. Similar to the

findings for HFD features in the N-REM 1 stage, these significant correlations did not involve the frontal or parietal EEG clusters or brain regions associated with the FAN or DMN.

Analysis of Variance (ANOVA) for Entropy and Fractal Features Across Sleep Stages in Sleep Stage-Specific EEG and BOLD Time Series Segments

The repeated measures ANOVAs and Friedman tests carried out for sleep-stage specific entropy and fractal dimension features averaged over all EEG channels, as well as the same features averaged over all fMRI ROIs, found significant variance in the means of all features across sleep stages, as shown in Table 1 and 2 and Figure 8. More specifically, in EEG-based features, the rANOVAs indicated significant differences for SampEn and ApEn means across sleep stages, with $F(2, 62) = 19.695$, $p < .001$, $\eta^2 = .178$ for SampEn, and $F(2, 62) = 17.199$, $p < .001$, $\eta^2 = .164$ for ApEn. Likewise, the analyses found significant differences in the means of HFD values across sleep stages, with $F(2, 62) = 17.659$, $p < .001$, $\eta^2 = .183$. As for the HE and alpha exponents from DFA, where the homogeneity of variances assumptions were violated, the Friedman tests also showed significant differences across sleep stages, with Hurst results at $\chi^2(2) = 22.750$, $p < .001$, Kendall's $W = .245$, and DFA at $\chi^2(2) = 17.438$, $p < .001$, Kendall's $W = .188$.

In the following step, the post-hoc Tukey HSD and Nemenyi tests provided further information about the specific pairwise comparisons. Considering SampEn, a significant difference was found between N-REM 1 and 2 ($p < .001$, 95% C.I. = [-0.07, -0.02]), with lower average in N-REM 2, as well as N-REM 2 and wakefulness ($p = .035$, 95% C.I. = [0.00, 0.05]), again with lower average in N-REM 2. However, no significant difference was observed between N-REM 1 and wakefulness ($p = .130$, 95% C.I. = [-0.04, 0.00]). Likewise, significant differences were found between the means of ApEn during N-REM 1 and 2 ($p < .001$, 95% C.I.

Table 1

Repeated Measures ANOVA Results for Sleep Stage-Specific Entropy and Fractal Features Averaged Over all EEG Channels

Feature	DFn	DFd	F (or χ^2)	p[a]	η^2 (or Kendall's W)
SampEn	2	62	19.695	<.001	.178
ApEn	2	62	17.199	<.001	.164
Hurst	2	62	22.750[b]	<.001	.245[c]
DFA	2	62	17.438[b]	<.001	.188[c]
Higuchi FD	2	62	17.659	<.001	.183

Note. DFn = degrees of freedom for the nominator, DFd = degrees of freedom for the denominator, a = p-value corrected with Benjamin-Hochberg procedure, b = χ^2 (chi-square) from Friedman test, c = effect size with Kendall's W.

Table 2

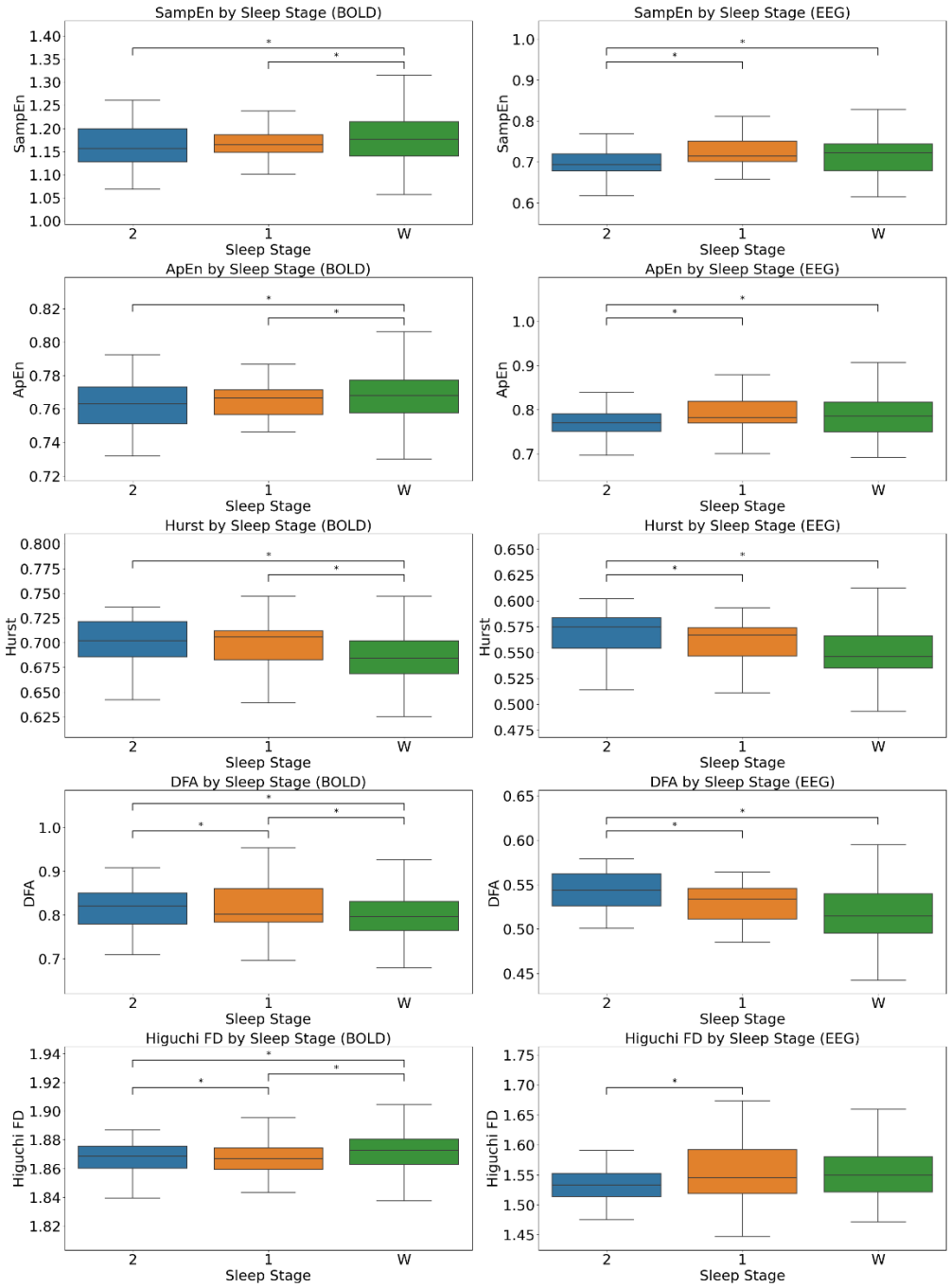
Repeated Measures ANOVA Results for Sleep Stage-Specific Entropy and Fractal Features Averaged Over all fMRI ROIs

Feature	DFn	DFd	F	p[a]	η^2
SampEn	2	62	11.599	<.001	.190
ApEn	2	62	6.213	.004	.116
Hurst	2	62	33.174	<.001	.381
DFA	2	62	17.969	<.001	.251
Higuchi FD	2	62	14.135	<.001	.214

Note. DFn = degrees of freedom for the nominator, DFd = degrees of freedom for the denominator, a = p-value corrected with Benjamin-Hochberg procedure.

Figure 8

Means Across Sleep Stages for Sleep Stage-Specific Entropy and Fractal Features Averaged Over all EEG Channels and fMRI ROIs with Post-Hoc Pairwise Comparison Results



Note. The asterisks indicate pairs from the post-hoc Tukey and Nemenyi tests found to be statistically significant with 5% chance level.

= [-0.06, -0.02]), with lower mean values in N-REM 2, and N-REM 2 and wakefulness (p = .049, 95% C.I. = [0.00, 0.04]), also having lower mean values in N-REM 2, but not between N-REM 1 and wakefulness (p = .156, 95% C.I. = [-0.04, 0.00]). Moreover, among the average HE values, the Nemenyi test found significant differences between N-REM 1 and 2 (p = .016); and N-REM 2 and wakefulness (p = .001), with higher mean values in N-REM 2, but no differences between N-REM 1 and wakefulness (p = .112). Similarly, for DFA, significant differences were observed between N-REM 1 and 2 (p = .024), with higher mean values in N-REM 2; and N-REM 2 and wakefulness (p = .001), with higher mean values in N-REM 2, while there was no difference between N-REM 1 and wakefulness (p = .291). Lastly, for HFD means, significant differences were observed between N-REM 1 and 2 (p < .001, 95% C.I. = [-0.05, -0.02]), with lower mean values in N-REM 2, but not between N-REM 1 and wakefulness (p = .079, 95% C.I. = [-0.03, 0.00]) or N-REM 2 and wakefulness (p = .050, 95% C.I. = [0.00, 0.04]).

In a similar manner, the repeated measures ANOVAs conducted for sleep-stage specific BOLD-based entropy and fractal dimension features averaged over all fMRI ROIs revealed significant differences in the means of all features across sleep stages, as depicted in Table 2 and Figure 8. More precisely, the analyses showed significant differences in the means of both SampEn, $F(2, 62) = 11.599$, $p < .001$, $\eta^2 = .190$, as well as ApEn, $F(2, 62) = 6.213$, $p = .004$, $\eta^2 = .116$, across wakefulness and sleep stages. The Hurst Exponent (HE) and alpha exponents from the Detrended Fluctuation Analysis (DFA) also displayed significant differences across sleep stages, with HE results at $F(2, 62) = 33.174$, $p < .001$, $\eta^2 = .381$, and DFA at $F(2, 62) = 17.969$, $p < .001$, $\eta^2 = .251$. Furthermore, significant differences were found in the means of Higuchi's Fractal Dimension (HFD) values across sleep stages, with $F(2, 62) = 14.135$, $p < .001$, $\eta^2 = .214$.

Following this, the post-hoc Tukey HSD tests provided detailed insights into the specific pairwise comparisons. For instance, a significant difference was observed between the mean of SampEn during N-REM 1 and wakefulness ($p < .001$, 95% C.I. = [0.02, 0.06]), as well as between N-REM 2 and wakefulness ($p = .001$, 95% C.I. = [0.01, 0.05]), with higher mean value in wakefulness in both cases. However, no significant difference was found between N-REM 1 and 2 ($p = .851$, 95% C.I. = [-0.02, 0.02]). For ApEn, significant differences were identified between N-REM 1 and wakefulness ($p = .016$, 95% C.I. = [0.00, 0.01]), and between N-REM 2 and wakefulness ($p = .006$, 95% C.I. = [0.00, 0.01]), with higher means in wakefulness in both comparisons. There was no significant difference between N-REM 1 and 2 ($p = .927$, 95% C.I. = [-0.01, 0.01]). In the case of HE, significant differences were found between N-REM 1 and wakefulness ($p < .001$, 95% C.I. = [-0.03, -0.02]), with higher average HE in N-REM 1, and between N-REM 2 and wakefulness ($p < .001$, 95% C.I. = [-0.03, -0.01]), also with higher mean HE in N-REM 2. Similarly to the previous cases, no significant difference was present between N-REM 1 and 2 ($p = .120$, 95% C.I. = [-0.02, 0.00]). Besides, DFA showed significant differences between N-REM 1 and 2 ($p = .047$, 95% C.I. = [-0.04, 0.00]), with lower mean value in N-REM 1; between N-REM 1 and wakefulness ($p < .001$, 95% C.I. = [-0.07, -0.03]), with wakefulness having lower mean, as well as between N-REM 2 and wakefulness ($p = .006$, 95% C.I. = [-0.05, -0.01]), with wakefulness having a lower mean value once again. Finally, for HFD, significant differences were detected between N-REM 1 and 2 ($p = .026$, 95% C.I. = [0.00, 0.01]), with higher average in N-REM 1; between N-REM 1 and wakefulness ($p < .001$, 95% C.I. = [0.01, 0.02]), with higher mean value in wakefulness; and between N-REM 2 and wakefulness ($p = .049$, 95% C.I. = [0.00, 0.01]), with higher mean value in wakefulness.

Machine Learning Classification of Sleep Stages Trained on Entropy and Fractal Features from Sleep Stage-Specific Segments of EEG, BOLD, and Combined EEG and BOLD

Despite the limited number of sleep-stage specific feature values, the performance of models trained on entropy and fractal features from sleep stage-specific segments of EEG, BOLD, and combined EEG and BOLD signals was still evaluated as can be seen in Figure 9. For models trained on purely EEG-based features, both the RF and SVM models achieved accuracies of 76.1%. The RF model had a precision of 0.579, recall of 0.584, and F1-score of 0.577, and AUC values of 0.60 for N-REM 1, 0.83 for N-REM 2, and 0.75 for wakefulness. Meanwhile, the SVM model showed different results by primarily predicting the majority class, which were wakeful stages. Specifically, the SVM classifier had a precision of 0.254, recall of 0.333, and F1-score of 0.288, and AUC values of 0.50 for N-REM 1, 0.50 for N-REM 2, and 0.50 for wakefulness.

