

Cross-participant prediction of vigilance stages through the combined use of wPLI and wSMI EEG functional connectivity metrics

Laura Sophie Imperatori^{1,2*}, Jacinthe Cataldi³, Monica Betta¹, Emiliano Ricciardi¹,

Robin A.A. Ince², Francesca Siclari³, Giulio Bernardi^{1,3*}

¹*MoMiLab, IMT School for Advanced Studies Lucca, Lucca, Italy,*

²*Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, United Kingdom,*

³*Center for Investigation and Research on Sleep, Lausanne University Hospital, Lausanne, Switzerland*

Accepted Manuscript

*** Correspondence**

Laura Sophie Imperatori
laurasophie.imperatori@gmail.com
IMT School for Advanced Studies
Piazza S. Francesco, 19
55100 Lucca, LU
Italy

Giulio Bernardi
giulio.bernardi@imtlucca.it
IMT School for Advanced Studies
Piazza S. Francesco, 19
55100 Lucca, LU
Italy

Accepted Manuscript

Abstract

Functional connectivity (FC) metrics describe brain inter-regional interactions and may complement information provided by common power-based analyses. Here we investigated whether the FC-metrics weighted Phase Lag Index (wPLI) and weighted Symbolic Mutual Information (wSMI) may unveil functional differences across four stages of vigilance – wakefulness (W), NREM-N2, NREM-N3 and REM sleep – with respect to each other and to power-based features. Moreover, we explored their possible contribution in identifying differences between stages characterized by distinct levels of consciousness (REM+W vs. N2+N3) or sensory disconnection (REM vs. W). Overnight sleep and resting-state wakefulness recordings from 24 healthy participants (27±6yrs, 13F) were analysed to extract power and FC-based features in six classical frequency bands. Cross-validated linear discriminant analyses (LDA) were applied to investigate the ability of extracted features to discriminate i) the four vigilance stages, ii) W+REM vs. N2+N3, and iii) W vs. REM. For the four-way vigilance stages classification, combining features based on power and both connectivity metrics significantly increased accuracy relative to considering only power, wPLI or wSMI features. Delta-power and connectivity (0.5-4Hz) represented the most relevant features for all the tested classifications, in line with a possible involvement of slow waves in consciousness and sensory disconnection. Sigma-FC, but not sigma-power (12-16Hz), was found to strongly contribute to the differentiation between states characterized by higher (W+REM) and lower (N2+N3) probabilities of conscious experiences. Finally, alpha-FC resulted as the most relevant FC-feature for distinguishing among wakefulness and REM sleep and may thus reflect the level of disconnection from the external environment.

Keywords:

EEG, Functional Connectivity, Sleep, Classification, NREM, REM, Consciousness, Disconnection.

Statement of Significance

This study shows that changes in brain activity across sleep stages may be better understood through the combined use of power and of functional connectivity (FC) metrics. Delta power and connectivity were found to be related to stage-dependent changes in consciousness and disconnection in line with previous works. In contrast to results obtained in subjects under anaesthesia or in patients with disorders of consciousness, we found that alpha-FC does not contribute to the classification of states with high and low levels of consciousness, and is instead highly relevant for distinguishing between wakefulness and REM sleep, a state typically characterized by vivid conscious experiences. Alpha-FC could therefore mark disconnection from the external environment rather than the level of consciousness per se.

Accepted Manuscript

Introduction

Classically, the transition from wakefulness to sleep and the alternation of the different sleep stages are described in terms of variations in electrophysiological brain activity, as measured using scalp electroencephalography (EEG). The falling asleep process is characterized by a shift from high-frequency, low-amplitude oscillations, to low-frequency, high-amplitude activity, with typical slow waves (0.5-4Hz) and spindles (12-16Hz)¹⁻³. An increase in the incidence and amplitude of regional slow waves then marks the deepening of Non-REM sleep. The occurrence of REM sleep, every 60-150min, is instead identified by a relative EEG desynchronization, with an increase in theta activity (4-8Hz), and by bursts of sawtooth-waves (2-5Hz) that commonly precede the occurrence of rapid eye movements⁴⁻⁶. All these changes may be efficiently captured through the analysis of signal power in classical frequency bands.

However, evidence indicates that the different stages of vigilance are also characterized by variations in the way distinct areas of the brain interact with each other (e.g.,⁷⁻¹¹). These differences can be explored using functional connectivity (FC) metrics, which identify statistical (undirected) associations among the EEG time-series of spatially distinct brain regions. Interestingly, distinct FC indices may allow the detection of partially different inter-regional interaction dynamics, with distinct weights of linear and nonlinear components, and may thus produce partially different results^{7,8}. However, it is unclear whether connectivity metrics, either analysed individually or in combination, can actually provide relevant or rather overlapping (i.e., redundant) information about the current stage of vigilance of an individual with respect to classical (power-based) measures of EEG brain activity.

The weighted Phase Lag Index (wPLI¹²) and the weighted Symbolic Mutual Information (wSMI¹³) are two robust FC approaches increasingly applied to MEG and EEG data because of their relative immunity to volume conduction. Thanks to this property, wPLI and wSMI are less likely to detect spurious functional interactions determined by the quasi-simultaneous conduction of real brain signals and of physiological or non-physiological noise to EEG-sensors distant from the actual source. Interestingly, both methods have been suggested to track the level of consciousness based on promising findings in patients with severe brain injury (*minimally conscious state, unresponsive wakefulness syndrome*) and in volunteers under anaesthesia¹³⁻¹⁷. Of note, we recently demonstrated that wPLI and wSMI present a different sensitivity to distinct FC dynamics and may thus provide complementary information regarding brain inter-regional interactions⁸. Specifically, while wPLI has an optimal sensitivity for dynamics presenting a mixture of linear and nonlinear components, wSMI has higher sensitivity to predominantly nonlinear dynamics.

In light of the above premises, here we set out to investigate whether wPLI and wSMI may allow the detection of FC differences across distinct stages of vigilance including wakefulness, light NREM sleep (N2), deep NREM sleep (N3) and REM sleep, and whether the provided information is redundant or complementary with respect to that granted by power-based indices of brain activity. To this aim, a supervised classification approach was used to identify the optimal combination of power- and FC-based features allowing to distinguish among vigilance stages. Crucially, this classification was cross-validated across different sets of participants, thus allowing to determine the ability of the predictive model to generalise across independent groups of subjects. For each individual feature, we also investigated its relevance, information content and degree of redundancy with other examined features. In addition, the present study offered us the opportunity to investigate the potential value of FC-derived indices as markers of the levels of consciousness and of sensory disconnection during

sleep. Indeed, the probability of having a conscious experience varies greatly between wakefulness and sleep as well as across the different sleep stages: subjects may report a conscious experience (a dream) in less than 30% of the awakenings from deep (N3) NREM sleep, while this percentage may get close to 100% in REM sleep¹⁸⁻²¹. Yet, while REM sleep is populated by rich and vivid conscious experiences that often resemble in many aspects those of wakefulness, it is also characterized by a relative disengagement (or ‘disconnection’) of the individual from the external environment^{18,22,23}. Therefore, we specifically investigated whether FC-based indices could significantly contribute to the classification of states characterized by high (wakefulness, REM) vs. low (NREM) levels of consciousness, as well as by ‘connected’ (wakefulness) and ‘disconnected’ (REM) consciousness²².

Methods

Participants. Twenty-four healthy adult participants (26.7 ± 5.7 yrs, 13F) underwent hd-EEG recordings (257 channels, 500Hz) during the night and while awake in bed after sleep (8am; 6min, eyes-closed). The overnight recordings were started at the usual bed-time of each subject and interrupted at 7am. Twelve subjects (25.5 ± 3.7 yrs, 6F; *dataset-1*) were studied in the context of a project aimed at exploring the effects of short-term visual deprivation on brain activity during sleep²⁴. Here we only included data from the control condition, in which subjects remained in the sleep laboratory and watched movies of their choice (selected from a pre-defined list) from 3 to 8pm. The other twelve subjects (27.9 ± 7.1 yrs, 7F; *dataset-2*) only arrived at the laboratory 2h prior to the beginning of the overnight recording.

All volunteers had a good sleep quality as assessed by the Pittsburgh Sleep Quality Index (PSQI score ≤ 5 ; ²⁵) and scored ≤ 10 points on the Epworth Sleepiness Scale (ESS; ²⁶). Volunteers were asked to maintain a regular sleep-wake schedule in the week preceding the overnight recording. Compliance was verified using wrist-worn actigraphy devices (MotionWatch 8, CamNtech). The study was approved by the ethical committee of the Lausanne University Hospital. Written informed consent was obtained from each subject.

Wake EEG recordings. Resting-state eyes-closed recordings collected in the morning, during wakefulness (W), were bandpass-filtered between 0.5 and 45 Hz. Each recording was divided into non-overlapping 5s epochs and visually inspected to identify and reject bad channels and epochs containing large artifacts (NetStation 5.3; Electrical Geodesic). An independent component analysis (ICA) was then performed to remove residual ocular, muscular, and electrocardiograph artifacts using EEGLAB²⁷. This procedure has been shown to reduce the impact of EEG artifacts on power computation and on the identification of individual graphoelements while producing negligible changes in physiological signals of interest^{28,29}. Rejected channels were interpolated using spherical splines. Finally, two 2s non-overlapping data segments were obtained from each 5s epoch by discarding the first 0.5s and the last 0.5s of data.

Sleep EEG recordings. Sleep EEG recordings were scored according to standard criteria using 30s epochs³⁰. For scoring purposes, electrodes located in the chin-cheek region were used to evaluate muscular activity, whereas four electrodes placed near the eyes were used to monitor horizontal and vertical eye movements¹. Recordings were bandpass filtered between 0.5 and 45 Hz. All epochs scored as N2, N3 (NREM) or REM were then extracted and visually inspected to remove bad channels (later replaced by spherical spline interpolation). N1 was not considered in the analysis, as it is a transitional state and therefore generally very short, accounting for only about 5% of total sleep time. As described above, an ICA-based procedure was used to reduce potential artifacts. Finally, each 30s epoch was divided into 15 non-overlapping 2s data segments.