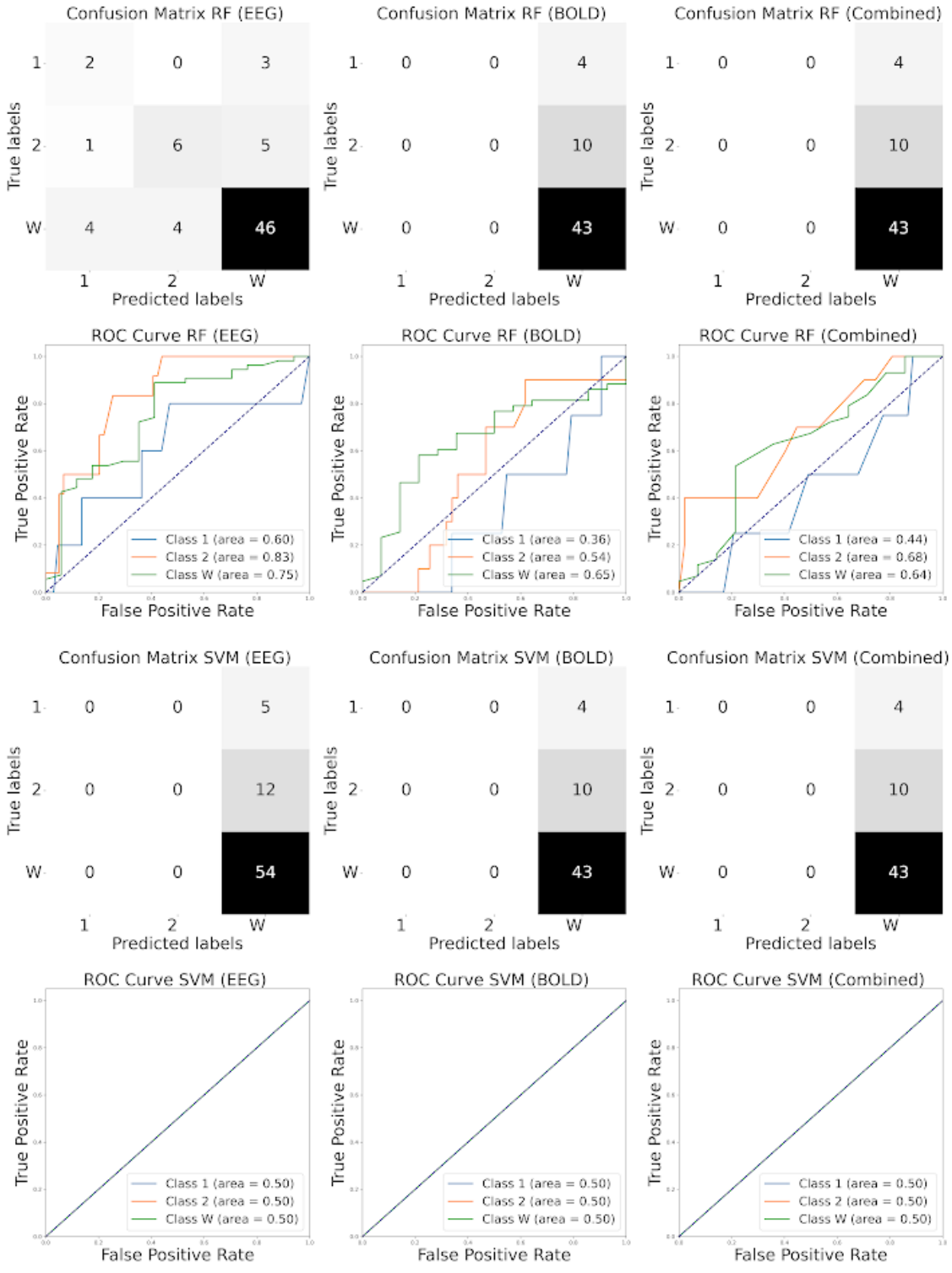
Considering the models trained on BOLD-based features, the performance was slightly lower. The RF model achieved an accuracy of 75.4%, precision of 0.251, recall of 0.333, and F1-score of 0.287. The AUC values were 0.36 for N-REM 1, 0.54 for N-REM 2, and 0.65 for wakefulness. Likewise, the SVM model also achieved an accuracy of 75.4%, with precision of 0.251, recall of 0.333, and F1-score of 0.287. However, for the SVM model, the AUC values were 0.50 for N-REM 1, 0.50 for N-REM 2, and 0.50 for wakefulness. In these cases as well, the models mainly learned to predict the majority class, which were wakeful stages.

Ultimately, for the models trained on sleep stage-specific entropy and fractal features from both EEG channels and fMRI ROIs, the combined features did not show improvement over the individual modalities. More precisely, the RF model trained on combined features achieved an accuracy of 75.4%, precision of 0.251, recall of 0.333, and F1-score of 0.287, with AUC

values of 0.44 for N-REM 1, 0.68 for N-REM 2, and 0.64 for wakefulness. Similarly, the SVM model trained on combined features also had an accuracy of 75.4%, precision of 0.251, recall of 0.333, and F1-score of 0.287, with AUC values of 0.50 for N-REM 1, 0.50 for N-REM 2, and 0.50 for wakefulness.

Figure 9

Confusion Matrices and ROC Curves for the Machine Learning Models Trained on EEG-Based, BOLD-Based, and Combined Entropy and Fractal Features from Sleep Stage Specific Segments.



Discussion

In light of the above, while the function of sleep and its stages remains unknown, the advent of simultaneous EEG-fMRI has enabled researchers to study brain processes underlying sleep with high spatio-temporal resolution. This is also highly relevant because millions of individuals are affected by sleep disorders and this advancement may provide new approaches to better understand and treat these conditions. Nevertheless, which characteristics of concurrently recorded EEG and BOLD time series are informative about sleep and its stages is largely unmapped. Specifically, how non-linear features, such as entropy and fractal features, relate across simultaneously collected EEG and BOLD time series during sleep, as well as the informativeness of these features for distinguishing sleep stages and classifying them with machine learning models, has not yet been tested in previous literature. Thus, this study investigated how individual entropy and fractal features relate between complete EEG signals from frontoparietal electrodes and same features of simultaneously recorded BOLD time series from fMRI ROIs associated with the FAN and DMN during sleep. Similarly, it was also examined how each of these features correlates between frontoparietal EEG channels and fMRI ROIs related to the DMN and FAN but in sleep stage-specific segments of simultaneous EEG-fMRI data. Finally, this study also tested how effective these features are in distinguishing and classifying wakeful and sleep stages with ANOVAs and supervised machine learning models, respectively.

Considering the first research question, between-modality correlations were limited. Specifically, for Sample Entropy (SampEn), three significant between-modality correlations were found, but none were between EEG electrodes in frontal or parietal clusters and ROIs linked to the frontoparietal attentional network (FAN) or default mode network (DMN).

Approximate Entropy (ApEn) showed 83 between-modality correlations, including significant negative correlations involving the Middle Frontal Gyrus (MFG) and F8, FC6, and P7, which are related to the FAN, and the Angular Gyrus (AG) and F8, FC6, that are DMN-related. The Hurst Exponent (HE) and the alpha exponents from Detrended Fluctuation Analysis (DFA) each had only one between-modality correlation, with no significant correlations between these features for fMRI ROIs linked to FAN or DMN and EEG channels from the frontal or parietal regions. Lastly, the analysis for Higuchi Fractal Dimension (HFD) revealed 26 between-modality correlations, with significant negative correlations related to the FAN between FC5 and MFG, and Fz and MFG, but no significant correlations were found for HFD features of fMRI ROIs linked to the DMN. Furthermore, taking into account the correlations of these features in sleep stage-specific segments, the findings indicated that, with the exception of ApEn, there were a higher number of significant correlations between all the other features from the frontal and parietal EEG channels and fMRI ROIs linked to the FAN and DMN during wakeful stages compared to the full time series. Among the N-REM sleep stages, the only case where there were a higher number of significant correlations between features of the frontal and parietal EEG channels and fMRI ROIs linked to the FAN and DMN relative to the features from the full time series was for DFA, which had one significant cross-modality correlation between SMGa and P8 in the N-REM 2 stage. In the other sleep stages, either there were no significant correlations between features of the frontal and parietal EEG channels and fMRI ROIs linked to the FAN and DMN, or there were fewer correlations relative to the full time series. Regarding the question of whether sleep stage-specific entropy and fractal features from EEG and BOLD time series can distinguish between sleep stages, the rANOVAs and Friedman tests found significant differences in all features across sleep stages. More specifically, post-hoc analyses showed that the mean

values of several EEG-based features, such as SampEn and ApEn, were significantly lower in N-REM 2 compared to both N-REM 1 and wakefulness, while HE and DFA values were higher in N-REM 2 compared to both N-REM 1 and wakefulness. Moreover, HFD also showed a lower mean in N-REM 2 compared to N-REM 1 but did not significantly differ between wakefulness and either N-REM 1 or N-REM 2. Although the average scores of EEG-based Hurst and DFA features were significantly different between N-REM 1 and N-REM 2, as well as N-REM 2 and wakefulness, an opposite trend was observed; the means of entropy features and HFD decreased from wakefulness to N-REM 1 and to N-REM 2 stages, while HE and DFA values increased. Similarly, in BOLD-based features, SampEn, ApEn, and HFD values had incrementally lower means from wakefulness to N-REM 2 sleep, while for HE and DFA, the means increased towards N-REM 2 sleep. Furthermore, significant differences were found in SampEn, ApEn, and Hurst between wakefulness and both N-REM 1 and N-REM 2, but without a significant difference between N-REM 1 and N-REM 2. Importantly, the average DFA and mean HFD values across all fMRI ROIs distinguished between all three sleep stages.

These results suggest that the relationships between EEG and BOLD signals are limited, complex and region-specific. Furthermore, the expected statistically significant negative correlations between HE features in the frontal and parietal EEG channels and fMRI ROIs related to FAN and DMN, as inspired by the findings of Tagliazucchi, Von Wegner, et al. (2013), were not observed. Although the present study had a different methodology and focus in both signals compared to the one of Tagliazucchi, Von Wegner, et al. (2013), these results might indicate the lack of a negative relationship between Hurst exponents of the frontoparietal EEG clusters and fMRI ROIs linked to the FAN and DMN. More likely, however, the differences in the approaches between the two studies or the limitations of the current study, which will be

discussed in the following sections, may have influenced these findings. Nevertheless, these results provide partial support at least for the hypothesis about ApEn and HFD showing statistically significant relationships across frontal and parietal EEG channels, and fMRI ROIs related to FAN and DMN. Moreover, in the sleep stage-specific segments the findings imply that correlations between these entropy and fractal features from the frontal and parietal EEG clusters and fMRI ROIs associated with the FAN and DMN are particularly apparent during wakeful stages. In contrast, the relationships between these features are less pronounced in other sleep stages, which may suggest that their dynamics between EEG and BOLD signals could be dependent on sleep stage. This, in turn, underlines the importance of conducting sleep stage-specific analyses to better understand the complex interactions between entropy and fractal features of EEG and BOLD signals. Considering the rANOVA and Friedman test results, it seems that entropy and fractal features from both EEG and BOLD data can distinguish between certain sleep stages and wakefulness. In fact, some fractal features, such as the alpha exponent from DFA and HFD features of BOLD time series, can differentiate between all three sleep stages examined. This aligns with previous findings by Acharya et al. (2015) and Chouvarda et al. (2011), which found that ApEn, SampEn, Hurst exponent, DFA, and HFD features from EEG data can distinguish sleep stages. In addition, these findings are extended by the current research even to BOLD-based features also delineating sleep stages, which is a new insight. Moreover, pairwise comparisons of entropy and HFD features further support previous research. It was hypothesized, based on prior studies, that all entropy and fractal features would have higher values during wakefulness and lower values in N-REM. This pattern was observed for entropy features and HFD, corroborating findings from Burioka et al. (2005), Chen (2017), De Miras et al. (2019), and Ge et al. (2007) in EEG data. Interestingly, the same trend was demonstrated in

simultaneous BOLD-based data, which is a new finding. Nevertheless, for HE and DFA values, an opposite trend was observed, with lower values in wakefulness and incrementally higher values in N-REM 1 and N-REM 2. Even though this contradicts the initial hypotheses, it was not entirely unexpected. As pointed out earlier, studies have shown inconsistent results regarding whether alpha exponents from DFA in EEG increase or decrease from wakefulness to deeper N-REM stages (Goshvarpour & Abbasi, 2013; Lee et al., 2001). This study, however, demonstrates an increasing trend in alpha exponents from wakefulness to deeper N-REM stages, not only in EEG but also in simultaneous BOLD time series, supporting the increasing nature of alpha exponents. Similarly, while previous literature on both EEG and fMRI suggested that HE values are lower in N-REM stages than in wakefulness, the results of the current research align with those of Weiss et al. (2009), showing an increase in HE values from wakefulness to deeper N-REM stages (Acharya et al., 2005; Tagliazucchi, Von Wegner, et al., 2013). This is corroborated by the concurrently obtained BOLD-based HE features. In summary, while some features showed trends opposite to those hypothesized based on the literature, this study supports the ability of entropy and fractal features in distinguishing sleep stages. Likewise, the inclusion of BOLD time series features alongside EEG features provides new insights and highlights the need for further investigation to resolve these inconsistencies.

As for the machine learning models, the best model overall turned out to be the RF classifier trained on purely EEG-based entropy and fractal features, which had a similar accuracy to the SVM model using EEG features but better AUC scores. In contrast, the RF and SVM models trained on BOLD-based entropy and fractal features had lower performance than the same models trained on EEG-based features. Moreover, the same accuracy scores were found for models trained on BOLD-based features and features from both modalities. Consequently,

integrating features from both EEG and BOLD did not show clear improvements, and since the predominant prediction of all models was wakeful stages, this indicates a need for more balanced data across all sleep stages. In addition, while the accuracy of the models with features from both modalities are similar to those of previous deep learning algorithms trained on EEG from simultaneous EEG-fMRI data, the models from the current study perform less well than expected (Zou et al., 2022). This is clear when compared to the previously discussed deep learning and simpler machine learning models for sleep staging that were trained exclusively on features of unimodal EEG time series (Boostani et al., 2017; Fiorillo et al., 2019; Gaiduk et al., 2023; Lambert & Peter-Derex, 2023; Li et al., 2022; Phan & Mikkelsen, 2022; Sri et al., 2022). Having that said, these findings suggest that while entropy and fractal features can help distinguish between sleep stages through ANOVAs, the machine learning models do not seem to perform as well as hypothesized, especially when combining entropy and fractal features from both EEG and BOLD data collected simultaneously. This might be due to several factors that could be considered among the limitations in this study.