Given that NREM and REM sleep are affected by homeostatic and circadian factors, and that REM sleep is especially represented toward the morning hours, further analyses were focused on NREM epochs extracted from the first half of the night and on REM epochs identified in the second half of the night. This approach also allowed us to minimize the impact of inter-subject differences in the relative distribution of NREM and REM epochs across the night (e.g., due to the lack of consistent REM sleep during the first sleep cycle in some participants).

Data bootstrapping. A bootstrapping procedure with replacement was applied to account for differences in the amount of data available for each stage of vigilance (see Supplementary Figure S1). Specifically, at each iteration of the bootstrapping procedure (Nboot=2000) and for each considered stage (W, N2, N3 and REM), fifteen 2s segments were randomly chosen, thus obtaining 30s-long data epochs. Of note, due to how wPLI is defined, connectivity can only be computed across multiple data segments¹². Thus, in our case, only one connectivity value was obtained for each set of 15 segments (and for each frequency band). For consistency, power and wSMI, were computed in each of the fifteen 2s segments and then averaged, thus obtaining one value for each set of 15 segments. A current-source-density transform was applied to all EEG data epochs using the CSD toolbox³¹, as described in previous works^{8,12,13}. This method provides a reference-independent signal and acts as a spatial filter, leading to a relatively improved spatial resolution³².

Power analysis. The Power Spectral Density (PSD; Welch's method, Hamming windows, 8 sections, 50% overlap) was calculated for each electrode in 2s data segments. Then, the signal power (POW) was computed for delta (0.5-4Hz), theta (4-8Hz), alpha (8-12Hz), sigma (12-16Hz), beta (18-25Hz), gamma (30-45Hz) and broad-band (0.5-45Hz) frequency bands by integrating the PSD values in the frequency ranges of interest. Relative power (relPOW) was computed by expressing the value of each frequency band as a percentage with respect to the total broad-band power. Finally, the mean was computed across the 15 segments of each 30s epoch and the median power was computed across all electrodes.

Connectivity analysis. For each 30s epoch, FC was computed in 2s data segments for each pair of electrodes, as detailed below. Then, the median connectivity across all channel pairs was obtained and used in subsequent analyses.

wPLI. The wPLI measures the extent to which phase angle differences between two time series $x(t)$ and $y(t)$ are distributed towards positive or negative parts of the imaginary axis in the complex plane (similar to the PLI^{12,33}). The underlying idea is that volume-conducted activity accounts for the greatest proportion of detected 0° or 180° phase differences between signals. Therefore, to obtain a conservative estimate for real, non-volume conducted activity, only phase angle distributions predominantly on the positive or negative side are considered. The PLI is defined as the absolute value of the sum of the signs of the imaginary part of the complex cross-spectral density S_{xy} of two real-valued signals $x(t)$ and $y(t)$ at trial t .

$$PLI = \left| \frac{\sum_{t=1}^n \text{sgn}(\text{imag}(S_{xy,t}))}{n} \right| \quad (1)$$

While PLI is already insensitive to zero-lag interactions, the weighted Phase-Lag Index¹² further addresses potential confounds caused by volume conduction, by scaling contributions of angle differences according to their distance from the real axis, as ‘almost-zero-lag’ interactions are considered as noise affecting real zero-lag interactions:

$$wPLI = \left| \frac{\sum_{t=1}^n |\text{imag}(S_{xy,t})| \text{sgn}(\text{imag}(S_{xy,t}))}{\sum_{t=1}^n |\text{imag}(S_{xy,t})|} \right| \quad (2)$$

The wPLI is based only on the imaginary component of the cross-spectrum, and thus implies robustness to noise compared to coherence, as uncorrelated noise sources cause an increase in signal power³⁴. Here wPLI was computed using the *Fieldtrip toolbox*³⁵ (multi-taper method fast Fourier transform, single Hanning taper¹², 0.5 Hz frequency resolution). For each frequency band, the mean value across the frequency-bins in the frequency range of interest was computed to obtain a single wPLI coupling value.

wSMI. The wSMI¹³ evaluates the extent to which two EEG signals present non-random joint fluctuations, suggesting sharing of information. The time series X and Y in all EEG channels are first transformed into sequences of discrete symbols (\hat{X} , \hat{Y}). The symbols are coded according to the trends in amplitudes of a specific predefined number of consecutive time points. We chose the kernel k to be 3, implying that the symbols are constituted of three elements, leading to $3!=6$ different potential symbols in total^{13,36}. The temporal separation of elements that constitute a symbol was set to be $\tau=[41, 21, 14, 10, 6, 3]$ frames, such that the maximum resolved frequency possible was $f_{max} = \frac{f_s}{k\tau}$. The

corresponding frequencies were $f_{max} = [4.1, 7.94, 11.90, 16.67, 27.78, 55.56]$ Hz, whereby - due to the band-pass-filtering of the signal up to 45Hz - the maximum resolved frequency for $\tau = 3$ was 45Hz.

Prior to wSMI computation, the signal was low-pass-filtered using a 6th order Butterworth IIR filter with an additional mirror padding ('*ft_preproc_padding*') of 1s before and after each individual segment to avoid potential filter edge-artifacts. For the analysis in frequency bands, a bandpass-filter (4th order Butterworth IIR filter) was used with the same padding scheme. The joint probability of each pair of symbols co-occurring in two different time-series is computed to estimate the symbolic mutual information (SMI) shared across two signals. To address volume conduction artifacts, the weighted symbolic mutual information disregards co-occurrences of identical or opposite-sign signals.

$$\text{wSMI}(\hat{X}, \hat{Y}) = \frac{1}{\log(k!)} \sum_{\hat{x} \in \hat{X}} \sum_{\hat{y} \in \hat{Y}} w(\hat{x}, \hat{y}) p(\hat{x}, \hat{y}) \log \frac{p(\hat{x}, \hat{y})}{p(\hat{x})p(\hat{y})} \quad (3)$$

The wSMI can lead to negative values, given that it is a weighted mutual information measure, a form of weighted relative entropy³⁷.

Statistical comparisons. As described above, for each bootstrap iteration, we computed the median power across electrodes and the median connectivity (wPLI/wSMI) across all pairs of channels in delta, theta, alpha, sigma, beta, and gamma frequency bands, thus obtaining 6 features per metric and 18 features in total (see Supplementary Figure 2 for frequency-resolved power and wPLI). For each individual feature, non-parametric paired permutation tests ($N_{perm_1} = 10,000$) were performed between mean values obtained from the bootstrap procedure for all vigilance stages. A Bonferroni-correction was applied for the number of tested comparisons, separately for power-, wPLI- and wSMI-based features ($N = 36$; *corrected* $p < 0.05$). Topographic comparisons between vigilance stages can be found in Supplementary Figures S3-6.

Classification Procedure. The subjects were divided into two equal-sized groups ($N = 12$) corresponding to *dataset-1* and *dataset-2*. A two-fold cross-validated LDA-based approach³⁸ was applied to investigate the ability of power- and FC-based features obtained from one dataset to discriminate vigilance stages in the other dataset (and vice-versa). Specifically, at each bootstrap iteration the classifier was trained on features extracted from a single 30s sampled epoch per stage of interest from each one of the 12 subjects of one dataset (1 or 2) and tested using features extracted from a single 30s epoch per stage of interest from each one of the 12 subjects of the other dataset (2 or 1). This procedure was repeated for all bootstrap iterations ($N_{boot} = 2,000$) to estimate the corresponding classification accuracy. The overall prediction accuracy was computed as the mean across all bootstrap iterations and then averaged across the two classification runs.

A forward selection procedure was performed to identify the best combination of features, henceforth referred to as the 'optimal feature set'. In particular, the procedure started with an empty feature set.

At each step of the forward selection procedure, the accuracies of classification models based on each available additional feature were computed. The feature giving rise to the highest classification accuracy was then added to the existing feature set. The selection was stopped when the added feature provided no further increase in either mean or median accuracy (accuracy variation ≤ 0). The final number of features included in each classification model is displayed in panel A of Figures 1, 3 and 4, together with the achieved mean accuracy. The full list of selected features for each optimal feature set is reported in Supplementary Table S1.

A permutation-based test was applied to determine whether accuracy values obtained from the optimal, selected feature-sets were significantly different from chance. Specifically, for each permutation ($N_{\text{perm}_2}=2,000$), the labels were shuffled in the training and in the test sets of the classifier and the mean accuracy estimate per permutation was obtained using the same bootstrapping procedure as in the original dataset. The true mean accuracy across bootstrap iterations was then compared to the null distribution to obtain the p-value expressing significance of accuracy relative to chance. A False Discovery Rate correction (FDR³⁹) was applied across tested models (FDR_q < 0.05; 7 comparisons). Moreover, the accuracies of models based on different feature-types (POW, wPLI, wSMI) were directly compared using multiple pairwise bootstrap tests. Specifically, for each pair of models, we computed the relative accuracy differences for every bootstrap iteration. We then computed the 2.5% and 97.5% confidence intervals (CI; $\alpha = 0.05$) for this distribution and determined whether zero (corresponding to no difference between models), was included (or not) in the CIs. In addition, p-values were derived from the bootstrap confidence intervals using the formula:

$$p = \min\left(\frac{2(1+\text{sum}(X \geq 0))}{1+N_{\text{boot}}}, \frac{2(1+\text{sum}(X \leq 0))}{1+N_{\text{boot}}}\right) \quad (4)$$

where X represents the distribution of differences⁴⁰. For each one of the three different comparisons of vigilance stages (W vs. N2 vs. N3 vs. REM, W+REM vs. N2+N3 and W vs. REM), an FDR correction was applied to account for multiple comparisons testing among all pairs of models (FDR_q < 0.05; 21 comparisons).

Mutual information analysis. In order to better understand which individual features carried more relevant information for distinguishing between the considered classes, we qualitatively compared the single-feature LDA performance (reflecting prediction across subjects) with the mutual information (MI) each feature contained about the classes (reflecting prediction within subjects). Specifically, MI was calculated and used to quantify how much information each individual feature contained about the three different comparisons of vigilance stages: W vs. N2 vs. N3 vs. REM, W+REM vs. N2+N3 and W vs. REM. Of note, we selected MI instead of a within-subject LDA approach, since the latter would have implied splitting the data of each subject in two sets, for training and testing respectively. Since, we only have few minutes of EEG recordings in wakefulness, this would have meant further reducing the usable data from this stage (Supplementary Figure S1).