One of the most critical limitations of this study is the lack of data for the studied sleep stages, as well as the large imbalance with predominantly wakeful data in the dataset. These may be attributable not only to the natural distribution of sleep stages but also probably to the short sleep scanning sessions, as well as the time series length needed for entropy and fractal feature calculations, which result in inadequate data for proper machine learning applications. To be more precise, even the simplest supervised machine learning models usually require at least 1000 samples per class for stable performance (Ramezan et al., 2021). However, this dataset had only 218 samples, with distributions of 167 from wakeful stages, 21 from N-REM 1, and 30 from N-REM 2. Moreover, the noisy data from the simultaneous EEG-fMRI recordings is another factor

to take into consideration as it made the analysis and interpretation of the results far more challenging (Bullock et al., 2021). More precisely, MRI-related artifacts heavily contaminated the EEG data, making it intrinsically biased and difficult to appropriately correct. Likewise, a related concern is the potential overcorrection of the EEG signal during pre- and postprocessing, especially when applying artifact subspace reconstruction (ASR). While this seems to be a realistic concern, artifacts were still sparsely detected in the data after using ASR. Similarly, there could also be potential overcorrection in the pre- and postprocessing of the fMRI data. Apart from these factors, the parameter selection for feature calculations for SampEn, ApEn, Hurst Exponent, alpha exponent for DFA, and Higuchi FD may have been rather arbitrary based on literature. The algorithm for k_{max} optimization in Higuchi FD also presents limiting factors, such as the maximum k_{max} having been set to 64 due to constraints in computational resources. The sliding window method for calculating sleep stage-specific features is yet another methodological limitation. As no relevant guidelines have been found in earlier studies, the step size for the sliding window was set to 100 based on additional tests in this study. These tests demonstrated that a smaller step size, such as one, does not result in windows with more informative features. Although this does provide some support for the decision about step sizes, it is important to acknowledge that the step-size selection process still contains elements of subjectivity that can also influence the final features and how they are interpreted. Finally, this research focused exclusively on non-linear features. However, entropy and fractal features alone, without additional features such as spectral properties, may be less informative and predictive in the context of sleep staging than expected in this study (Fell et al., 1996). These limitations point out the need for collecting larger datasets, as well as developing more robust and refined

analytical approaches to better understand the relationships between features of simultaneously collected EEG and BOLD signals across different sleep stages.

In spite of these limitations, the current study has several strengths, too. Firstly, based on the reviewed literature, this study may represent the first-ever concurrent analysis of SampEn, ApEn, Hurst Exponent, alpha exponent from DFA, and Higuchi FD in simultaneous EEG-fMRI data, providing a novel and detailed approach to understanding how entropy and fractal features relate within these modalities and how informative they are regarding sleep stages. More specifically, this study offers an unprecedented, multi-faceted paradigm for entropy- and fractal feature-based sleep analysis, incorporating correlational analyses for both full and sleep stage-specific segments, ANOVA, and machine learning methods. Yet another strength of the current study is that, with the exception of manually selecting the cutoff parameter of ASR in the preprocessing, as well as disabling the motion artifact segment censoring in the postprocessing, all pre- and postprocessing procedures were based on standardized and validated methods and parameters in an attempt to ensure replicability on other datasets. Assumptions were also rigorously tested at all steps of the analysis, further strengthening the study's findings. In fact, despite the computational constraints and the lack of precedents, even the k_{max} optimization for HFD was carefully considered, which demonstrates a new and thoughtful approach to parameter selection. Thus, all these aspects strengthen the robustness and novelty of the present study.

These results have important implications across many domains. On the one hand, they contribute to our theoretical understanding of entropy and fractal features in neurophysiological signals like EEG and BOLD during sleep. In particular, despite the challenges posed by severely contaminated and insufficient data for appropriate machine learning classification, the study demonstrated statistically significant correlations across several entropy and fractal features from

both full-length and sleep stage-specific EEG and BOLD time series, showing multiple links between frontoparietal EEG channels and the FAN and DMN. Besides, the findings also suggest that wakeful sleep stages can be distinguished based on entropy and fractal features as demonstrated by the results of the rANOVAs, Friedman tests, and post-hoc analyses. Even though collecting simultaneous EEG-fMRI data with minimal artifacts is still inherently difficult, future research should replicate or extend this study by examining other non-linear or various time, frequency, or time-frequency domain features in larger and cleaner datasets (Bullock et al., 2021). On the other hand, these findings may have practical applications in clinical, pharmaceutical, and medical settings as well. Particularly, using these non-linear features from simultaneous EEG-fMRI, researchers and practitioners may gain insights into the underlying mechanisms of sleep disorders afflicting millions of people in our society (Benjafield et al., 2019; Kerkhof, 2017; Sateia, 2014). With that in mind, future research should look at the applicability of the paradigm presented in this study, as well as the predictive power of entropy and fractal features from simultaneous EEG-fMRI in patients with sleep disorders.

To conclude, the purpose of the present study was to uncover the within- and cross-modality correlations between several entropy and fractal features in simultaneous EEG-fMRI data, their effectiveness in distinguishing sleep stages, as well as the performance of machine learning models trained on these features for sleep stage classification. In spite of having mixed results and many limitations, including insufficient and noisy data, the methods and findings of this research provide a new approach to sleep analysis, as well as perhaps the diagnosis of sleep disorders. Future research could attempt to replicate these preliminary findings in larger and cleaner datasets. Alternatively, new studies could extend the scope of the current research by using other non-linear, time, frequency, or time-frequency domain features. Thus, this work,

among other related publications, may be one of the forerunners to many future studies that could focus on entropy and fractal features in simultaneous EEG-fMRI data in order to gain a better understanding of the nature and function of sleep and sleep stages.

Data and code availability

The Python and MATLAB codes will become available in the GitHub repository of this project at https://github.com/hunorbartalis/FractalEntropySleepEEGfMRI once permission is granted. To request access to the data and further references, please contact the author.

Acknowledgments

The main supervisor of this thesis project, Dr. Wouter De Baene from the Department of Cognitive Neuropsychology at Tilburg University, is much appreciated by the author of this thesis for his outstanding mentorship, exceptional suggestions, and unwavering support throughout this project. Likewise, special thanks are extended to Dr. Marijn van Wingerden from the Department of Cognitive Science and Artificial Intelligence at Tilburg University, who took the role of a second supervisor and provided valuable insights and constant support during this project. Lastly, special thanks go to Dr. Yameng Gu from United Imaging in North America for supplying the dataset and valuable guidance, Dr. Jacobus A. de Zwart from the National Institutes of Health in the United States of America for software contributions, and Federico Zamberlan from the Department of Cognitive Science and Artificial Intelligence at Tilburg University for his insights.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*. https://doi.org/10.3389/fninf.2014.00014

Acharya, U. R., Bhat, S., Faust, O., Adeli, H., Chua, E. C. P., Lim, W. J. E., & Koh, J. E. W. (2015). Nonlinear dynamics measures for automated EEG-based sleep stage detection. *European Neurology*, *74*(5–6), 268–287. https://doi.org/10.1159/000441975

Acharya, U. R., Faust, O., Kannathal, N., Chua, T., & Laxminarayan, S. (2005). Non-linear analysis of EEG signals at various sleep stages. *Computer Methods and Programs in Biomedicine*, *80*(1), 37–45. https://doi.org/10.1016/j.cmpb.2005.06.011

Agarwal, R., & Gotman, J. (2001). Computer-assisted sleep staging. *IEEE Transactions on Biomedical Engineering*, *48*(12), 1412–1423. https://doi.org/10.1109/10.966600

Aserinsky, E., & Kleitman, N. (1953). Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science*, *118*(3062), 273–274. https://doi.org/10.1126/science.118.3062.273

Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41. https://doi.org/10.1016/j.media.2007.06.004

Avants, B., Tustison, N. J., & Song, G. (2009). Advanced normalization tools (ANTS). *The Insight Journal*, *2*(365), 1–35. https://doi.org/10.54294/uvnhin

Bal, A., Ganguly, D., & Chatterjee, K. (2021). Stationarity and self-similarity determination of time series data using Hurst exponent and R/S ration analysis. In A. E. Hassanien, S.

Bhattacharyya, S. Chakrabati, A. Bhattacharya, & S. Dutta (Eds.), *Emerging technologies in data mining and information security* (Vol. 1300, pp. 601–612). Springer. https://doi.org/10.1007/978-981-33-4367-2_57

Bandettini, P. A., Wong, E. C., Hinks, R. S., Tikofsky, R. S., & Hyde, J. S. (1992). Time course EPI of human brain function during task activation. *Magnetic Resonance in Medicine*, *25*(2), 390–397. https://doi.org/10.1002/mrm.1910250220

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, *37*(1), 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042

Benjafield, A., Ayas, N., Eastwood, P. R., Heinzer, R., Ip, M. S., Morrell, M. J., Nuñez, C. M., Patel, S. R., Penzel, T., Pépin, J., Peppard, P. E., Sinha, S., Tufik, S., Valentine, K., & Malhotra, A. (2019). Estimation of the global prevalence and burden of obstructive sleep apnoea: A literature-based analysis. *The Lancet Respiratory Medicine*, *7*(8), 687–698. https://doi.org/10.1016/s2213-2600(19)30198-5

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Berger, H. (1929). Über das Elektrenkephalogramm des Menschen. *Archiv Für Psychiatrie Und Nervenkrankheiten*, *87*(1), 527–570. https://doi.org/10.1007/bf01797193

Betta, M., Handjaras, G., Leo, A., Federici, A., Farinelli, V., Ricciardi, E., Siclari, F., Meletti, S., Ballotta, D., Benuzzi, F., & Bernardi, G. (2021). Cortical and subcortical hemodynamic changes during sleep slow waves in human light sleep. *NeuroImage*, *236*, 118117. https://doi.org/10.1016/j.neuroimage.2021.118117

Blanca, M. J., Arnau, J., García-Castro, F. J., Alarcón, R., & Bono, R. (2023). Non-normal data in repeated measures ANOVA: Impact on type I error and power. *Psicothema*, *35*(1), 21–29. https://doi.org/10.7334/psicothema2022.292

Boostani, R., Karimzadeh, F., & Nami, M. (2017). A comparative review on sleep stage classification methods in patients and healthy individuals. *Computer Methods and Programs in Biomedicine*, *140*, 77–91. https://doi.org/10.1016/j.cmpb.2016.12.004

Boynton, G. M., Engel, S. A., Glover, G. H., & Heeger, D. J. (1996). Linear systems analysis of functional magnetic resonance imaging in human V1. *The Journal of Neuroscience*, *16*(13), 4207–4221. https://doi.org/10.1523/jneurosci.16-13-04207.1996

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145–1159. https://doi.org/10.1016/s0031-3203(96)00142-2

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/a:1010933404324

Brett, M., Markiewicz, C. J., Hanke, M., Côté, M.-A., Cipollini, B., McCarthy, P., Jarecka, D., Cheng, C. P., Halchenko, Y. O., Cottaar, M., Larson, E., Ghosh, S., Wassermann, D., Gerhard, S., Lee, G. R., Wang, H.-T., Kastman, E., Kaczmarzyk, J., Guidotti, R., ... Reddam, V. R. (2022). *Nipy/nibabel* (Version 4.0.0rc0) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.6617127

Bruce, E. N., Bruce, M. C., & Vennelaganti, S. (2009). Sample entropy tracks changes in electroencephalogram power spectrum with sleep state and aging. *Journal of Clinical Neurophysiology*, *26*(4), 257–266. https://doi.org/10.1097/wnp.0b013e3181b2f1e3

Bryce, R., & Sprague, K. (2012). Revisiting detrended fluctuation analysis. *Scientific Reports*, *2*(1). https://doi.org/10.1038/srep00315

Bullock, M., Jackson, G. D., & Abbott, D. F. (2021). Artifact reduction in simultaneous EEG-fMRI: A systematic review of methods and contemporary usage. *Frontiers in Neurology*, *12*. https://doi.org/10.3389/fneur.2021.622719

Burioka, N., Miyata, M., Cornélissen, G., Halberg, F., Takeshima, T., Kaplan, D. T., Suyama, H., Endo, M., Maegaki, Y., Nomura, T., Tsutsumi, Y., Kenji, N., & Shimizu, E. (2005). Approximate entropy in the electroencephalogram during wake and sleep. *Clinical EEG and Neuroscience*, *36*(1), 21–24. https://doi.org/10.1177/155005940503600106

Buzsáki, G., Anastassiou, C. A., & Koch, C. (2012). The origin of extracellular fields and currents — EEG, ECoG, LFP and spikes. *Nature Reviews Neuroscience*, *13*(6), 407–420. https://doi.org/10.1038/nrn3241

Cantero, J. L., Atienza, M., & Salas, R. M. (2002). Human alpha oscillations in wakefulness, drowsiness period, and REM sleep: Different electroencephalographic phenomena within the alpha band. *Neurophysiologie Clinique*, *32*(1), 54–71. https://doi.org/10.1016/s0987-7053(01)00289-1

Ceballos, R. F., & Largo, F. F. (2018). On the estimation of the Hurst exponent using adjusted rescaled range analysis, detrended fluctuation analysis and variance time plot: A case of exponential distribution. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1805.08931

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, G., Taylor, P. A., Reynolds, R. C., Leibenluft, E., Pine, D. S., Brotman, M. A., Pagliaccio, D., & Haller, S. P. (2023). BOLD response is more than just magnitude: Improving detection sensitivity through capturing hemodynamic profiles. *NeuroImage*, *277*, 120224. https://doi.org/10.1016/j.neuroimage.2023.120224

Chen, J. (2017). Characterizing sleep stages by the fractal dimensions of electroencephalograms. *Biostatistics and Biometrics Open Access Journal*, *2*(2). https://doi.org/10.19080/bboaj.2017.02.555584

Chouvarda, I., Rosso, V., Méndez, M. O., Bianchi, A. M., Parrino, L., Grassi, A., Terzano, M. G., & Cerutti, S. (2011). Assessment of the EEG complexity during activations from sleep. *Computer Methods and Programs in Biomedicine*, *104*(3), e16–e28. https://doi.org/10.1016/j.cmpb.2010.11.005

Cirelli, C., & Tononi, G. (2015). Cortical development, electroencephalogram rhythms, and the sleep/wake cycle. *Biological Psychiatry*, *77*(12), 1071–1078. https://doi.org/10.1016/j.biopsych.2014.12.017

Ciric, R., Rosen, A. F. G., Erus, G., Cieslak, M., Adebimpe, A., Cook, P. A., Bassett, D. S., Davatzikos, C., Wolf, D. H., & Satterthwaite, T. D. (2018). Mitigating head motion artifact in functional connectivity MRI. *Nature Protocols*, *13*(12), 2801–2826. https://doi.org/10.1038/s41596-018-0065-y

Ciric, R., Thompson, W. H., Lorenz, R., Goncalves, M., MacNicol, E. E., Markiewicz, C. J., Halchenko, Y. O., Ghosh, S. S., Gorgolewski, K. J., Poldrack, R. A., & Esteban, O. (2022). TemplateFlow: FAIR-sharing of multi-scale, multi-species brain models. *Nature Methods*, *19*(12), 1568–1571. https://doi.org/10.1038/s41592-022-01681-2

Ciric, R., Wolf, D. H., Power, J. D., Roalf, D. R., Baum, G. L., Ruparel, K., Shinohara, R. T., Elliott, M. A., Eickhoff, S. B., Davatzikos, C., Gur, R. C., Gur, R. E., Bassett, D. S., & Satterthwaite, T. D. (2017). Benchmarking of participant-level confound regression strategies for the control of motion artifact in studies of functional connectivity. *NeuroImage*, *154*, 174–187. https://doi.org/10.1016/j.neuroimage.2017.03.020

Clausius, R. (1867). *The mechanical theory of heat: With its applications to the steam-engine and to the physical properties of bodies* (T. Archer Hirst, Ed.). John van Voorst.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/bf00994018

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/10.1006/cbmr.1996.0014

Cox, R. W., & Hyde, J. S. (1997). Software tools for analysis and visualization of fMRI data. *NMR in Biomedicine*, *10*(4–5), 171–178. https://doi.org/10.1002/(SICI)1099-1492(199706/08)10:4/5%3C171::AID-NBM453%3E3.0.CO;2-L

Da Costa, M., Goldberger, A. L., & Peng, C. (2002). Multiscale entropy analysis of complex physiologic time series. *Physical Review Letters*, *89*(6). https://doi.org/10.1103/physrevlett.89.068102

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. *NeuroImage*, *9*(2), 179–194. https://doi.org/10.1006/nimg.1998.0395

Danker-Hopfe, H., Anderer, P., Zeitlhofer, J., Boeck, M., Dorn, H., Gruber, G., Heller, E., Loretz, E., Moser, D., Parapatics, S., Saletu, B., Schmidt, A., & Dorffner, G. (2009). Interrater reliability for sleep scoring according to the Rechtschaffen & Kales and the

new AASM standard. *Journal of Sleep Research*, *18*(1), 74–84. https://doi.org/10.1111/j.1365-2869.2008.00700.x

De Araújo, D. B., Tedeschi, W., Santos, A. C. D., Elías, J., Neves, U. P. C., & Baffa, O. (2003). Shannon entropy applied to the analysis of event-related fMRI time series. *NeuroImage*, *20*(1), 311–317. https://doi.org/10.1016/s1053-8119(03)00306-9

De Miras, J. R., Soler, F., Iglesias-Parro, S., Ibáñez-Molina, A. J., Casali, A. G., Laureys, S., Massimini, M., Esteban, F. J., Navas, J., & Langa, J. A. (2019). Fractal dimension analysis of states of consciousness and unconsciousness using transcranial magnetic stimulation. *Computer Methods and Programs in Biomedicine*, *175*, 129–137. https://doi.org/10.1016/j.cmpb.2019.04.017

Delgado-Bonal, A., & Marshak, A. (2019). Approximate entropy and sample entropy: A comprehensive tutorial. *Entropy*, *21*(6), 541. https://doi.org/10.3390/e21060541

Delignières, D., Ramdani, S., Lemoine, L., Torre, K., Fortes, M., & Ninot, G. (2006). Fractal analyses for 'short' time series: A re-assessment of classical methods. *Journal of Mathematical Psychology*, *50*(6), 525–544. https://doi.org/10.1016/j.jmp.2006.07.004

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dement, W. C., & Kleitman, N. (1957). Cyclic variations in EEG during sleep and their relation to eye movements, body motility, and dreaming. *Electroencephalography and Clinical Neurophysiology*, *9*(4), 673–690. https://doi.org/10.1016/0013-4694(57)90088-3

Duyn, J. H. (2012). EEG-fMRI methods for the study of brain networks during sleep. *Frontiers in Neurology*, *3*. https://doi.org/10.3389/fneur.2012.00100

Esteban, O., Markiewicz, C., Blair, R., Moodie, C., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*(1), 111–116. https://doi.org/10.1038/s41592-018-0235-4

Esteban, O., Markiewicz, C. J., Goncalves, M., Provins, C., Kent, J. D., DuPre, E., Salo, T., Ciric, R., Pinsard, B., Blair, R. W., Poldrack, R. A., & Gorgolewski, K. J. (2023). *fMRIPrep: a robust preprocessing pipeline for functional MRI* (Version 23.1.4) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.8206595

Esteller, R., Vachtsevanos, G., Echauz, J., & Litt, B. (2001). A comparison of waveform fractal dimension algorithms. *IEEE Transactions on Circuits and Systems*, *48*(2), 177–183. https://doi.org/10.1109/81.904882

Farag, A. F., & El-Metwally, S. M. (2012). Detrended fluctuation analysis features for automated sleep staging of sleep EEG. *International Journal of Systems Biology and Biomedical Technologies*, *1*(4), 47–59. https://doi.org/10.4018/ijsbbt.2012100104

Fell, J., Röschke, J., Mann, K., & Schäffner, C. (1996). Discrimination of sleep stages: A comparison between spectral and nonlinear EEG measures. *Electroencephalography and Clinical Neurophysiology*, *98*(5), 401–410. https://doi.org/10.1016/0013-4694(96)95636-9

Fiorillo, L., Puiatti, A., Papandrea, M., Ratti, P. L., Favaro, P., Roth, C., Bargiotas, P., Bassetti, C., & Faraci, F. D. (2019). Automated sleep scoring: A review of the latest approaches. *Sleep Medicine Reviews*, *48*, 101204. https://doi.org/10.1016/j.smrv.2019.07.007

Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M. E., Haselgrove, C., Van Der Kouwe, A., Killiany, R. J., Kennedy, D. N., Klaveness, S., Montillo, A., Makris, N., Rosen, B. R., & Dale, A. M. (2002). Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, *33*(3), 341–355. https://doi.org/10.1016/s0896-6273(02)00569-x

Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, *47*, S102. https://doi.org/10.1016/s1053-8119(09)70884-5

Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association*, *32*(200), 675–701. https://doi.org/10.1080/01621459.1937.10503522

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J., Frith, C., & Frackowiak, R. S. J. (1994). Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, *2*(4), 189–210. https://doi.org/10.1002/hbm.460020402

Gaiduk, M., Alarcón, Á. S., Seepold, R., & Madrid, N. M. (2023). Current status and prospects of automatic sleep stages scoring: Review. *Biomedical Engineering Letters*, *13*(3), 247–272. https://doi.org/10.1007/s13534-023-00299-3

Ge, J., Zhou, P., Zhao, X., & Wang, M. (2007). Sample entropy analysis of sleep EEG under different stages. *IEEE/ICME International Conference on Complex Medical Engineering*. https://doi.org/10.1109/iccme.2007.4381996

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C.

(2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178. https://doi.org/10.1038/nature18933

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., & Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124. https://doi.org/10.1016/j.neuroimage.2013.04.127

Glover, G. H. (2011). Overview of functional magnetic resonance imaging. *Neurosurgery Clinics of North America*, *22*(2), 133–139. https://doi.org/10.1016/j.nec.2010.11.001

Gneiting, T., & Schlather, M. (2004). Stochastic models that separate fractal dimension and the Hurst effect. *Siam Review*, *46*(2), 269–282. https://doi.org/10.1137/s0036144501394387

Goldberger, A. L., Amaral, L. a. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., Mietus, J. E., Moody, G. B., Peng, C., & Stanley, H. E. (2000). PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*, *101*(23). https://doi.org/10.1161/01.cir.101.23.e215

Gómez, C., Mediavilla, Á., Hornero, R., Abásolo, D., & Fernández, A. (2009). Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients. *Medical Engineering & Physics*, *31*(3), 306–313. https://doi.org/10.1016/j.medengphy.2008.06.010

Gomolka, R., Kampusch, S., Kaniušas, E., Thürk, F., Széles, J. C., & Klonowski, W. (2018). Higuchi fractal dimension of heart rate variability during percutaneous auricular vagus nerve stimulation in healthy and diabetic subjects. *Frontiers in Physiology*, *9*. https://doi.org/10.3389/fphys.2018.01162

Gonzalez, C. E., Mak-McCully, R. A., Rosen, B. Q., Cash, S. S., Chauvel, P., Bastuji, H., Rey, M., & Halgren, E. (2018). Theta bursts precede, and spindles follow, cortical and thalamic downstates in human NREM sleep. *The Journal of Neuroscience*, *38*(46), 9989–10001. https://doi.org/10.1523/jneurosci.0476-18.2018

González, C. M., Jensen, E. W., Gambús, P. L., & Vallverdú, M. (2019). Entropy measures as descriptors to identify apneas in rheoencephalographic signals. *Entropy*, *21*(6), 605. https://doi.org/10.3390/e21060605

Gordon, E. M., Laumann, T. O., Adeyemo, B., Huckins, J. F., Kelley, W. M., & Petersen, S. E. (2016). Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral Cortex*, *26*(1), 288–303. https://doi.org/10.1093/cercor/bhu239

Gorgolewski, K. J., Auer, T., Calhoun, V. D., Craddock, R. C., Das, S., Duff, E. P., Flandin, G., Ghosh, S., Glatard, T., Halchenko, Y. O., Handwerker, D. A., Hanke, M., Keator, D., Li, X., Michael, Z., Maumet, C., Nichols, B. N., Nichols, T. E., Pellman, J., . . . Poldrack, R. A. (2016). The brain imaging data structure, a format for organizing and describing outputs of neuroimaging experiments. *Scientific Data*, *3*(1), 160044. https://doi.org/10.1038/sdata.2016.44

Gorgolewski, K. J., Burns, C. D., Madison, C. M., Clark, D., Halchenko, Y. O., Waskom, M. L., & Ghosh, S. S. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*, 13. https://doi.org/10.3389/fninf.2011.00013

Gorgolewski, K. J., Esteban, O., Markiewicz, C. J., Ziegler, E., Ellis, D. G., Notter, M. P., Jarecka, D., Burns, C., Goncalves, M., Berleant, S., Pinsard, B., Madison, C., Waskom, M., Clark, D., Manhães-Savio, A., Clark, D., Jordan, K., Dayan, M., Halchenko, Y. O., ...

Ghosh, S. (2018). *Nipype* (Version 1.8.3) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.596855.

Goshvarpour, A., & Abbasi, A. (2013). Analysis of electroencephalogram signals in different sleep stages using detrended fluctuation analysis. *International Journal of Image, Graphics and Signal Processing*, *5*(12), 49–55. https://doi.org/10.5815/ijigsp.2013.12.07

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00267

Gratton, C., Dworetsky, A., Coalson, R. S., Adeyemo, B., Laumann, T. O., Wig, G. S., Kong, T. S., Gratton, G., Fabiani, M., Barch, D. M., Tranel, D., Miranda-Dominguez, O., Fair, D. A., Dosenbach, N. U., Snyder, A. Z., Perlmutter, J. S., Petersen, S. E., & Campbell, M. C. (2020). Removal of high frequency contamination from motion estimates in single-band fMRI saves data without biasing functional connectivity. *NeuroImage*, *217*, 116866. https://doi.org/10.1016/j.neuroimage.2020.116866

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060

Gu, Y., Han, F., Sainburg, L. E., Schade, M. M., Buxton, O. M., Duyn, J. H., & Liu, X. (2022). An orderly sequence of autonomic and neural events at transient arousal changes. *NeuroImage*, *264*, 119720. https://doi.org/10.1016/j.neuroimage.2022.119720

Gu, Y., Sainburg, L. E., Han, F., & Liu, X. (2023). Simultaneous EEG and functional MRI data during rest and sleep from humans. *Data in Brief*, *48*, 109059. https://doi.org/10.1016/j.dib.2023.109059

Gutiérrez-Tobal, G. C., Álvarez, D., Gómez-Pilar, J., Del Campo, F., & Hornero, R. (2015). Assessment of time and frequency domain entropies to detect sleep apnoea in heart rate variability recordings from men and women. *Entropy*, *17*(1), 123–141. https://doi.org/10.3390/e17010123

Hardstone, R., Poil, S., Schiavone, G., Jansen, R. C., Nikulin, V. V., Mansvelder, H. D., & Linkenkaer-Hansen, K. (2012). Detrended fluctuation analysis: A scale-free view on neuronal oscillations. *Frontiers in Physiology*, *3*. https://doi.org/10.3389/fphys.2012.00450

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E. S., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Van Kerkwijk, M. H., Brett, M., Haldane, A., Del Río, J. F., Wiebe, M. W., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hicks, S. A., Strümke, I., Thambawita, V., Hammou, M., Riegler, M. A., Halvorsen, P., & Parasa, S. (2022). On evaluation metrics for medical applications of artificial intelligence. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-09954-8

Higuchi, T. (1988). Approach to an irregular time series on the basis of the fractal theory. *Physica D: Nonlinear Phenomena*, *31*(2), 277–283. https://doi.org/10.1016/0167-2789(88)90081-4

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/mcse.2007.55

Hurst, H. E. (1951). Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, *116*(1), 770–799. https://doi.org/10.1061/taceat.0006518