The continuous feature values were binned based on ranks into 4 equal-population bins, and the standard discrete formulation of MI was used^{41,42}. A Miller-Madow bias correction was also applied⁴³. A data augmentation procedure was used to ensure that a minimum of 100 values was available in each subject for enabling MI computation. Specifically, 50 data epochs were repeatedly

sampled with replacement (2,000 iterations) from the previously generated data distribution (see “Data bootstrapping” section). Of note, this procedure was especially necessary for wakefulness data, since only few minutes were recorded in all subjects, whereas more than 25 min of data were available for all other stages in 23 of the 24 studied volunteers. Mean MI values (across bootstrap iterations) of each subject were averaged to obtain a group-level MI value. Finally, the obtained MI values were plotted against LDA accuracies in order to allow for a qualitative evaluation of the relative contribution of each feature to within- or between-subjects classification of the conditions of interest (Figure 5). Specifically, for these plots the single-feature LDA classification procedure was repeated by performing the training on the same augmented datasets obtained for MI computation across all 12 subjects in dataset-1 (2) and tested on one 30s epoch per subject in dataset-2 (1).

Partial information decomposition. We used partial information decomposition (PID)^{44,45} to determine the level of information about the stage of vigilance that was shared by all pairs of examined features. In fact, two features could convey the same, shared information, which is quantified as redundancy. Each feature could convey unique information that is not available from the other feature. Moreover, two features could also convey more information together than they do separately, i.e. if the relationship between the feature values is itself informative about the vigilance stage. This is quantified as synergy. PID was computed for each subject on the same extracted data as for MI.

Results

Classification of vigilance stages. For the classification across all considered vigilance stages (W, N2, N3, REM) the accuracy estimate obtained with each single metric was 66.63% for relPOW, 58.24% for wPLI and 60.31% for wSMI (Figure 1). The combination of two feature types increased accuracy to 75.90% for relPOW+wPLI, 74.90% for relPOW+wSMI, and 70.22% for wPLI+wSMI. The highest accuracy (80.49%) was obtained when all types of features were considered together (relPOW+wPLI+wSMI; accuracy relative to chance level: $FDRq < 0.05$). Figure 1B shows the relative differences between tested models. Of note, while the model including all feature types performed significantly better than relPOW, wPLI or wSMI alone ($FDRq < 0.05$), no significant differences were found with respect to models combining relPOW and one connectivity metric (wPLI or wSMI; $P > 0.05$, *uncorrected*), or combining wPLI and wSMI ($P < 0.05$, *uncorrected*). The three most relevant features for the relPOW+wPLI+wSMI model included elements from each of the considered metrics, namely delta-POW, sigma-wPLI and sigma-wSMI (Figure 1C). Statistical comparisons based on paired non-parametric permutation tests ($P < 0.05$, Bonferroni correction; Figure 2) showed that both NREM (N2/N3) and REM sleep were characterised by increased delta-POW ($N3 > N2 > REM > W$), with respect to W, while sigma-wPLI ($N2 > N3 > REM = W$) and sigma-wSMI ($N2 = N3 > REM = W$) were higher in NREM relative to both wakefulness and REM sleep. Additional relevant features included alpha-wSMI ($W > N3 > N2 > REM$) and delta-wPLI ($N2 > N3 = REM > W$).

[Figures 1, 2]

Classification of high vs. low level of consciousness. For the classification between stages associated with higher (W+REM) and lower (N2+N3) probability of conscious experiences, the classification accuracy estimate obtained with each single metric was 81.58% for relPOW, 75.53% for wPLI and 81.81% for wSMI (Figure 3). The combination of two feature types increased accuracy to 88.99% for relPOW+wPLI, 89.28% for relPOW+wSMI, and 83.94% for wPLI+wSMI. The highest accuracy (92.05%) was obtained when all feature types were included (relPOW+wPLI+wSMI; accuracy relative to chance level: $FDR_q < 0.05$). Figure 3B shows the relative differences between tested models. While the model including all feature types performed significantly better than relPOW or wPLI (but not wSMI; $P < 0.05$, *uncorrected*), no significant differences were found with respect to models combining relPOW and one connectivity metric (wPLI or wSMI; $P > 0.05$, *uncorrected*), or combining wPLI and wSMI ($P < 0.05$, *uncorrected*). The most relevant features for the relPOW+wPLI+wSMI model were sigma-wSMI, delta-POW and sigma-wPLI. Delta-wPLI also provided a minor contribution to overall model accuracy. Statistical comparisons based on paired non-parametric permutation tests ($P < 0.05$, Bonferroni correction; Figure 2) showed that both NREM (N2/N3) and REM-sleep were characterised by increased delta-POW (N3>N2> REM>W) and higher sigma-FC (wPLI/wSMI) in N2/N3-sleep with respect to both W and REM-sleep (no differences between W and REM).

[Figure 3]

Classification of consciousness with and without sensory disconnection. For the comparison between wakefulness ('connected' consciousness) and REM sleep ('disconnected' consciousness), the maximum classification accuracy obtained with each single metric was 92.91% for relative POW (relPOW), 77.29% for wPLI and 87.03% for wSMI (Figure 4). The combination of two feature types increased accuracy to 92.91% for relPOW+wPLI, 93.59% for relPOW+wSMI, and 87.03% for wPLI+wSMI. The model including all feature types (POW+wPLI+wSMI) achieved an accuracy similar to the one obtained using relPOW+wSMI (93.59%; accuracy relative to chance level: $FDR_{q_p} < 0.05$). Of note, the POW+wPLI+wSMI model did not perform significantly better than any of the other tested models, although it tended to achieve a higher accuracy relative to the one based on wPLI alone ($P < 0.05$, *uncorrected*; Figure 4B). In fact, wPLI-features also tended to perform worse than relPOW, relPOW+wPLI and relPOW+wSMI ($P < 0.05$, *uncorrected*). The most relevant features for the POW+wPLI+wSMI model resulted to be delta-POW, theta-POW and alpha-wSMI. Statistical comparisons based on paired non-parametric permutation tests ($P < 0.05$, Bonferroni correction; Figure 2) showed increased delta-POW and theta-POW and decreased alpha-wSMI in REM sleep with respect to wakefulness.

[Figure 4]

Contribution of individual features. In order to further evaluate which individual features carried more relevant information for distinguishing between considered classes (i.e., vigilance stages), we contrasted the single-feature LDA performance (regarding the predictability across subjects) with the mutual information each feature contained about the classes (within subjects' analysis). On the one hand, LDA tests the generalisation of predictive accuracy of a specific feature to a new set of subjects. On the other hand, MI is a within-subject analysis, determining the predictive power of the features within each individual. In this light, the two approaches provide related, but complementary information. Results obtained from this analysis are shown in Figure 5. It is important to note that most features were distributed along the diagonal of the plots, indicating a relative consistency in the relevance of tested features for within- and across-subjects classifications. In line with this, the Spearman's ρ between LDA and MI for the three plots were 0.89, 0.90 and 0.78, respectively ($p < 0.01$).

For the differentiation between all stages of vigilance (W, N2, N3, REM), delta-POW resulted as the best feature based on both its individual classification performance and MI value (59.83%, MI=0.93bits; Figure 5A). For the comparison between NREM (N2, N3) sleep and W+REM, sigma-wSMI resulted as the best feature (81.30%, MI=0.53bits; Figure 5B). Finally, for the comparison of W and REM sleep, alpha-wSMI alone achieved 87.17% accuracy (with the greatest recorded MI value of 0.73bits for this comparison), although a higher LDA accuracy was obtained by delta-POW.

[Figure 5]

Redundancy Analysis. In order to determine to what extent the statistical relationship in two features is common or overlapping, we used an information theoretic framework called *partial information decomposition* (PID)^{46,47}. Here, we assessed whether the information that any particular feature contains regarding the label (W vs. N2 vs. N3 vs. REM in Figure 6A, W+REM vs. N2+N3 in Figure 6B, W vs. REM in Figure 6C) is shared among two features or is unique to that feature (see Supplementary Figure S7 for corresponding synergy results). Results shown in Figure 6 indicate that especially power-based features were highly redundant with each other. Indeed, non-parametric permutation tests ($n\text{Perm}_3=10,000$) showed that redundancy among power-features was higher than that among wPLI and wSMI features, respectively ($42.98 \pm 7.51\%$ vs. $16.77 \pm 3.41\%$ vs. $28.28 \pm 7.62\%$; $p=0.0001$) for the comparison among all the four vigilance stages (W, N2, N3, REM). Power features were more redundant with respect to wSMI features and less redundant relative to wPLI features ($27.43 \pm 5.10\%$ vs. $19.56 \pm 3.30\%$; $p=0.0001$). Of note, wPLI features showed a low redundancy with respect to wSMI features ($18.54 \pm 3.30\%$), thus confirming that the two metrics are sensitive to distinct FC-dynamics. Similar differences were observed for the W+REM vs. N2+N3 and W vs. REM comparisons. Interestingly, in the W vs. REM case, both alpha-POW and alpha-wSMI appeared to be highly redundant with respect to delta-POW (83.44% and 72.27%, respectively) and with each other (72.42%). This may explain both why alpha-POW was not included in the forward selection procedure and the fact that the inclusion of alpha-wSMI determined only a marginal increase in overall classification accuracy (Figure 4).

[Figure 6]

Discussion

Power-based features are known to differ significantly across vigilance stages and may be used for the development of automated sleep scoring algorithms with a satisfactory level of accuracy⁴⁸⁻⁵⁰. However, previous work also showed that vigilance stages may differ in terms of functional connectivity (FC), and that distinct FC-metrics may also reveal different variations in inter-regional neural communication as a function of their specific sensitivity for linear and nonlinear interaction dynamics^{7,8}. Yet it was still unclear whether FC metrics actually provide relevant or rather redundant information with respect to classical quantitative measures of brain EEG activity. By using a linear discriminant analysis (LDA) with features based on signal power and two connectivity metrics with known distinct sensitivity for linear and nonlinear interactions (i.e., wPLI and wSMI), here we showed that the inclusion of connectivity-based features significantly increases the classification accuracy of vigilance stages. Moreover, we determined which specific individual features provide the best accuracy not only for distinguishing among all vigilance stages, but also for comparisons contrasting states associated with lower or higher probabilities of having a conscious experience (typically in the form of dreams during sleep), and states associated with more or less 'disconnection' from the external environment. In line with previous evidence^{16,17,51,52}, our results showed that delta-power carries relevant information for all of these comparisons. Moreover, our results indicate that sigma-FC may be especially relevant for distinguishing among vigilance stages respectively characterized by higher and lower probability of conscious experiences during sleep, while changes in alpha-FC may mainly mark the degree of sensory disconnection rather than the level of consciousness per se.