Iannetti, G. D., Niazy, R. K., Wise, R. G., Jezzard, P., Brooks, J., Zambreanu, L., Vennart, W., Matthews, P. M., & Tracey, I. (2005). Simultaneous recording of laser-evoked brain potentials and continuous, high-field functional magnetic resonance imaging in humans. *NeuroImage*, *28*(3), 708–719. https://doi.org/10.1016/j.neuroimage.2005.06.060

Iber, C., Ancoli-Israel, S., L. Chesson, A., & F. Quan, S. (2007). *The AASM manual for the scoring of sleep and associated events: Rules, terminology, and technical specification* (1st ed.). American Academy of Sleep Medicine.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1006/nimg.2002.1132

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., & Smith, S. M. (2012). FSL. *NeuroImage*, *62*(2), 782–790. https://doi.org/10.1016/j.neuroimage.2011.09.015

Jiang, G. J. A., Fan, S., Abbod, M., Huang, H., Lan, J. Y., Tsai, F., Chang, H. C., Yang, Y. W., Chuang, F. L., Chiu, Y. F., Jen, K., Wu, J. F., & Shieh, J. (2015). Sample entropy analysis of EEG signals via artificial neural networks to model patients' consciousness level based on anesthesiologists experience. *BioMed Research International*, *2015*, 1–8. https://doi.org/10.1155/2015/343478

Jiang, L., & Zuo, X. (2016). Regional homogeneity: A multimodal, multiscale neuroimaging marker of the human connectome. *Neuroscientist*, *22*(5), 486–505. https://doi.org/10.1177/1073858415595004

Katz, M. J. (1988). Fractals and the analysis of waveforms. *Computers in Biology and Medicine*, *18*(3), 145–156. https://doi.org/10.1016/0010-4825(88)90041-8

Kemp, B., Zwinderman, A., Tuk, B., Kamphuisen, H., & Oberye, J. (2000). Analysis of a sleep-dependent neuronal feedback loop: The slow-wave microcontinuity of the EEG. *IEEE Transactions on Biomedical Engineering*, *47*(9), 1185–1194. https://doi.org/10.1109/10.867928

Kendall, M. G., & Smith, B. B. (1939). The problem of m rankings. *Annals of Mathematical Statistics*, *10*(3), 275–287. https://doi.org/10.1214/aoms/1177732186

Kerkhof, G. (2017). Epidemiology of sleep and sleep disorders in The Netherlands. *Sleep Medicine*, *30*, 229–239. https://doi.org/10.1016/j.sleep.2016.09.015

Kesić, S., & Spasić, S. (2016). Application of Higuchi's fractal dimension from basic to clinical neurophysiology: A review. *Computer Methods and Programs in Biomedicine*, *133*, 55–70. https://doi.org/10.1016/j.cmpb.2016.05.014

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., Lee, N., Rossa, B., Reuter, M., Neto, E. C., & Keshavan, A. (2017). Mindboggling morphometry of human brains. *PLoS Computational Biology*, *13*(2), e1005350. https://doi.org/10.1371/journal.pcbi.1005350

Kothe, C., & Jung, T.-P. (2016). *Artifact removal techniques with signal reconstruction* (Patent No. US20160113587A1). U.S. Patent and Trademark Office.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., Cheng, H., Brady, T. J., & Rosen, B. R. (1992). Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(12), 5675–5679. https://doi.org/10.1073/pnas.89.12.5675

Lambert, I., & Peter-Derex, L. (2023). Spotlight on sleep stage classification based on EEG. *Nature and Science of Sleep*, *15*, 479–490. https://doi.org/10.2147/nss.s401270

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, *1*(1), 76–85. https://doi.org/10.1137/0701007

Lau, Z. J., Pham, T. N., Chen, S. H. A., & Makowski, D. (2022). Brain entropy, fractal dimensions and predictability: A review of complexity measures for EEG in healthy and neuropsychiatric populations. *European Journal of Neuroscience*, *56*(7), 5047–5069. https://doi.org/10.1111/ejn.15800

Lee, J., Kim, D., Kim, I., & Kim, S. I. (2001). Analysis of scaling exponents of waken and sleeping stage in EEG. In *Connectionist models of neurons, learning processes, and artificial intelligence* (pp. 450–456). https://doi.org/10.1007/3-540-45720-8_53

Lemaitre, G., Nogueira, F., & Aridas, C. K. (2016). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced datasets in machine learning. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1609.06570

Levene, H. (1960). Robust tests for equality of variances. In I. Olkin & H. Hotelling (Eds.), *Contributions to probability and statistics: Essays in honor of Harold Hotelling* (pp. 278–292). Stanford University Press. https://ci.nii.ac.jp/naid/10007628681

Li, C., Qi, Y., Ding, X., Zhao, J., Sang, T., & Lee, M. (2022). A deep learning method approach for sleep stage classification with EEG spectrogram. *International Journal of Environmental Research and Public Health*, *19*(10), 6322. https://doi.org/10.3390/ijerph19106322

Li, H., Cheng, P., & Ye, D. (2015). A study of sleep staging based on a sample entropy analysis of electroencephalogram. *Bio-medical Materials and Engineering*, *26*(s1), S1149–S1156. https://doi.org/10.3233/bme-151411

Liu, Z., De Zwart, J. A., Van Gelderen, P., Kuo, L. W., & Duyn, J. H. (2012). Statistical feature extraction for artifact removal from concurrent fMRI-EEG recordings. *NeuroImage*, *59*(3), 2073–2087. https://doi.org/10.1016/j.neuroimage.2011.10.042

Lo, Y., Hsiao, Y., & Chang, F. (2022). Use electroencephalogram entropy as an indicator to detect stress-induced sleep alteration. *Applied Sciences*, *12*(10), 4812. https://doi.org/10.3390/app12104812

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, *453*(7197), 869–878. https://doi.org/10.1038/nature06976

Lomb, N. R. (1976). Least-squares frequency analysis of unequally spaced data. *Astrophysics and Space Science*, *39*(2), 447–462. https://doi.org/10.1007/bf00648343

Loomis, A. L., Harvey, E. N., & Hobart, G. (1935). Potential rhythms of the cerebral cortex during sleep. *Science*, *81*(2111), 597–598. https://doi.org/10.1126/science.81.2111.597

Loomis, A. L., Harvey, E. N., & Hobart, G. (1937). Cerebral states during sleep, as studied by human brain potentials. *Journal of Experimental Psychology*, *21*(2), 127–144. https://doi.org/10.1037/h0057431

Louppe, G. (2014). Understanding random forests: From theory to practice. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1407.7502

Ma, Y., Shi, W., Peng, C., & Yang, A. C. (2018). Nonlinear dynamical analysis of sleep electroencephalography using fractal and entropy approaches. *Sleep Medicine Reviews*, *37*, 85–93. https://doi.org/10.1016/j.smrv.2017.01.003

Makowski, D., Pham, T. N., Lau, Z. J., Brammer, J. C., Lespinasse, F., Phạm, H., Schölzel, C., & Chen, S. H. A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior Research Methods*, *53*(4), 1689–1696. https://doi.org/10.3758/s13428-020-01516-y

Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *Siam Review*, *10*(4), 422–437. https://doi.org/10.1137/1010093

Mandelbrot, B. B., & Wallis, J. R. (1969). Some long-run properties of geophysical records. *Water Resources Research*, *5*(2), 321–340. https://doi.org/10.1029/wr005i002p00321

Mandelbrot, B. B., & Wheeler, J. A. (1983). The fractal geometry of nature. *American Journal of Physics*, *51*(3), 286–287. https://doi.org/10.1119/1.13295

Markett, S., Reuter, M., Montag, C., Voigt, G., Lachmann, B., Rudorf, S., Elger, C. E., & Weber, B. (2013). Assessing the function of the fronto-parietal attention network: Insights from resting-state fMRI and the attentional network test. *Human Brain Mapping*, *35*(4), 1700–1709. https://doi.org/10.1002/hbm.22285

Mauchly, J. W. (1940). Significance test for sphericity of a normal n-variate distribution. *Annals of Mathematical Statistics*, *11*(2), 204–209. https://doi.org/10.1214/aoms/1177731915

Mayeli, A., Zoubi, O. A., Henry, K., Wong, C. K., White, E. J., Luo, Q., Zotev, V., Refai, H. H., & Bodurka, J. (2021). Automated pipeline for EEG artifact reduction (APPEAR) recorded during fMRI. *Journal of Neural Engineering*, *18*(4), 0460b4. https://doi.org/10.1088/1741-2552/ac1037

Mayer, C. C., Bachler, M., Hörtenhuber, M., Stocker, C., Holzinger, A., & Wassertheurer, S. (2014). Selection of entropy-measure parameters for knowledge discovery in heart rate

variability data. *BMC Bioinformatics*, *15*(S6). https://doi.org/10.1186/1471-2105-15-s6-s2

Mehta, K., Salo, T., Madison, T., Adebimpe, A., S. Bassett, D., Bertolero, M., Cieslak, M., Covitz, S., Houghton, A., S. Keller, A., Luo, A., Miranda-Dominguez, O., M. Nelson, S., Shafiei, G., Shanmugan, S., T. Shinohara, R., J. Sydnor, V., Feczko, E., A. Fair, D., & D. Satterthwaite, T. (2023). XCP-D: A robust pipeline for the post-processing of fMRI data. *bioRxiv*. https://doi.org/10.1101/2023.11.20.567926

Mishra, P., Pandey, C. K., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive statistics and normality tests for statistical data. *Annals of Cardiac Anaesthesia*, *22*(1), 67. https://doi.org/10.4103/aca.aca_157_18

Miskovic, V., MacDonald, K. J., Rhodes, L. J., & Côté, K. A. (2018). Changes in EEG multiscale entropy and power-law frequency scaling during the human sleep cycle. *Human Brain Mapping*, *40*(2), 538–551. https://doi.org/10.1002/hbm.24393

Moser, D., Anderer, P., Gruber, G., Parapatics, S., Loretz, E., Boeck, M., Kloesch, G., Heller, E., Schmidt, A., Danker-Hopfe, H., Saletu, B., Zeitlhofer, J., & Dorffner, G. (2009). Sleep classification according to AASM and Rechtschaffen & Kales: Effects on sleep scoring parameters. *Sleep*, *32*(2), 139–149. https://doi.org/10.1093/sleep/32.2.139

Murphy, M., Riedner, B. A., Huber, R., Massimini, M., Ferrarelli, F., & Tononi, G. (2009). Source modeling sleep slow waves. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(5), 1608–1613. https://doi.org/10.1073/pnas.0807933106

Namdari, A., & Li, Z. (2019). A review of entropy measures for uncertainty quantification of stochastic processes. *Advances in Mechanical Engineering*, *11*(6), 1–14. https://doi.org/10.1177/1687814019857350

Nazarychev, S. A., Zagretdinov, A. R., Ziganshin, S. G., & Vankov, Y. (2019). Classification of time series using the Hurst exponent. *Journal of Physics. Conference Series*, *1328*(1), 012056. https://doi.org/10.1088/1742-6596/1328/1/012056

Nemenyi, P. B. (1963). *Distribution-free multiple comparisons* [Ph.D. dissertation, Princeton University]. Princeton University Library. https://catalog.princeton.edu/catalog/9920813653506421

Niazy, R. K., Beckmann, C. F., Iannetti, G. D., Brady, J. M., & Smith, S. M. (2005). Removal of FMRI environment artifacts from EEG data using optimal basis sets. *NeuroImage*, *28*(3), 720–737. https://doi.org/10.1016/j.neuroimage.2005.06.067

Ogawa, S., Lee, T. M., Kay, A. R., & Tank, D. W. (1990). Brain magnetic resonance imaging with contrast dependent on blood oxygenation. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(24), 9868–9872. https://doi.org/10.1073/pnas.87.24.9868

Patriat, R., Reynolds, R. C., & Birn, R. M. (2017). An improved model of motion-related signal changes in fMRI. *NeuroImage*, *144*, 74–82. https://doi.org/10.1016/j.neuroimage.2016.08.051

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2012). Scikit-learn: Machine learning in Python. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.1201.0490

Peng, C., Buldyrev, S. V., Havlin, S., Simons, M., Stanley, H. E., & Goldberger, A. L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E*, *49*(2), 1685–1689. https://doi.org/10.1103/physreve.49.1685

Peng, C., Havlin, S., Stanley, H. E., & Goldberger, A. L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *5*(1), 82–87. https://doi.org/10.1063/1.166141

Phan, H., & Mikkelsen, K. B. (2022). Automatic sleep staging of EEG signals: Recent development, challenges, and future directions. *Physiological Measurement*, *43*(4). https://doi.org/10.1088/1361-6579/ac6049

Pincus, S. M. (1991). Approximate entropy as a measure of system complexity. *Proceedings of the National Academy of Sciences of the United States of America*, *88*(6), 2297–2301. https://doi.org/10.1073/pnas.88.6.2297

Pincus, S. M. (1995). Approximate entropy (ApEn) as a complexity measure. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *5*(1), 110–117. https://doi.org/10.1063/1.166092

Pincus, S. M., & Goldberger, A. L. (1994). Physiological time-series analysis: What does regularity quantify? *American Journal of Physiology-heart and Circulatory Physiology*, *266*(4), H1643–H1656. https://doi.org/10.1152/ajpheart.1994.266.4.h1643

Pincus, S. M., & Huang, W. (1992). Approximate entropy: Statistical properties and applications. *Communications in Statistics - Theory and Methods*, *21*(11), 3061–3077. https://doi.org/10.1080/03610929208830963

Porcaro, C., Moaveninejad, S., D'Onofrio, V., & DiIeva, A. (2024). Fractal time series: Background, estimation methods, and performances. In *Advances in neurobiology* (pp. 95–137). https://doi.org/10.1007/978-3-031-47606-8_5

Power, J. D., Mitra, A., Laumann, T. O., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048

Ramezan, C. A., Warner, T. A., Maxwell, A. E., & Price, B. S. (2021). Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data. *Remote Sensing*, *13*(3), 368. https://doi.org/10.3390/rs13030368