Contribution of connectivity-based metrics to the classification of vigilance stages. Relative variations in brain activity that characterize distinct vigilance stages may be efficiently captured through the analysis of signal power in classical frequency bands. Of note, however, evidence indicates that the different stages of vigilance are also accompanied by relative variations in the way distinct areas of the brain interact with each other⁷. Here we showed that information granted by FC analysis, as determined using wPLI and wSMI metrics, provides complementary information with respect to power-based features. In fact, while individual FC metrics showed a lower accuracy (58-60%) at discriminating the four examined stages of vigilance (W, N2, N3, REM) with respect to power-based features (~67%), their inclusion in the model considerably increased the overall accuracy, which reached ~80%. The adopted LDA classifier, at each one of its two folds, was trained on one group of subjects and tested on a separate group of individuals, thus supporting the generalisability of our observations.

Importantly, combining power-based features and features based on only one connectivity metric (either wPLI or wSMI) led to a ~10% increase in accuracy relative to power-features alone, while the addition of a second connectivity metric only determined a further ~5% gain in accuracy. However, only the combination of power-based features with both wPLI- and wSMI-based features determined a significant accuracy increase with respect to the classifier based on relative power alone, thus suggesting that combining multiple FC-features may be advantageous for the purpose of sleep-stage

classification. It should be noted however, that our main aim was to directly evaluate and compare the information content of simple power and connectivity features, without considering other indices often employed by previous studies on automated sleep stage classification⁵³. Such indices include EEG features extracted from either the time domain, the frequency domain, or the time-frequency domain, as well as complexity and non-linear indices. In this respect, we cannot exclude that other types of features could provide a similar information content with respect to connectivity-based indices. Further studies will be necessary to clarify this issue. Another aspect to consider is that different classifiers may have a different performance and provide partially different results. However, an exploratory analysis performed in our dataset, showed overall similar performance levels for LDA, quadratic discriminant analysis (QDA) and Naïve Bayes classifiers (Supplementary Figure S8), thus implying that the selected classification approach did not have a major impact on our results.

Interestingly, a specific evaluation of the contribution provided by individual features showed that delta-power is the most relevant feature for the cross-participant classification for all vigilance stages and carries the most information at the within-subject level. Other relevant features included several FC-features based on both wPLI (sigma-wPLI, delta-wPLI) and wSMI (sigma-wSMI, alpha-wSMI).

Overall, these results indicate that the inclusion of FC-based features in the classification of vigilance stages can lead to a higher accuracy with respect to using power-based features alone. Indeed, the model including features based on power and both connectivity metrics led to a significantly higher accuracy relative to models based on single feature types (power, wPLI, wSMI); however, we found no significant differences between the model encompassing all feature types and models including features based on power and only one FC-metric. Nevertheless, it should be kept in mind that a lack of statistical significance does not necessarily imply an equivalence of the compared models, and that our results also confirmed previous evidence indicating that wPLI and wSMI provide partially distinct and complementary information. In fact, we found that the two FC-metrics have a relatively low redundancy with respect to each other and a significantly different degree of redundancy with respect to signal power.

Classification of states with high or low levels of consciousness or disconnection. Previous works reported differences in delta^{15,17,54} and alpha^{17,55–57} connectivity between states characterized by high or low levels of consciousness. These studies compared healthy adult individuals and patients in *unresponsive wakefulness syndrome* (UWS) or *minimally conscious state* (MCS), or the same subjects before, after and during anaesthesia. However, any valid metric of consciousness should be able to identify differences between wakefulness with awareness and all states associated with a reduced level of consciousness, including dreamless sleep^{58,59}. Thus, here we investigated whether, in line with previous studies, FC-based indices could be used to distinguish between states characterized by a higher level of consciousness, that are wakefulness and REM sleep, and states associated with a relatively reduced probability of having a conscious experience, corresponding to N2 and N3 sleep²⁰. The obtained classification model led to a high accuracy (~92%), which was mainly driven by a subset of FC- and power-based features. In particular, the strongest contribution to accuracy was provided by sigma-wSMI and delta-power, while sigma-wPLI and delta-wPLI provided a smaller contribution. Of note, the contribution of delta power and connectivity is especially consistent with previous findings indicating a relationship between these parameters and changes in the level of consciousness during sleep and in pathological states or anaesthesia^{15,17,54,60}. Variations in sigma-connectivity have been less often reported, but may reflect here the occurrence and cortical ‘propagation’ of sleep spindles^{61,62}. In fact,

previous work indicates that changes in spindle parameters may be related to the presence of conscious experiences during NREM sleep¹⁹. It should be noted, though, that sigma-power did not significantly contribute to classification accuracy.

The additional classification between wakefulness and REM-sleep, respectively representing states of connected and disconnected consciousness, revealed an important contribution of alpha-FC (wSMI) to classification accuracy (other features included delta-power and theta-power). As mentioned above, this feature did not contribute instead to the distinction between states characterized by high or low levels of consciousness. Based on this finding, we suggest that alpha connectivity may mainly mark disconnection or disengagement from the external environment rather than the current level of consciousness, in line with recent findings describing a suppression in TMS-evoked alpha activity during disconnected consciousness in REM sleep²².

Interestingly, delta-power appeared among the most relevant features for all tested classifications. Stage-related variations in this parameter may depend on the occurrence and cortical spreading of sleep slow waves (0.5-4Hz). Indeed, the sleep slow wave, which marks temporary neuronal *off-periods* at cortical level, has been suggested to interrupt the causal flow of information among brain areas, thus impairing the possible emergence of conscious experiences in both NREM and REM sleep^{21,63,64}. Moreover, sensory-evoked slow-wave-like events (i.e., K-complexes) have been suggested to also have a direct role in the relative disconnection that characterize sleep through the quenching of incoming sensory information⁶⁵⁻⁶⁸. Of note, evidence indicates that delta waves may have a role in sensory disconnection also during REM sleep⁶⁹. Indeed, slow waves similar to those of NREM have been described in primary sensory cortices during this stage^{4,69,70}. In addition, recent work demonstrated that the processing of informative stimuli is significantly reduced during rapid eye movements⁷¹, which are in turn associated with peculiar centro-parietal delta waves called sawtooth waves⁴. Importantly, while artifacts related to eye movements could represent a non-neural source of delta activity in REM sleep, a relevant contribution to our results is unlikely. First, the ICA-based artifact rejection allowed to strongly reduce all sources of artifactual activity. Second, global power and FC values were computed using the median across electrodes, thus reducing the possible impact of residual extreme artifactual activity. Finally, a control analysis indicated that similar classification accuracies were obtained for power-based features computed using all electrodes or after exclusion of electrodes located around the eyes, on the cheeks and on the neck. Specifically, the accuracy obtained for the full set of electrodes (N=257) was $92.9 \pm 3.9\%$, while the value obtained using the same features and the reduced set of electrodes (N=165) was $92.1 \pm 4.0\%$ ($p > 0.05$; CI [-0.042, 0.063]). These observations indicate that the key contribution of delta activity to the distinction between wakefulness and REM sleep was likely due to the presence in the latter of sawtooth waves and/or slow-wave like activity.

Limitations of the study. In the four-way classification of vigilance stages, the model including features based on power and both connectivity metrics led to a significantly higher accuracy relative to models based on single feature types (power, wPLI, wSMI). However, we found no significant differences between the model encompassing all feature types and models including features based on power and only one FC-metric, as well as between these latter models and the model relying on power-based features alone. In this light, while our findings are consistent with previous observations, indicating that wPLI and wSMI provide useful and complementary information⁸, further investigations will be necessary to clarify the unique contribution of each of the two metrics in characterizing brain connectivity changes across vigilance stages.

While it has been shown that the probability of dreaming in NREM sleep is around 50% (across N2 and N3) as compared to 80-100% in REM sleep and wakefulness²¹, we did not directly assess the presence of conscious experiences and the level of consciousness in the present data. Similarly, we did not directly assess the level of relative sensory disconnection. Therefore, analyses investigating the power- and connectivity-based correlates of consciousness and disconnection should be considered as preliminary. Future studies should include within-stage analyses comparing conditions with or without conscious experiences and with low or high responsiveness to external stimuli (high/low arousal threshold). Moreover, here only the whole-brain, median power and connectivity across channels were used, and topographic differences in the distribution of evaluated features were not included in the LDA classifiers. However, recent studies clarified that local variations in brain activity may determine significant changes in the level of consciousness within and across sleep stages^{19,21}. Thus, the inclusion of topographic information could be expected to increase classification accuracy. Relatedly, it would be also important for future studies to investigate whether considering a smaller sensor layout would lead to similar results with respect to those obtained here using hd-EEG with 257 electrodes⁷².

Importantly, accuracy values obtained in our analyses are likely to represent an overestimation of true accuracies for the tested models. In fact, the same data sets used for feature selection were also used for model testing. We note however that our present aim was not to obtain an exact estimate of accuracy to determine the absolute validity of our tested models, but rather to evaluate the relative contribution of distinct power and connectivity features. Accuracy values obtained using distinct training (N=8), validation (N=4) and testing (N=12) sets are reported in Supplementary Figure S9. Importantly, while this procedure led to overall lower accuracy values (potentially also due to the smaller number of subjects included in the training set), the main selected features are largely consistent with those obtained in our main analyses reported in Figures 1, 3 and 4 (also see Supplementary Figure S10).

Conclusions

Our results demonstrate that changes in brain activity across sleep stages are better characterised by the combined use of the signal power and the two connectivity metrics wPLI and wSMI, relative to both individual connectivity approaches and power-based indices alone. In line with previous studies regarding the neural correlates of consciousness, present results also showed that variations in delta power and connectivity are among the most relevant classification features. On the other hand, in contrast to previous works, alpha-connectivity did not contribute to the classification of states with high and low levels of consciousness, and was instead found to represent one of the best features for distinguishing between wakefulness and REM sleep, a state typically characterized by vivid conscious experiences. In this light, present results suggest that connectivity changes in the alpha range could mark disconnection from the external environment rather than the level of consciousness per se.