Rechtschaffen, A. (1998). Current perspectives on the function of sleep. *Perspectives in Biology and Medicine*, *41*(3), 359–390. https://doi.org/10.1353/pbm.1998.0051

Rechtschaffen, A., & Kales, A. (1968). *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects*. https://ci.nii.ac.jp/naid/10023901533

Richman, J., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, *278*(6), H2039–H2049. https://doi.org/10.1152/ajpheart.2000.278.6.h2039

Rodríguez-Sotelo, J. L., Osorio-Forero, A., Jiménez-Rodríguez, A., Cuesta–Frau, D., Cirugeda-Roldán, E., & Peluffo, D. (2014). Automatic sleep stages classification using EEG entropy features and unsupervised pattern analysis techniques. *Entropy*, *16*(12), 6573–6589. https://doi.org/10.3390/e16126573

Rojas, G. M., Alvarez, C., Montoya, C. E., De La Iglesia-Vayá, M., Cisternas, J. E., & Gálvez, M. (2018). Study of resting-state functional connectivity networks using EEG electrodes position as seed. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00235

Rosenberg, R. S., & Van Hout, S. (2013). The American Academy of Sleep Medicine inter-scorer reliability program: Sleep stage scoring. *Journal of Clinical Sleep Medicine*, *09*(01), 81–87. https://doi.org/10.5664/jcsm.2350

Sateia, M. J. (2014). International classification of sleep disorders-third edition. *Chest*, *146*(5), 1387–1394. https://doi.org/10.1378/chest.14-0970

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., Eickhoff, S. B., Hakonarson, H., Gur, R. C., Gur, R. E., & Wolf, D. H. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, *64*, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052

Scargle, J. D. (1982). Studies in astronomical time series analysis. II - statistical aspects of spectral analysis of unevenly spaced data. *The Astrophysical Journal*, *263*, 835-853. https://doi.org/10.1086/160554

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, *28*(9), 3095–3114. https://doi.org/10.1093/cercor/bhx179

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia*, *126*(5), 1763–1768. https://doi.org/10.1213/ane.0000000000002864

Scornet, E. (2017). Tuning parameters in random forests. *ESAIM: Proceedings and Surveys*, *60*, 144–162. https://doi.org/10.1051/proc/201760144

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, *52*(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Siegel, J. M. (2005). Clues to the functions of mammalian sleep. *Nature*, *437*(7063), 1264–1271. https://doi.org/10.1038/nature04285

Siegel, J. M. (2008). Do all animals sleep? *Trends in Neurosciences*, *31*(4), 208–213. https://doi.org/10.1016/j.tins.2008.02.001

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P., De Luca, M., Drobnjak, I., Flitney, D., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*, S208–S219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Song, C., Boly, M., Tagliazucchi, E., Laufs, H., & Tononi, G. (2022). fMRI spectral signatures of sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(30). https://doi.org/10.1073/pnas.2016732119

Sri, T. R., A, R. R., Madala, J., Duddukuru, S. L., Reddipalli, R., & Polasi, P. K. (2022). A systematic review on deep learning models for sleep stage classification. *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*. https://doi.org/10.1109/icoei53556.2022.9776965

Stiller, J. W., & Postolache, T. T. (2005). Sleep-wake and other biological rhythms: Functional neuroanatomy. *Clinics in Sports Medicine*, *24*(2), 205–235. https://doi.org/10.1016/j.csm.2004.12.008

Sturzbecher, M. J., & De Araújo, D. B. (2012). Simultaneous EEG-fMRI: Integrating spatial and temporal resolution. In *The relevance of the time domain to neural network models* (pp. 199–217). https://doi.org/10.1007/978-1-4614-0724-9_11

Tagliazucchi, E., Behrens, M., & Laufs, H. (2013). Sleep neuroimaging and models of consciousness. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00256

Tagliazucchi, E., & Laufs, H. (2014). Decoding wakefulness levels from typical fMRI resting-state data reveals reliable drifts between wakefulness and sleep. *Neuron*, *82*(3), 695–708. https://doi.org/10.1016/j.neuron.2014.03.020

Tagliazucchi, E., Von Wegner, F., Morzelewski, A., Borisov, S. V., Jahnke, K., & Laufs, H. (2012). Automatic sleep staging using fMRI functional connectivity data. *NeuroImage*, *63*(1), 63–72. https://doi.org/10.1016/j.neuroimage.2012.06.036

Tagliazucchi, E., Von Wegner, F., Morzelewski, A., Brodbeck, V., Jahnke, K., & Laufs, H. (2013). Breakdown of long-range temporal dependence in default mode and attention networks during deep sleep. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15419–15424. https://doi.org/10.1073/pnas.1312848110

Taylor, A., Kim, J., & Ress, D. (2018). Characterization of the hemodynamic response function across the majority of human cerebral cortex. *NeuroImage*, *173*, 322–331. https://doi.org/10.1016/j.neuroimage.2018.02.061

Taylor, P. A., Chen, G., Glen, D. R., Reynolds, R. C., & Cox, R. W. (2018). Lomb-Scargle your way to RSFC parameter estimation in AFNI-FATCAT. *International Society for Magnetic Resonance in Medicine*. https://archive.ismrm.org/2018/2830.html

Taylor, P. A., & Saad, Z. S. (2013). FATCAT: (An efficient) functional and tractographic connectivity analysis toolbox. *Brain Connectivity*, *3*(5), 523–535. https://doi.org/10.1089/brain.2013.0154

Tian, Y., Margulies, D. S., Breakspear, M., & Zalesky, A. (2020). Topographic organization of the human subcortex unveiled with functional connectivity gradients. *Nature Neuroscience*, *23*(11), 1421–1432. https://doi.org/10.1038/s41593-020-00711-6

Tononi, G., Boly, M., Gosseries, O., & Laureys, S. (2016). The neurology of consciousness: An overview. In *The neurology of consciousness: Cognitive neuroscience and neuropathology* (2nd ed., pp. 407–461). https://doi.org/10.1016/b978-0-12-800948-2.00025-x

Townsend, R. H. D. (2010). Fast calculation of the Lomb-Scargle periodogram using graphics processing units. *The Astrophysical Journal Supplement Series*, *191*(2), 247–253. https://doi.org/10.1088/0067-0049/191/2/247

Tukey, J. W. (1949). Comparing individual means in the analysis of variance. *Biometrics*, *5*(2), 99. https://doi.org/10.2307/3001913

Tukey, J. W. (1977). Schematic summaries. In *Exploratory data analysis* (pp. 43–44). Addison-Wesley Publishing Company.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., & Gee, J. C. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*(6), 1310–1320. https://doi.org/10.1109/tmi.2010.2046908

Utevsky, A. V., Smith, D. V., & Huettel, S. A. (2014). Precuneus is a functional core of the default-mode network. *The Journal of Neuroscience*, *34*(3), 932–940. https://doi.org/10.1523/jneurosci.4227-13.2014

Varley, T. F., Craig, M., Adapa, R., Finoia, P., Williams, G. B., Allanson, J., Pickard, J. D., Menon, D. K., & Stamatakis, E. A. (2020). Fractal dimension of cortical functional connectivity networks & severity of disorders of consciousness. *PloS One*, *15*(2), e0223812. https://doi.org/10.1371/journal.pone.0223812

Vega, C. F., & Noel, J. (2015). Parameters analyzed of Higuchi's fractal dimension for EEG brain signals. *2015 Signal Processing Symposium (SPSympo)*. https://doi.org/10.1109/sps.2015.7168285

Vijayan, S., Lepage, K. Q., Kopell, N., & Cash, S. S. (2017). Frontal beta-theta network during REM sleep. *eLife*, *6*, e18894. https://doi.org/10.7554/elife.18894

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Van Der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wainer, J., & Fonseca, P. (2020). How to tune the RBF SVM hyperparameters?: An empirical evaluation of 18 search algorithms. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2008.11655

Wang, S., Tepfer, L. J., Taren, A. A., & Smith, D. V. (2017). Functional parcellation of the default mode network: A large-scale meta-analysis. *bioRxiv (Cold Spring Harbor Laboratory)*. https://doi.org/10.1101/225375

Wang, Z., Yin, L., Childress, A. R., & Detre, J. A. (2014). Brain entropy mapping using fMRI. *PLOS ONE*, *9*(3), e89948. https://doi.org/10.1371/journal.pone.0089948

Warach, S., Ives, J. R., Schlaug, G., Patel, M. D., Darby, D., Thangaraj, V., Rr, E., & Schomer, D. L. (1996). EEG-triggered echo-planar functional MRI in epilepsy. *Neurology*, *47*(1), 89–93. https://doi.org/10.1212/wnl.47.1.89

Warbrick, T. (2022). Simultaneous EEG-fMRI: What have we learned and what does the future hold? *Sensors*, *22*(6), 2262. https://doi.org/10.3390/s22062262

Weaver, K. F., Morales, V., Dunn, S. L., Godde, K., & Weaver, P. F. (2017). Anova. In *An introduction to statistical analysis in research: With applications in the biological and life sciences* (pp. 227–296). John Wiley & Sons. https://doi.org/10.1002/9781119454205.ch6

Weiss, B., Clemens, Z., Bódizs, R., Vágó, Z., & Halász, P. (2009). Spatio-temporal analysis of monofractal and multifractal properties of the human sleep EEG. *Journal of Neuroscience Methods*, *185*(1), 116–124. https://doi.org/10.1016/j.jneumeth.2009.07.027

Xie, H., He, W., & Liu, H. (2008). Measuring time series regularity using nonlinear similarity-based sample entropy. *Physics Letters A*, *372*(48), 7140–7146. https://doi.org/10.1016/j.physleta.2008.10.049

Yang, A. C., Tsai, S., Lin, C., & Peng, C. (2018). A strategy to reduce bias of entropy estimates in resting-state fMRI signals. *Frontiers in Neuroscience*, *12*. https://doi.org/10.3389/fnins.2018.00398

Yarkoni, T., Markiewicz, C., De La Vega, A., Gorgolewski, K., Salo, T., Halchenko, Y., McNamara, Q., DeStasio, K., Poline, J., Petrov, D., Hayot-Sasson, V., Nielson, D., Carlin, J., Kiar, G., Whitaker, K., DuPre, E., Wagner, A., Tirrell, L., Jas, M., . . . Blair, R. (2019). PyBIDS: Python tools for BIDS datasets. *Journal of Open Source Software*, *4*(40), 1294. https://doi.org/10.21105/joss.01294

Yentes, J. M., Hunt, N. H., Schmid, K. K., Kaipust, J. P., McGrath, D., & Stergiou, N. (2012). The appropriate use of approximate entropy and sample entropy with short data sets. *Annals of Biomedical Engineering*, *41*(2), 349–365. https://doi.org/10.1007/s10439-012-0668-3

Zhang, Y., Brady, M., & Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, *20*(1), 45–57. https://doi.org/10.1109/42.906424

Zhou, F., Huang, S., Gao, L., Zhuang, Y., Ding, S., & Gong, H. (2016). Temporal regularity of intrinsic cerebral activity in patients with chronic primary insomnia: A brain entropy study using resting-state fMRI. *Brain and Behavior*, *6*(10). https://doi.org/10.1002/brb3.529

Zou, G., Liu, J., Zou, Q., & Gao, J. (2022). A-PASS: An automated pipeline to analyze simultaneously acquired EEG-fMRI data for studying brain activities during sleep. *Journal of Neural Engineering*, *19*(4), 046031. https://doi.org/10.1088/1741-2552/ac83f2

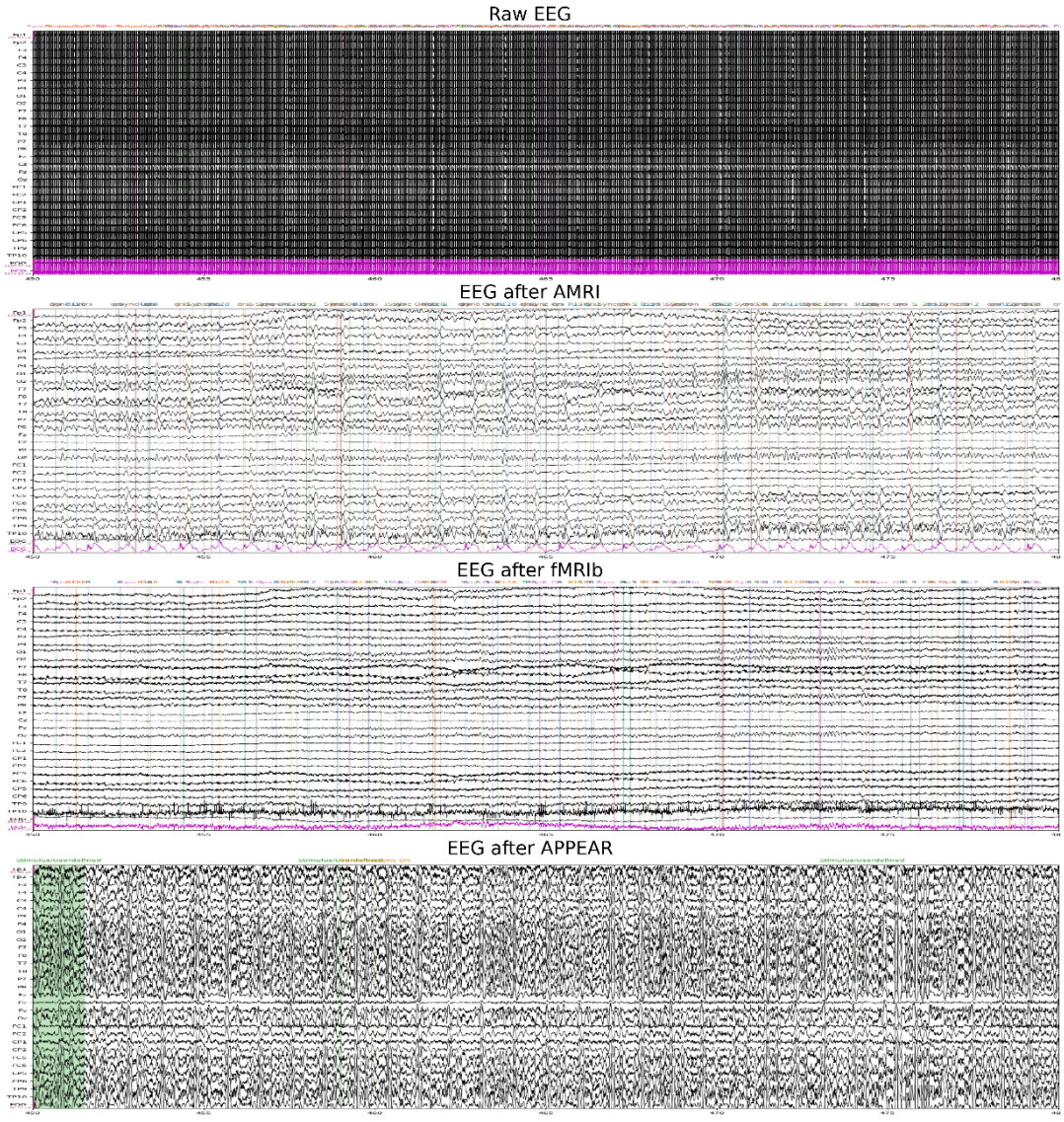
Zou, Q., Zhu, C., Yang, Y., Zuo, X., Long, X., Cao, Q., Wang, Y., & Zang, Y. (2008). An improved approach to detection of amplitude of low-frequency fluctuation (ALFF) for resting-state fMRI: Fractional ALFF. *Journal of Neuroscience Methods*, *172*(1), 137–141. https://doi.org/10.1016/j.jneumeth.2008.04.012

Appendix A

Comparison of the raw EEG Signal and the Output of the Tested Algorithms and Pipelines for FMRI Gradient and BCG Artifact Removal

Figure A1

Comparison of the Raw EEG Signal and the Output of the Tested Algorithms and Pipelines for FMRI Gradient and BCG Artifact Removal.



Note. The graph shows the raw EEG and EEG cleaned with various algorithms (AMRI, fMRIb, APPEAR) from all channels for the same 30-second long segments.

Appendix B

Report of fMRIprep on Preprocessing Procedures

Results included in this manuscript come from preprocessing performed using fMRIPrep 23.1.4 (Esteban et al. (2023); Esteban et al. (2018); RRID:SCR_016216), which is based on Nipype 1.8.6 (K. Gorgolewski et al. (2011); K. J. Gorgolewski et al. (2018); RRID:SCR_002502).

Anatomical data preprocessing

A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset. The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4BiasFieldCorrection (Tustison et al. 2010), distributed with ANTs (version unknown) (Avants et al. 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a Nipype implementation of the antsBrainExtraction.sh workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (FSL (version unknown), RRID:SCR_002823, Zhang, Brady, and Smith 2001). Brain surfaces were reconstructed using recon-all (FreeSurfer 7.3.2, RRID:SCR_001847, Dale, Fischl, and Sereno 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (RRID:SCR_002438, Klein et al. 2017). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with antsRegistration (ANTs (version unknown)), using brain-extracted versions of both T1w reference and the T1w template. The following template was were selected for spatial

normalization and accessed with TemplateFlow (23.0.0, Ciric et al. 2022): ICBM 152 Nonlinear Asymmetrical template version 2009c [Fonov et al. (2009), RRID:SCR_008796; TemplateFlow ID: MNI152NLin2009cAsym].

Functional data preprocessing

For each of the BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using mcflirt (FSL , Jenkinson et al. 2002). BOLD runs were slice-time corrected to 1.01s (0.5 of slice acquisition range 0s-2.02s) using 3dTshift from AFNI (Cox and Hyde 1997, RRID:SCR_005927). The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying the transforms to correct for head-motion. These resampled BOLD time-series will be referred to as preprocessed BOLD in original space, or just preprocessed BOLD. The BOLD reference was then co-registered to the T1w reference using bbregister (FreeSurfer) which implements boundary-based registration (Greve and Fischl 2009). Co-registration was configured with six degrees of freedom. Several confounding time-series were calculated based on the preprocessed BOLD: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in Nipype (following the definitions by Power et al. 2014). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a

set of physiological regressors were extracted to allow for component-based noise correction (CompCor, Behzadi et al. 2007). Principal components are estimated after high-pass filtering the preprocessed BOLD time-series (using a discrete cosine filter with 128s cut-off) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels on BOLD space, a mask of pixels that likely contain a volume fraction of GM is subtracted from the aCompCor masks. This mask is obtained by dilating a GM mask extracted from the FreeSurfer's aseg segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the k components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each (Satterthwaite et al. 2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardized DVARS were annotated as motion outliers. Additional nuisance timeseries are calculated by means of principal components analysis of the signal found within a thin band (crown) of voxels around the edge of the brain, as proposed by (Patriat, Reynolds, and Birn 2017). The BOLD

time-series were resampled into standard space, generating a preprocessed BOLD run in MNI152NLin2009cAsym space. First, a reference volume and its skull-stripped version were generated using a custom methodology of fMRIPrep. All resamplings can be performed with a single interpolation step by composing all the pertinent transformations (i.e. head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using antsApplyTransforms (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos 1964). Non-gridded (surface) resamplings were performed using mri_vol2surf (FreeSurfer).

Many internal operations of fMRIPrep use Nilearn 0.10.1 (Abraham et al. 2014, RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in fMRIPrep's documentation.

Copyright Waiver

Appendix C

Post-Processing of fmriprep Outputs

The eXtensible Connectivity Pipeline- DCAN (XCP-D) (Ciric et al. 2018; Satterthwaite et al. 2013) was used to post-process the outputs of fMRIPrep version 23.1.4 (Esteban et al. 2018, 2023, RRID:SCR_016216). XCP-D was built with Nipype version 1.8.6 (Gorgolewski et al. 2011, RRID:SCR_002502). Native-space T1w images were transformed to MNI152NLin2009cAsym space at 1 mm3 resolution. For each of the five BOLD runs found per subject (across all tasks and sessions), the following post-processing was performed. Non-steady-state volumes were extracted from the preprocessed confounds and were discarded from both the BOLD data and nuisance regressors. In order to identify high-motion outlier volumes, the six translation and rotation head motion traces were low-pass filtered below 6.0 breaths-per-minute using a(n) fourth-order Butterworth filter, based on Gratton et al. (2020). Next, framewise displacement was calculated using the formula from Power et al. (2014), with a head radius of 50.0 mm. Volumes with filtered framewise displacement greater than 0.3 mm were flagged as high-motion outliers for the sake of later censoring (Power et al. 2014). In total, 36 nuisance regressors were selected from the preprocessing confounds, according to the '36P' strategy. These nuisance regressors included six filtered motion parameters, mean global signal, mean white matter signal, mean cerebrospinal fluid signal with their temporal derivatives, and quadratic expansion of six motion parameters, tissue signals and their temporal derivatives (Ciric et al. 2017; Satterthwaite et al. 2013). The BOLD data were despiked with AFNI's 3dDespike. Nuisance regressors were regressed from the BOLD data using linear regression, as implemented in Nilearn. Any volumes censored earlier in the workflow were then interpolated in the residual time series produced by the regression. The interpolated timeseries were then band-pass filtered

using a(n) second-order Butterworth filter, in order to retain signals between 0.01-0.08 Hz. The filtered, interpolated time series were then re-censored to remove high-motion outlier volumes. The denoised BOLD was smoothed using Nilearn with a Gaussian kernel (FWHM=6.0 mm). The amplitude of low-frequency fluctuation (ALFF) (Zou et al. 2008) was computed by transforming the mean-centered, standard deviation-normalized, denoised BOLD time series to the frequency domain using the Lomb-Scargle periodogram (Lomb 1976; Scargle 1982; Townsend 2010; Taylor et al. 2018). The power spectrum was computed within the 0.01-0.08 Hz frequency band and the mean square root of the power spectrum was calculated at each voxel to yield voxel-wise ALFF measures. The resulting ALFF values were then multiplied by the standard deviation of the denoised BOLD time series to retain the original scaling. The ALFF maps were smoothed with Nilearn using a Gaussian kernel (FWHM=6.0 mm). Regional homogeneity (ReHo) (Jiang and Zuo 2016) was computed with neighborhood voxels using AFNI's 3dReHo (Taylor and Saad 2013).

Processed functional timeseries were extracted from the residual BOLD signal with Nilearn's NiftiLabelsMasker for the following atlases: the Schaefer Supplemented with Subcortical Structures (4S) atlas (Schaefer et al. 2018,@pauli2018high,@king2019functional,@najdenovska2018vivo,@glasser2013minimal) at 10 different resolutions (156, 256, 356, 456, 556, 656, 756, 856, 956, and 1056 parcels), the Glasser atlas (Glasser et al. 2016), the Gordon atlas (Gordon et al. 2016), the Tian subcortical atlas (Tian et al. 2020), and the HCP CIFTI subcortical atlas (Glasser et al. 2013). Corresponding pair-wise functional connectivity between all regions was computed for each atlas, which was operationalized as the Pearson's correlation of each parcel's unsmoothed timeseries. In cases of

partial coverage, uncovered voxels (values of all zeros or NaNs) were either ignored (when the parcel had >50.0% coverage) or were set to zero (when the parcel had <50.0% coverage).

Many internal operations of XCP-D use AFNI (Cox 1996; Cox and Hyde 1997), ANTS (Avants et al. 2009), TemplateFlow version 24.2.0 (Ciric et al. 2022), matplotlib version 3.8.2 (Hunter 2007), Nibabel version 5.2.0 (Brett et al. 2022), Nilearn version 0.10.3 (Abraham et al. 2014), numpy version 1.26.2 (Harris et al. 2020), pybids version 0.16.4 (Yarkoni et al. 2019), and scipy version 1.11.4 (Virtanen et al. 2020). For more details, see the XCP-D website (https://xcp-d.readthedocs.io).

Copyright Waiver

Appendix D

Approximate Entropy (ApEn) Calculation

ApEn can be measured in a time-series, that has N number of data points, by breaking it down into smaller segments that have fixed lengths, m, which are shorter than N. Then, one can compare the resulting segments and estimate the chance that sequences of data points which are similar over a certain number of observations (m) will continue to be similar when another data point is added, within a certain pre-specified limit r (Delgado-Bonal & Marshak, 2019; Pincus, 1991). This is reflected in Equation 2:

$$ApEn(m,r,N) = \Phi^{m}(r) - \Phi^{m+1}(r)\,, \tag{2}$$

where m and m+1 are the lengths of these smaller segments that are being compared, and r corresponds to the previously mentioned limit or, in other words, tolerance (Delgado-Bonal & Marshak, 2019; González et al., 2019; Pincus, 1991). Furthermore, $\Phi^{m}(r)$ encapsulates these, representing the average of the logarithmic chances of finding similar segments of length m within the tolerance r, as shown in Equation 3:

$$\Phi^{m}(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} log\, C_i^{m}(r)\,, \tag{3}$$

where $C_i^{m}(r)$ is the correlation integral estimation (Delgado-Bonal & Marshak, 2019; González et al., 2019; Pincus, 1991).

Appendix E

Sample Entropy (SampEn) Calculation

Formally, SampEn can be defined in the following way. Considering a time-series data with N datapoints, one can begin calculating SampEn in a similar manner to ApEn, except that this time self-comparisons are omitted from the correlation integral estimation, after which one takes the natural logarithmic values of $\Phi^{m+1}(r)$ and subtracts them from the natural logarithms of $\Phi^{m}(r)$ (González et al., 2019). As outlined by González and colleagues (2019), the resulting equations are Equation 4, Equation 5, and Equation 6:

$$C_i^m(r) = \frac{\text{instances of } j \text{ with distance } [x(i), x(j)] \leq r \text{ and } i \neq j}{N - \text{m} + 1}\ , \tag{4}$$

$$\Phi^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} \log C_i^m(r)\ , \tag{5}$$

$$SampEn(m,r,N) = \log\Phi^m(r) - \log\Phi^{m+1}(\text{r})\ , \tag{6}$$

where m specifies the length of the compared sequences, r denotes the tolerance for accepting matching segments, and $\log\Phi^m(r)$ and $\log\Phi^{m+1}(r)$ correspond to the logarithmic chance of the segments matching for m and m+1 data points, respectively. As can be seen, $i \neq j$ indicates that identical pairs are not included in the process.

Appendix F

Higuchi's Fractal Dimension (HFD) Calculation

The calculation of HFD for a time series that has length N involves multiple steps that have been set forth by Higuchi (1988). First, a number of k new time series are made, given the original time series with length N, as illustrated in Equation 7:

$$X_k^m: x(m), x(m+k), x(m+2k), \ldots, x(m+[\frac{N-m}{k}]\cdot k), \tag{7}$$

where *m* denotes the starting time points, while *k* indicates the interval between points (Gómez et al., 2009; Higuchi, 1988). After that, the curve length of X_k^m is determined for each of these time series through the formula in Equation 8:

$$L_m(k)=\frac{\left\{\left(\sum_{i=1}^{\left[\frac{N-m}{k}\right]}|x(m+ik)-x(m+(i-1)\cdot k|\right)\frac{N-1}{\left[\frac{N-m}{k}\right]\cdot k}\right\}}{k}, \tag{8}$$

in which case the $\frac{N-1}{\left[\frac{N-m}{k}\right]\cdot k}$ is the factor used for normalizing the length of the curve of the resulting time series based on the number of intervals and the series length (Gómez et al., 2009; Higuchi, 1988). This process is then repeated for various values of k, averaging the segment lengths for each k to find the curve length associated with that interval, which can be seen in Equation 9:

$$L(k)=\frac{1}{k}\times\sum_{m=1}^{k}L_m(k)\ . \tag{9}$$

As a result, HFD corresponds to the slope of least squares best fitting line over all sets of average curve lengths ($L(k)$) and inverses of k ($1/k$), on a double logarithmic scale, where *k* varies up to k_{max} (Gómez et al., 2009; Higuchi, 1988).