Overall, our results indicate that connectivity metrics provide relevant, complementary information with respect to metrics based on signal power, and also show that the application of connectivity approaches with different sensitivity to linear and nonlinear interaction dynamics may allow to achieve a more complete description of brain activity across different behavioural and vigilance states.

References

1. Siclari F, Bernardi G, Riedner BA, LaRocque JJ, Benca RM, Tononi G. Two Distinct Synchronization Processes in the Transition to Sleep: A High-Density Electroencephalographic Study. *Sleep*. 2014;37(10):1621-1637. doi:10.5665/sleep.4070
2. Marzano C, Moroni F, Gorgoni M, Nobili L, Ferrara M, De Gennaro L. How we fall asleep: Regional and temporal differences in electroencephalographic synchronization at sleep onset. *Sleep Med*. 2013. doi:10.1016/j.sleep.2013.05.021
3. Ogilvie RD. The process of falling asleep. *Sleep Med Rev*. 2001;5(3):247-270. doi:10.1053/smr.2001.0145
4. Bernardi G, Betta M, Ricciardi E, Pietrini P, Tononi G, Siclari F. Regional Delta Waves In Human Rapid Eye Movement Sleep. *J Neurosci*. 2019;39(14):2686-2697. doi:10.1523/JNEUROSCI.2298-18.2019
5. Aserinsky E, Kleitman N. Regularly occurring periods of eye motility, and concomitant phenomena, during sleep. *Science (80-)*. 1953. doi:10.1126/science.118.3062.273
6. Dement W, Kleitman N. The relation of eye movements during sleep to dream activity: An objective method for the study of dreaming. *J Exp Psychol*. 1957;53(5):339-346. doi:10.1037/h0048189
7. Migliorelli C, Bachiller A, Andrade AG, et al. Alterations in EEG connectivity in healthy young adults provide an indicator of sleep depth. *Sleep*. 2019;42(6). doi:10.1093/sleep/zsz081
8. Imperatori LS, Betta M, Cecchetti L, et al. EEG functional connectivity metrics wPLI and wSMI account for distinct types of brain functional interactions. *Sci Reports* 2019 91. 2019;9(1):8894. doi:10.1038/s41598-019-45289-7
9. Rusterholz T, Achermann P, Dürri R, Koenig T, Tarokh L. Global field synchronization in gamma range of the sleep EEG tracks sleep depth: Artifact introduced by a rectangular analysis window. *J Neurosci Methods*. 2017;284:21-26. doi:10.1016/j.jneumeth.2017.04.002
10. Vecchio F, Miraglia F, Gorgoni M, et al. Cortical connectivity modulation during sleep onset: A study via graph theory on EEG data. *Hum Brain Mapp*. 2017;38(11):5456-5464. doi:10.1002/hbm.23736
11. Langheim FJP, Murphy M, Riedner BA, Tononi G. Functional connectivity in slow-wave

- sleep: Identification of synchronous cortical activity during wakefulness and sleep using time series analysis of electroencephalographic data. *J Sleep Res.* 2011. doi:10.1111/j.1365-2869.2011.00911.x
12. Vinck M, Oostenveld R, Van Wingerden M, Battaglia F, Pennartz CMA. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage.* 2011;55(4):1548-1565. doi:10.1016/j.neuroimage.2011.01.055
 13. King JR, Sitt JD, Faugeras F, et al. Information sharing in the brain indexes consciousness in noncommunicative patients. *Curr Biol.* 2013;23:1914-1919. doi:10.1016/j.cub.2013.07.075
 14. Chennu S, O'Connor S, Adapa R, Menon DK, Bekinschtein TA. Brain Connectivity Dissociates Responsiveness from Drug Exposure during Propofol-Induced Transitions of Consciousness. *PLoS Comput Biol.* 2016;12(1):1-17. doi:10.1371/journal.pcbi.1004669
 15. Chennu S, Finioia P, Kamau E, et al. Spectral Signatures of Reorganised Brain Networks in Disorders of Consciousness. *PLoS Comput Biol.* 2014;10(10). doi:10.1371/journal.pcbi.1003887
 16. Lee M, Sanders RD, Yeom S-K, et al. Network Properties in Transitions of Consciousness during Propofol-induced Sedation. *Sci Rep.* 2017;7(1):16791. doi:10.1038/s41598-017-15082-5
 17. Lee M, Sanders RD, Yeom SK, et al. Change in functional networks for transitions between states of consciousness during midazolam-induced sedation. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS.* ; 2017:958-961. doi:10.1109/EMBC.2017.8036984
 18. Nir Y, Tononi G. Dreaming and the brain: from phenomenology to neurophysiology. *Trends Cogn Sci.* 2010;14(2):88-100. doi:10.1016/j.tics.2009.12.001
 19. Siclari F, Bernardi G, Cataldi J, Tononi G. Dreaming in NREM sleep: A high-density EEG study of slow waves and spindles. *J Neurosci.* 2018. doi:10.1523/JNEUROSCI.0855-18.2018
 20. Siclari F, LaRocque JJ, Postle BR, Tononi G. Assessing sleep consciousness within subjects using a serial awakening paradigm. *Front Psychol.* 2013. doi:10.3389/fpsyg.2013.00542
 21. Siclari F, Baird B, Perogamvros L, et al. The neural correlates of dreaming. *Nat Neurosci.* 2017;20(6):872-878. doi:10.1038/nn.4545
 22. Darracq M, Funk CM, Polyakov D, et al. Evoked Alpha Power is Reduced in Disconnected Consciousness During Sleep and Anesthesia. *Sci Rep.* 2018;8(1):16664. doi:10.1038/s41598-018-34957-9
 23. Bonnet MH, Moore SE. The threshold of sleep: Perception of sleep as a function of time asleep and auditory threshold. *Sleep.* 1982. doi:10.1093/sleep/5.3.267
 24. Bernardi G, Betta M, Cataldi J, et al. Visual imagery and visual perception induce similar changes in occipital slow waves of sleep. *J Neurophysiol.* 2019;121(6):2140-2152. doi:10.1152/jn.00085.2019

25. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: A new instrument for psychiatric practice and research. *Psychiatry Res.* 1989. doi:10.1016/0165-1781(89)90047-4
26. Johns MW. A New Method for Measuring Daytime Sleepiness: The Epworth Sleepiness Scale. *Sleep.* 1991;14(6):540-545. doi:10.1093/sleep/14.6.540
27. Delorme A, Makeig S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods.* 2004;134(1):9-21. doi:10.1016/j.jneumeth.2003.10.009
28. Iriarte J, Urrestarazu E, Valencia M, et al. Independent component analysis as a tool to eliminate artifacts in EEG: A quantitative study. *J Clin Neurophysiol.* 2003. doi:10.1097/00004691-200307000-00004
29. Romero S, Mañanas MA, Clos S, Gimenez S, Barbanoj MJ. Reduction of EEG Artifacts by ICA in Different Sleep Stages. In: *Annual International Conference of the IEEE Engineering in Medicine and Biology - Proceedings.* ; 2003. doi:10.1109/iembs.2003.1280467
30. Iber C, Ancoli-Israel S, A C. *The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology and Technical Specifications.*; 2007.
31. Kayser J, Tenke CE. Principal components analysis of Laplacian waveforms as a generic method for identifying ERP generator patterns: I. Evaluation with auditory oddball tasks. *Clin Neurophysiol.* 2006;117(2):348-368. doi:10.1016/j.clinph.2005.08.034
32. Nunez PL, Srinivasan R. *Electric Fields of the Brain: The Neurophysics of EEG.*; 2009. doi:10.1093/acprof:oso/9780195050387.001.0001
33. Stam CJ, Nolte G, Daffertshofer A. Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Hum Brain Mapp.* 2007;28(11):1178-1193. doi:10.1002/hbm.20346
34. Peraza LR, Asghar AUR, Green G, Halliday DM. Volume conduction effects in brain network inference from electroencephalographic recordings using phase lag index. *J Neurosci Methods.* 2012;207(2):189-199. doi:10.1016/j.jneumeth.2012.04.007
35. Oostenveld R, Fries P, Maris E, et al. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data, FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Comput Intell Neurosci.* 2011. doi:10.1155/2011/156869, 10.1155/2011/156869
36. Sitt JD, King JR, El Karoui I, et al. Large scale screening of neural signatures of consciousness in patients in a vegetative or minimally conscious state. *Brain.* 2014;137(8):2258-2270. doi:10.1093/brain/awu141
37. Kvålseth TO. The relative useful information measure: Some comments. *Inf Sci (Ny).* 1991;56(1-3):35-38. doi:10.1016/0020-0255(91)90022-M
38. Jamalabadi H, Alizadeh S, Schönauer M, Leibold C, Gais S. Classification based hypothesis testing in neuroscience: Below-chance level classification rates and overlooked statistical properties of linear parametric classifiers. *Hum Brain Mapp.* 2016. doi:10.1002/hbm.23140

39. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995;57(1):289-300. doi:10.1111/j.2517-6161.1995.tb02031.x
40. Davison A. Bootstrap Methods and Their Application| Computational Statistics, Machine Learning and Information Science. 2018.
41. Ince RAA, Giordano BL, Kayser C, Rousselet GA, Gross J, Schyns PG. A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Hum Brain Mapp*. 2017;38(3):1541-1573. doi:10.1002/hbm.23471
42. Cover TM, Thomas JA. Entropy, Relative Entropy and Mutual Information. In: *Elements of Information Theory*. ; 2003. doi:10.1002/0471200611.ch2
43. Miller. Note on the bias of information estimates. *Inf theory Psychol Probl methods*. 1955. <https://ci.nii.ac.jp/naid/10015446623/>. Accessed December 8, 2019.
44. Williams PL, Beer RD. *Nonnegative Decomposition of Multivariate Information*. <https://arxiv.org/pdf/1004.2515.pdf>. Accessed December 8, 2019.
45. Ince R. Measuring Multivariate Redundant Information with Pointwise Common Change in Surprisal. *Entropy*. 2017;19(7):318. doi:10.3390/e19070318
46. Allen B, Stacey B, Bar-Yam Y. Multiscale Information Theory and the Marginal Utility of Information. *Entropy*. 2017;19(6):273. doi:10.3390/e19060273
47. G. James R, J. Ellison C, P. Crutchfield J. dit: a Python package for discrete information theory. *J Open Source Softw*. 2018. doi:10.21105/joss.00738
48. Koley B, Dey D. An ensemble system for automatic sleep stage classification using single channel EEG signal. *Comput Biol Med*. 2012. doi:10.1016/j.combiomed.2012.09.012
49. Berthomier C, Drouot X, Herman-Stoica M, et al. Automatic analysis of single-channel sleep EEG: Validation in healthy individuals. *Sleep*. 2007. doi:10.1093/sleep/30.11.1587
50. Zoubek L, Charbonnier S, Lesecq S, Buguet A, Chapotot F. Feature selection for sleep/wake stages classification using data driven methods. *Biomed Signal Process Control*. 2007. doi:10.1016/j.bspc.2007.05.005
51. Massimini M, Huber R, Ferrarelli F, Hill S, Tononi G. The Sleep Slow Oscillation as a Traveling Wave. *J Neurosci*. 2004;24(31):6862-6870. doi:10.1523/JNEUROSCI.1318-04.2004
52. Chennu S, Finoia P, Kamau E, et al. Spectral Signatures of Reorganised Brain Networks in Disorders of Consciousness. *PLoS Comput Biol*. 2014;10(10). doi:10.1371/journal.pcbi.1003887
53. Aboalayon KAI, Faezipour M, Almuhammadi WS, Moslehpour S. Sleep stage classification using EEG signal analysis: A comprehensive survey and new investigation. *Entropy*. 2016. doi:10.3390/e18090272
54. Lee M, Sanders RD, Yeom S-K, et al. Network properties in transitions of consciousness

- during propofol-induced sedation. *Sci Rep.* 2017;7(1):16791.
55. Chennu S, O'Connor S, Adapa R, Menon DK, Bekinschtein TA. Brain connectivity dissociates responsiveness from drug exposure during propofol-induced transitions of consciousness. *PLoS Comput Biol.* 2016;12(1):e1004669.
 56. Blain-Moraes S, Lee U, Ku S, Noh G, Mashour GA. Electroencephalographic effects of ketamine on power, cross-frequency coupling, and connectivity in the alpha bandwidth. *Front Syst Neurosci.* 2014;8:114. doi:10.3389/fnsys.2014.00114
 57. Comsa IM, Bekinschtein TA, Chennu S. Transient Topographical Dynamics of the Electroencephalogram Predict Brain Connectivity and Behavioural Responsiveness During Drowsiness. *Brain Topogr.* 2018;0(0):0. doi:10.1007/s10548-018-0689-9
 58. Demertzi A, Tagliazucchi E, Dehaene S, Deco G, Barttfeld P. Human consciousness is supported by dynamic complex patterns of brain signal coordination. 2019;(February):1-12.
 59. Sarasso S, Rosanova M, Casali AG, et al. Quantifying cortical EEG responses to TMS in (un)consciousness. *Clin EEG Neurosci.* 2014;45(1):40-49. doi:10.1177/1550059413513723
 60. Massimini M, Ferrarelli F, Huber R, Esser SK, Singh H, Tononi G. Breakdown of cortical effective connectivity during sleep. *Science.* 2005;309(5744):2228-2232. doi:10.1126/science.1117256
 61. Andrillon T, Nir Y, Staba RJ, et al. Sleep spindles in humans: Insights from intracranial EEG and unit recordings. *J Neurosci.* 2011. doi:10.1523/JNEUROSCI.2604-11.2011
 62. Muller L, Piantoni G, Koller D, Cash SS, Halgren E, Sejnowski TJ. Rotating waves during human sleep spindles organize global patterns of activity that repeat precisely through the night. *Elife.* 2016. doi:10.7554/eLife.17267.001
 63. Tononi G, Massimini M. Why does consciousness fade in early sleep? In: *Annals of the New York Academy of Sciences.* ; 2008. doi:10.1196/annals.1417.024
 64. Pigorini A, Sarasso S, Proserpio P, et al. Bistability breaks-off deterministic responses to intracortical stimulation during non-REM sleep. *Neuroimage.* 2015;112:105-113. doi:10.1016/j.neuroimage.2015.02.056
 65. Laurino M, Menicucci D, Piarulli A, et al. Disentangling different functional roles of evoked K-complex components: Mapping the sleeping brain while quenching sensory processing. *Neuroimage.* 2014. doi:10.1016/j.neuroimage.2013.10.030
 66. Halász P. The K-complex as a special reactive sleep slow wave - A theoretical update. *Sleep Med Rev.* 2016;29:34-40. doi:10.1016/j.smrv.2015.09.004
 67. Cash SS, Halgren E, Dehghani N, et al. The human K-complex represents an isolated cortical down-state. *Science (80-).* 2009. doi:10.1126/science.1169626
 68. Laurino M, Piarulli A, Menicucci D, Gemignani A. Local Gamma Activity During Non-REM Sleep in the Context of Sensory Evoked K-Complexes. *Front Neurosci.* 2019. doi:10.3389/fnins.2019.01094

69. Funk CM, Honjoh S, Rodriguez AV, Cirelli C, Tononi G. Local Slow Waves in Superficial Layers of Primary Cortical Areas during REM Sleep. *Curr Biol*. 2016;26(3):396-403. doi:10.1016/J.CUB.2015.11.062
70. Baird B, Castelnovo A, Riedner BA, et al. Human Rapid Eye Movement Sleep Shows Local Increases in Low-Frequency Oscillations and Global Decreases in High-Frequency Oscillations Compared to Resting Wakefulness. *eNeuro*. 2018;5(4). doi:10.1523/ENEURO.0293-18.2018
71. Koroma M, Lacaux C, Andrillon T, Legendre G, Léger D, Kouider S. Sleepers Selectively Suppress Informative Inputs during Rapid Eye Movements. *Curr Biol*. 2020. doi:10.1016/j.cub.2020.04.047
72. Engemann DA, Raimondo F, King J-R, et al. Robust EEG-based cross-site and cross-protocol classification of states of consciousness. *Brain*. 2018;141(11):3179-3192. doi:10.1093/brain/awy251

Accepted Manuscript

Figure Captions

Figure 1. Four-way classification of W, N2, N3 and REM vigilance stages. (A) LDA accuracy for tested feature sets in the classification of W vs. N2 vs. N3 vs. REM sleep. The top values reported in each box indicate the mean accuracy across bootstrap iterations (marked by a white dot), while the bottom values indicate the number of included features F based on the forward selection procedure. (B) Differences among models based on different feature-sets (bootstrap tests). The dark-colored red (row > column) cells indicate significant differences between compared models ($FDRq < 0.05$), while the light-colored cells indicate $P < 0.05$, *uncorrected*. For these cells, the corresponding confidence intervals (2.5% and 97.5% CI; $\alpha = 0.05$) are shown. The grey cells indicate non-significant differences ($P > 0.05$). (C) Forward feature selection for the classification of W vs. N2 vs. N3 vs. REM sleep (whole-brain analysis). The blue continuous line represents the mean classification accuracy across bootstrap iterations, while the dashed lines represent the 2.5% and 97.5% confidence interval boundaries on the accuracy estimate. The dashed grey vertical line indicates the threshold for which inclusion of additional features did not improve either mean or median classification accuracy.

Figure 2. Statistical comparisons between stages of vigilance in delta, theta, alpha, sigma, beta and gamma frequency bands for whole-brain (A) relative power, (B) wPLI and (C) wSMI. Red dots indicate subjects of dataset-1, while blue dots indicate subjects of dataset-2. * $p < 0.05$, Bonferroni correction across all performed comparisons for each considered metric.

Figure 3. Two-way classification of W+REM vs. N2+N3 vigilance stages. (A) LDA accuracy for tested feature sets in the classification of W+REM vs. N2+N3 vigilance stages. The top values reported in each box indicate the mean accuracy across bootstrap iterations (marked by a white dot), while the bottom values indicate the number of included features F based on the forward selection procedure. (B) Differences among models based on different feature-sets (bootstrap tests). The dark-colored red (row > column) cells indicate significant differences between compared models ($FDRq < 0.05$), while the light-colored cells indicate $P < 0.05$, *uncorrected*. For these cells, the corresponding confidence intervals (2.5% and 97.5% CI; $\alpha = 0.05$) are shown. The grey cells indicate non-significant differences ($P > 0.05$). (C) Forward feature selection for the classification of W+REM vs. N2+N3 vigilance stages. The green continuous line represents mean classification accuracy across bootstrap iterations, while the dashed lines represent the 2.5% and 97.5% confidence interval boundaries on the accuracy estimate. The dashed grey vertical line indicates the threshold for which inclusion of additional features did not improve either mean or median classification accuracy.

Figure 4. Two-way classification of W and REM vigilance stages. (A) LDA accuracy for tested feature sets in the classification of W vs. REM-sleep. The top values reported in each box indicate the mean accuracy across bootstrap iterations (marked by a white dot), while the bottom values indicate the number of included features F based on the forward selection procedure. (B) Differences among models based on different feature-sets (bootstrap tests). The light-red (row > column) and the light-blue (row < column) cells indicate $P < 0.05$, *uncorrected* (no effects survived after FDR correction for multiple comparisons). For these cells, the corresponding confidence intervals (2.5% and 97.5% CI; $\alpha = 0.05$) are shown. The grey cells indicate non-significant differences ($P > 0.05$). (C) Forward feature selection for the classification of W vs. REM-sleep. The red continuous line represents mean classification accuracy across bootstrap iterations, while the dashed lines represent the 2.5% and

97.5% confidence interval boundaries on the accuracy estimate. The dashed grey vertical line indicates the threshold for which inclusion of additional features did not improve either mean or median classification accuracy.

Figure 5. Mutual Information vs. LDA classifier accuracy. Each plot represents the single-feature LDA performance (y-axis) with respect to the mean within-participant MI conveyed by the feature about the vigilance classes (x-axis) for all three tested comparisons: (A) W vs. N2 vs. N3 vs. REM, (B) W+REM vs. N2+N3, (C) W vs. REM. For each point, the horizontal black lines represent the standard error of the mean for MI, while the vertical lines indicate the standard deviation across LDA bootstrap estimates.