Table F1

Selected and Mean k_{max} Values for Each EEG Channel

EEG channel	Selected k_{max} for Complete Time Series (and Mean)	Selected k_{max} for Sleep Stage-Specific Segment (and Mean)
C3	16 (15.56)	19 (18.28)
C4	17 (16.25)	19 (18.56)
CP1	18 (17.39)	20 (19.36)
CP2	20 (19.40)	21 (20.31)
CP5	17 (16.40)	19 (18.60)
CP6	15 (14.83)	18 (17.01)
Cz	21 (20.28)	22 (21.33)
F3	15 (14.59)	16 (15.59)
F4	17 (16.31)	18 (17.69)
F7	15 (14.10)	16 (15.47)
F8	18 (17.27)	20 (19.07)
FC1	19 (18.31)	20 (19.97)
FC2	19 (18.61)	20 (19.73)
FC5	15 (14.65)	17 (16.83)
FC6	16 (15.98)	19 (18.02)
Fp1	18 (17.02)	19 (18.27)
Fp2	18 (17.14)	19 (18.95)
Fz	19 (18.30)	20 (19.64)
O1	23 (22.35)	24 (23.52)
O2	21 (20.60)	22 (21.60)
Oz	20 (19.85)	20 (19.40)
P3	19 (18.75)	21 (20.61)
P4	18 (17.78)	21 (20.40)
P7	21 (20.54)	23 (22.13)
P8	19 (18.96)	22 (21.02)
Pz	20 (19.55)	21 (20.68)
T7	17 (16.11)	18 (17.65)
T8	17 (16.11)	18 (17.86)
TP10	16 (16.00)	18 (17.17)
TP9	20 (19.99)	21 (20.67)

Note. The selected and mean values of the optimal k_{max} for each EEG channel are shown, calculated for complete time series and sleep stage-specific segments.

Table F2

Selected and Mean k_{max} Values for Each fMRI ROI

fMRI ROI	Selected k_{max} for Complete Time Series (and Mean)	Selected k_{max} for Sleep Stage-Specific Segment (and Mean)
AG	13 (12.69)	25 (24.67)
COC	13 (12.40)	25 (24.01)
CGa	13 (12.55)	25 (24.10)
CGp	14 (13.03)	25 (24.44)
CuC	14 (13.02)	25 (24.83)
FMC	13 (12.20)	25 (24.56)
FOC	13 (12.20)	25 (24.12)
FrOC	13 (12.36)	25 (24.73)
FrP	13 (12.87)	25 (24.23)
HG	13 (12.85)	24 (23.80)
IFGpo	13 (12.46)	24 (23.80)
IFGpt	13 (12.35)	25 (24.50)
ITGa	13 (12.46)	25 (24.26)
ITGp	13 (12.09)	25 (24.49)
ITGt	13 (12.89)	25 (24.02)
IC	13 (12.73)	26 (25.50)
ICC	13 (12.64)	25 (24.18)
JLC	13 (12.75)	25 (24.67)
LOCi	14 (13.34)	25 (24.55)
LOCs	13 (12.51)	24 (23.40)
LG	13 (12.78)	25 (24.26)
MFG	13 (12.42)	24 (23.96)
MTGa	13 (12.69)	24 (23.68)
MTGp	14 (13.37)	25 (24.05)
MTGt	13 (12.54)	24 (23.83)
OFG	13 (12.49)	25 (24.32)
OP	13 (12.41)	24 (23.25)
PCG	13 (12.03)	24 (23.73)
PHGa	13 (12.25)	25 (24.33)
PHGp	13 (12.48)	26 (25.25)
POC	13 (12.43)	25 (24.07)
PP	13 (12.51)	24 (23.86)
PT	13 (12.79)	24 (23.30)
PoCG	13 (12.99)	25 (24.54)
PrCG	13 (12.70)	24 (23.12)
PC	13 (12.89)	26 (25.17)
SC	12 (11.96)	26 (25.64)
SFG	13 (12.86)	26 (25.51)
SPL	13 (12.28)	25 (24.84)
STGa	13 (12.46)	25 (24.14)
STGp	13 (12.77)	24 (23.29)

fMRI ROI	Selected k_{max} for Complete Time Series (and Mean)	Selected k_{max} for Sleep Stage-Specific Segment (and Mean)
SC	13 (12.98)	24 (23.51)
SMGa	13 (12.64)	24 (23.23)
SMGp	13 (12.67)	24 (23.55)
TFCa	13 (12.23)	26 (25.14)
TFCp	13 (12.50)	24 (23.77)
TOFC	13 (12.36)	24 (23.88)
TP	13 (12.72)	25 (24.13)

Note. Selected and mean values of the optimal k_{max} for each fMRI ROI are shown, calculated for complete time series and sleep stage-specific segments. The ROIs are the following: AG (Angular Gyrus), COC (Central Opercular Cortex), CGa (Cingulate Gyrus, anterior division), CGp (Cingulate Gyrus, posterior division), CuC (Cuneal Cortex), FMC (Frontal Medial Cortex), FOC (Frontal Opercular Cortex), FrOC (Frontal Orbital Cortex), FrP (Frontal Pole), HG (Heschl's Gyrus (includes H1 and H2)), IFGpo (Inferior Frontal Gyrus, pars opercularis), IFGpt (Inferior Frontal Gyrus, pars triangularis), ITGa (Inferior Temporal Gyrus, anterior division), ITGp (Inferior Temporal Gyrus, posterior division), ITGt (Inferior Temporal Gyrus, temporooccipital part), IC (Insular Cortex), ICC (Intracalcarine Cortex), JLC (Juxtapositional Lobule Cortex (formerly Supplementary Motor Cortex)), LOCi (Lateral Occipital Cortex, inferior division), LOCs (Lateral Occipital Cortex, superior division), LG (Lingual Gyrus), MFG (Middle Frontal Gyrus), MTGa (Middle Temporal Gyrus, anterior division), MTGp (Middle Temporal Gyrus, posterior division), MTGt (Middle Temporal Gyrus, temporooccipital part), OFG (Occipital Fusiform Gyrus), OP (Occipital Pole), PCG (Paracingulate Gyrus), PHGa (Parahippocampal Gyrus, anterior division), PHGp (Parahippocampal Gyrus, posterior division), POC (Parietal Opercular Cortex), PP (Planum Polare), PT (Planum Temporale), PoCG (Postcentral Gyrus), PrCG (Precentral Gyrus), PC (Precuneous Cortex), SC (Subcallosal Cortex), SFG (Superior Frontal Gyrus), SPL (Superior Parietal Lobule), STGa (Superior Temporal Gyrus, anterior division), STGp (Superior Temporal Gyrus, posterior division), SC (Supracalcarine Cortex), SMGa (Supramarginal Gyrus, anterior division), SMGp (Supramarginal Gyrus, posterior division), TFCa (Temporal Fusiform Cortex, anterior division), TFCp (Temporal Fusiform Cortex, posterior division), TOFC (Temporal Occipital Fusiform Cortex), TP (Temporal Pole) (Jenkinson et al., 2012).

Appendix G

Hurst Exponent (HE) Calculation Through the R/S and DFA Approaches

To obtain HE with the rescaled-range analysis, the process begins with splitting up a time series of length N into smaller segments that are not overlapping and have a certain length n, that is an integer, which evenly divides N (Ceballos & Largo, 2018). Then, each of the shorter segments undergoes a series of steps to calculate an average R/S value. To be more specific, the average of each segment is calculated first, as demonstrated in Equation 10:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ . \tag{10}$$

In the next steps, shown in Equation 11 and Equation 12, the computed average is subtracted from each value within the segment to achieve a series adjusted to the mean that can then be used for generating a series of cumulative deviations from the mean for each data point t (Z_t):

$$Y_t = X_t - \overline{X}, \tag{11}$$

$$Z_t = \sum_{i=1}^{t} Y_i \ , \tag{12}$$

where Y_t represents the deviation of data point X_t from the mean $\overline{X}$, and Z_t is the sum of these deviations up to each point of t = 1,2, …, n (Bal et al., 2021; Ceballos & Largo, 2018). After calculating the cumulative deviation, the range (R) of the series can be determined by the difference between the maximum and minimum values of the cumulative deviation, which is demonstrated in Equation 13:

$$R(n) = \text{maximum}(Z_1, Z_2, ..., Z_n) - \text{minimum}(Z_1, Z_2, ..., Z_n) \ . \tag{13}$$

The rescaled range (R/S) is then obtained by dividing the range (R) by the standard deviation (S), calculated for the previously mentioned, shorter segments of the time series of increasing length to observe how it scales with the size of the segment, as seen by Equation 14 and Equation 15:

$$S(n) = \sqrt{\left(\frac{1}{n}\right)\sum_{i=1}^{n}(X_i\text{-}m)^2} \ , \tag{14}$$

$$R/S = \frac{R(n)}{S(n)} \ , \tag{15}$$

where segment lengths usually increase in a logarithmic scale. Subsequently, for each segment length the R/S ratio is averaged over all segments (Bal et al., 2021).

Ultimately, through Equation 16, one can compute the HE by looking at the logarithmic ratio of the rescaled range (R/S) relative to the logarithm of the segment size (n):

$$\log(R/S)_n = \log(c) + HE\log(n) \ , \tag{16}$$

where c is the constant in the least square regression (Ceballos & Largo, 2018). More specifically, in this case, the HE is the slope of the straight line that one gets when applying the following power law to the data, which is shown in Equation 17:

$$E[R/S]=cn^{HE} \ , \tag{17}$$

where $E[R/S]$ is the value expected when $n \to \infty$ (Bal et al., 2021; Ceballos & Largo, 2018).

The other method mentioned in this study that relates to the HE is DFA (Peng et al., 1994; Peng et al., 1995). In this method, first, a cumulative sum is calculated for the time series using its mean value to obtain what is called the signal profile, demonstrated in Equation 18:

$$X_t = \sum_{i=1}^{t}(X_i\text{-}\overline{X}) \ \ , \tag{18}$$

where X_t corresponds the signal profile and $\overline{X}$ to the mean (Hardstone et al., 2012; Peng et al., 1994). Second, one has to determine a collection of evenly distributed and logarithmically spaced segment sizes with a minimum of four samples, as suggested by Peng et al. (1994), up to the signal's total length, as can be seen in Equation 19:

$$T = \{t_1, t_2, ..., t_k\} , \quad (19)$$

Then, for each segment size within this range, the signal profile X_t is divided into overlapping segments, each by 50%, creating a collection of time series of identical length (Hardstone et al., 2012; Peng et al., 1994). From each segment, a linear trend is removed via a least-squares fit, resulting in a detrended time series, from which the standard deviation is calculated, as shown in Equation 20:

$$\sigma(w_{detrend}) = \sqrt{\frac{1}{N\text{-}1}\sum_{i=1}^{N}(w_{detrend}(i) - \mu_{w_{detrend}})^2} , \quad (20)$$

in which case $\sigma(w_{detrend})$ is the standard deviation of the detrended time series $w_{detrend}$ (Hardstone et al., 2012; Peng et al., 1994). Furthermore, across all segments of equal size, the average of these standard deviations are calculated, leading to a fluctuation function, which can then be plotted on a double logarithmic plot against segment sizes T (Bryce & Sprague, 2012; Peng et al., 1994). Finally, DFA exponent, α, showing the time-series' scaling behavior, can be obtained from this plot as the slope of the trend line across the relevant time scales using linear regression (Hardstone et al., 2012; Peng et al., 1994).

Appendix H

Evaluation Metrics for Machine Learning Models

In machine learning, a classifier's accuracy is defined as the number of samples correctly classified by the model in relation to the total number of samples given to it, as shown in Equation 21:

$$\text{accuracy} = \frac{\text{number of samples classified correctly}}{\text{total number of samples}} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (21)$$

where TP and TN represent the true positives and negatives, whereas FP and FN stand for false positives and negatives (Hicks et al., 2022). Likewise, recall, which is known as the "true positive rate", measures the ratio of correctly made positive predictions against all positive instances, whereas precision is the ratio of correctly identified samples in a category and the number of samples classified in that category, detailed in Equation 22 and Equation 23:

$$\text{recall} = \frac{\text{number of true positives}}{\text{number of samples classified as positives}} = \frac{TP}{TP + FN}, \quad (22)$$

$$\text{precision} = \frac{\text{number of samples classified correctly in one class}}{\text{number of samples assigned to that class}} = \frac{TC}{TC + FC}, \quad (23)$$

where C represents the class, that may be positive (P), known as positive predictive value (PPV), or negative (N), called negative predictive value (NPV) (Hicks et al., 2022). This study applied the more prevalent interpretation of precision, PPV. Finally, the F1-score is yet another important metric in machine learning calculated with Equation 24:

$$\text{F1} = 2 * \frac{\text{precision} * \text{recall}}{\text{precision} + \text{recall}} = \frac{2 * TP}{2 * TP + FP + FN}, \quad (24)$$

where the F1 is a "harmonic mean of precision and recall" (Hicks et al., 2022).

Appendix I

Parameter Grid and Best Model Parameters for Machine Learning

Table I1

Parameter Grid and Best Model Parameters for Random Forest

Parameters	Optimal Parameter for EEG Features	Optimal Parameter for BOLD Features	Optimal Parameter for Combined Features
bootstrap: (True, False)	False	False	False
criterion: (gini, entropy)	entropy	gini	gini
max_features: (None, sqrt, log2)	log2	log2	sqrt
min_samples_split: (2, 5, 10)	2	2	2
min_samples_leaf: (1, 2, 5)	1	1	1
max_depth: (10, 25, 50)	10	10	10

Table I2

Parameter Grid and Best Model Parameters for Support Vector Machine

Parameters	Optimal Parameter EEG Features	Optimal Parameter for BOLD Features	Optimal Parameter for Combined Features
C:	1	1	1
(0.01, 0.1, 1)			
gamma:	0.1	0.01	0.01
(0.01, 0.1, 1)			
kernel:	rbf	rbf	rbf
(linear, rbf, poly)			