Figure 6. Relative Redundancy among features (redundancy as computed by partial information decomposition divided by the sum of redundancy, synergy, unique X and unique Y information values). The three matrices display the results based on partial information decomposition (PID) for the three tested comparisons: A) W vs. N2 vs. N3 vs. REM, B) W+REM vs. N2+N3, C) W vs. REM. A higher value (darker colour) indicates higher redundancy between two features.

Accepted Manuscript

Acknowledgments

This work was supported by the Swiss National Science Foundation (Ambizione Grant [PZ00P3_173955] to F.S.), the Divesa Foundation Switzerland (to F.S.), the Pierre-Mercier Foundation for Science (to F.S.), the Bourse Pro-Femme of the University of Lausanne (to F.S.), the Foundation for the University of Lausanne (to F.S and G.B.), the Wellcome Trust [214120/Z/18/Z] (to R.A.A.I.) and an International Brain Research Organization Pan-European Regional Committee short-term postdoctoral fellowship (to G.B.). The authors wish to thank Giacomo Handjaras and Denis Engemann for their contribution to the definition of methodologies adopted in the present work.

Author Contributions

Conceptualization: L.S.I., G.B.; *Data acquisition:* J.C., G.B., M.B., F.S.; *Data preprocessing:* J.C., G.B., M.B., L.S.I.; *Data analysis:* L.S.I.; *Methodology:* L.S.I., R.A.A.I., G.B.; *Results interpretation:* L.S.I., G.B., F.S., R.A.A.I.; *Wrote the manuscript:* L.S.I., G.B.; *Revised and edited the manuscript:* All authors.

Disclosure Statement

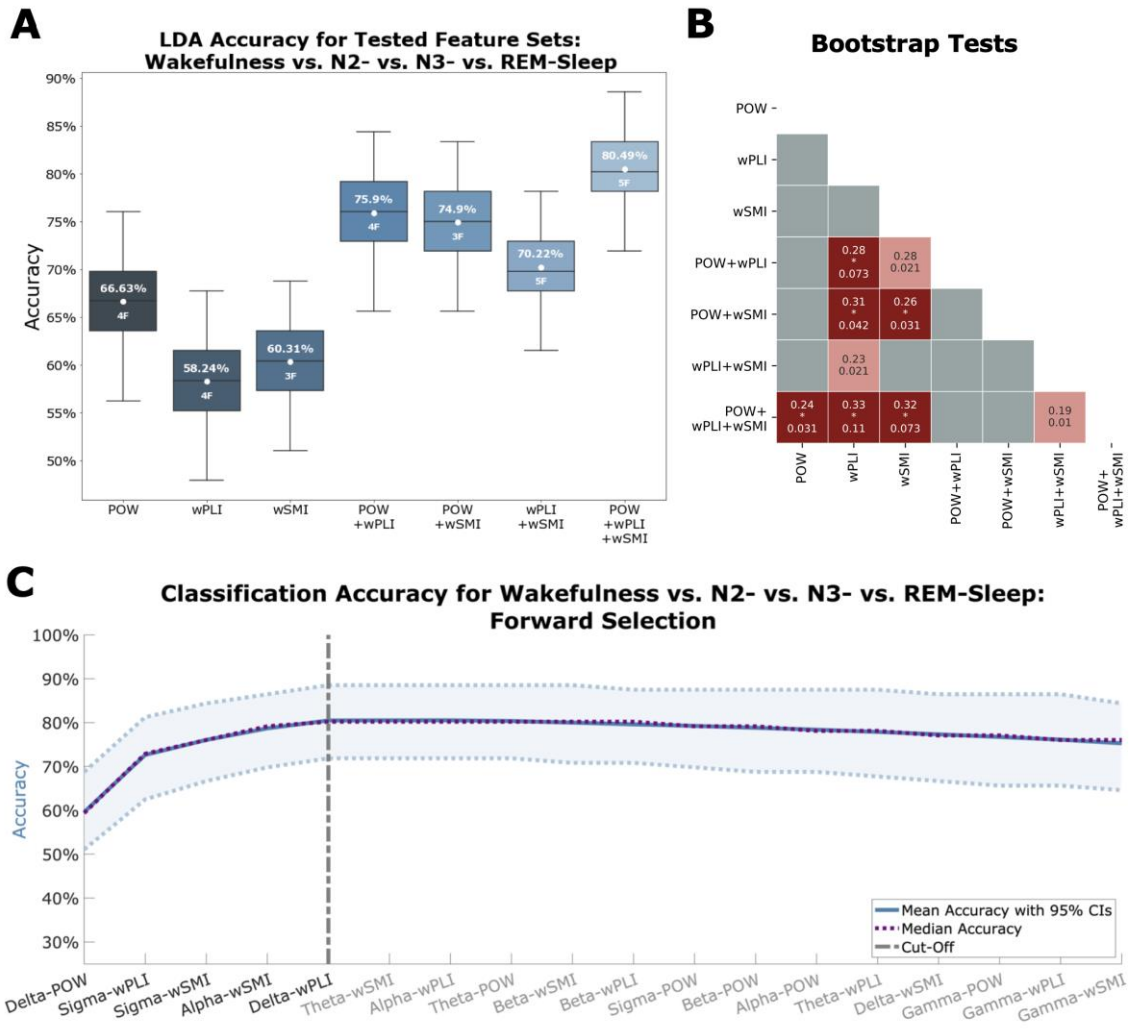
Non-financial Disclosure: none. Financial Disclosure: none.

Data Availability

The data and scripts that support the findings of this study are available on request from the first corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Accepted Manuscript

Figure 1



Accepted

Figure 2

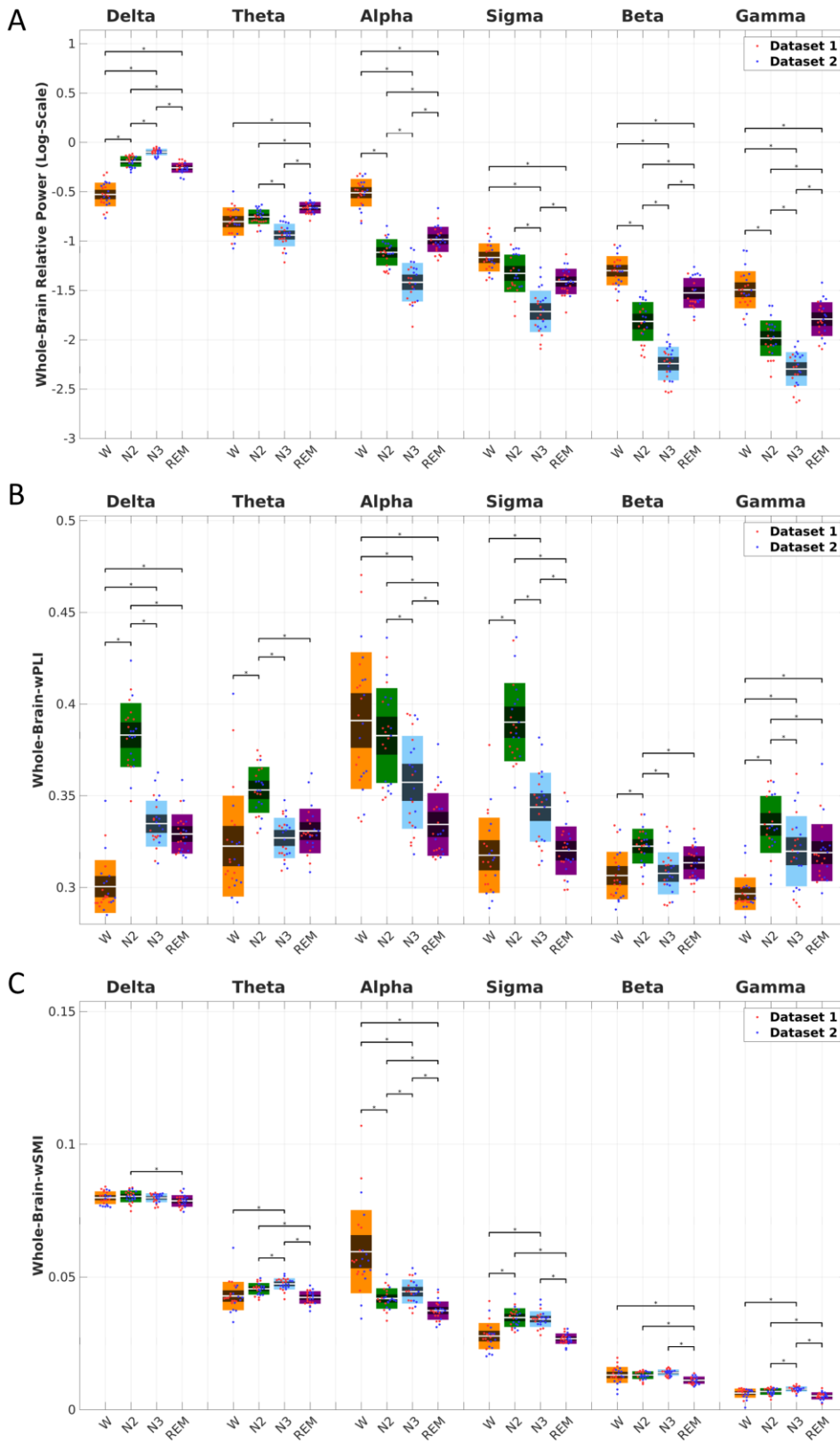
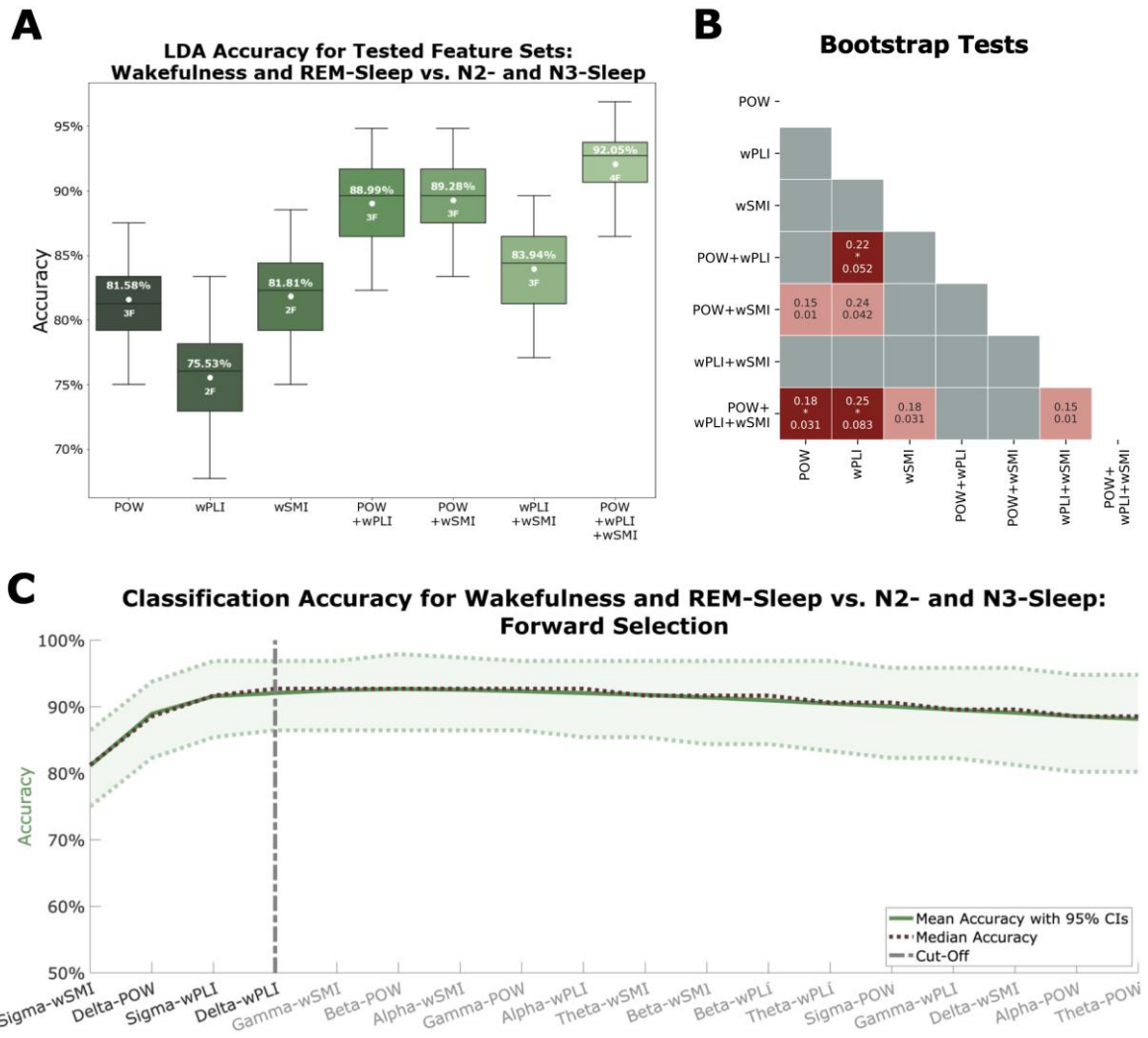
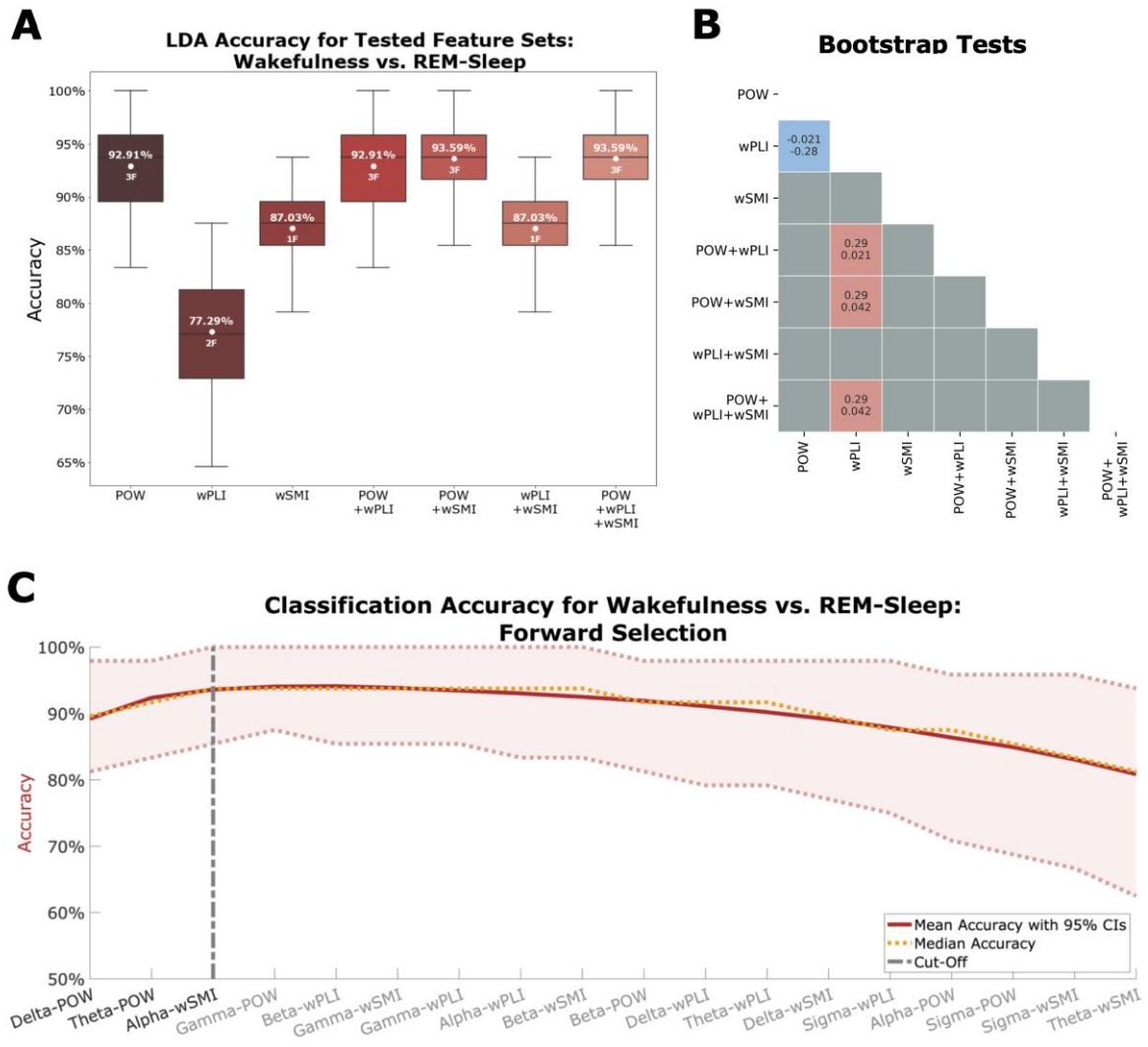


Figure 3



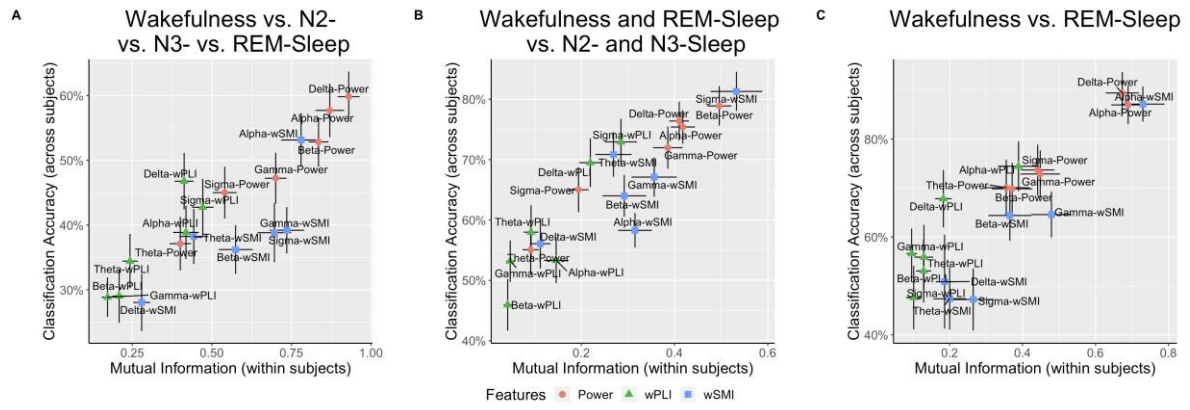
Accepted

Figure 4



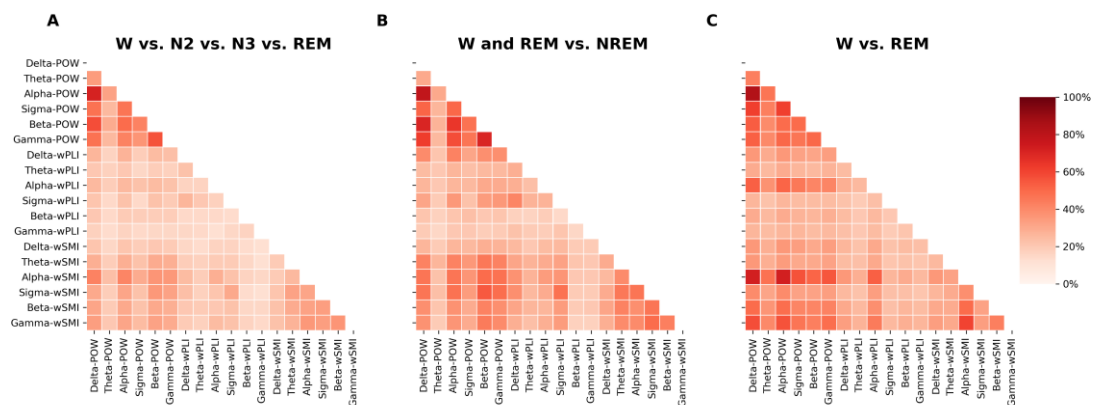
Accepted

Figure 5



Accepted Manuscript

Figure 6



Accepted Manuscript