Predicting Microsleep States Using EEG Inter-Channel Relationships

Abdul Baseer Buriro, *Student Member, IEEE*, Reza Shoorangiz, *Member, IEEE*, Stephen J. Weddell, *Senior Member, IEEE*, and Richard D. Jones, *Fellow, IEEE*

***Abstract*— A microsleep is a brief and an involuntary sleep-related loss of consciousness of up to 15 s. We investigated the performances of seven pairwise inter-channel relationships–covariance, Pearson's correlation coefficient, wavelet cross-spectral power, wavelet coherence, joint entropy, mutual information, and phase synchronization index–in continuous prediction of microsleep states from EEG. These relationships were used as the feature sets of a linear discriminant analysis (LDA) and a linear support vector machine classifiers. Priors for both classifiers were incorporated to address the class imbalance in the training data sets. Each feature set was extracted from a 5-s window of EEG with the step of 0.25 s and was demeaned with respect to the mean of first 2 min. The sequential forward selection (SFS) method, based on a serial combination of the correlation coefficient, Fisher score-based filter, and an LDA-based wrapper, was used to select features from each training set. The comparison was based on 16-channel EEG data from eight subjects who had performed a 1-D visuomotor task for two 1-h sessions. The prediction performances were evaluated using leave-one-subject-out cross-validation. For both classifiers, non-normalized feature sets were found to perform better than normalized feature sets. Furthermore, demeaning the non-normalized features considerably improved the prediction performance. Overall, the LDA classifier with joint entropy features resulted in the best average prediction performances (phi, AUC_{PR}, and AUC_{ROC}) of (0.47, 0.50, and 0.95). Joint entropy between O1 and O2 from theta frequency band was the most informative feature.**

***Index Terms*— EEG, microsleep, inter-channel relationships, LDA, LSVM, class imbalance.**

I. Introduction

WE DEFINE microsleeps as complete and unintentional sleep-related losses of consciousness of up to 15 s. They are accompanied by behavioral signs of eye closure, droopy eyes, and total loss of visuomotor responsiveness [1], [2], which are quite distinctive from the more tonic states of drowsiness (tendency to fall asleep) and fatigue (disinclination to responsiveness) [3], [4]. Studies have shown that healthy and non-sleep-deprived can also have frequent microsleeps [2], [4]. Furthermore, the high correlation between the probability of accidents and the duration of microsleeps has been reported [5]. In extended-attention monotonous activities such as driving, the consequences of microsleeps are often fatal. Microsleep related accidents can potentially be avoided if they are noninvasively and accurately predicted.

EEG is being used for detection of microsleeps due to its high temporal resolution of changes in brain activity. Spectral power features have frequently been used in EEG-based microsleep state detection [1], [6]–[8]. Compared to other neural networks, a long-short-term-memory (LSTM) recurrent neural network (RNN) resulted in the best performance metrics (phi, AUC_{PR}, AUC_{ROC}) of (0.38, 0.43, 0.84) [1]. Stacking of 7 linear discriminant analysis (LDA) classifiers resulted in performance metrics (phi, AUC_{PR}, AUC_{ROC}) of (0.39, 0.43, 0.84) on pruned EEG data [6]. Whereas, stacking of 7 echo state networks (ESN) resulted in performance metrics (phi, AUC_{PR}, AUC_{ROC}) of (0.44, 0.45, 0.88) [8]. These studies used a 2-s window to extract features from EEG, principal component analysis (PCA) to reduce feature space and had a detection resolution of 1.0 s for microsleep states.

Similarly, spectral power features have also been used in EEG-based microsleep state prediction [9], [10]. With single LDA classifier, performance metrics (phi, AUC_{PR}, AUC_{ROC}) of (0.33, 0.36, 0.90) were achieved with mutual information-based greedy-forward-feature-selection algorithm [9], and of (0.34, 0.36, 0.90) with Bayesian multi-subject factor analysis for feature reduction [10]. In both studies, a ~5-s window was used to extract features. The prediction time (τ) was 0.25 s ahead with a prediction temporal resolution of 0.25 s.

Golz *et al.* [7] fused spectral and delay vector variances of 7 EEG, 2 EOG, and 3 eye-tracking signals per eye (pupil size, x and y gaze coordinates) and achieved an accuracy of ~0.91 with radial basis function support vector machine (RBF-SVM) on classification of microsleep and alert events. This promising but erroneous accuracy was achieved by balancing the test data and cross-validation was performed on concatenated data from all the subjects. In doing so, independence of the test and training data, and hence generalization accuracy of the system, were lost. Similarly, with spectral features and a claimed accuracy of 0.88 on the prediction of

Manuscript received March 18, 2018; revised July 2, 2018 and October 20, 2018; accepted October 20, 2018. Date of publication October 30, 2018; date of current version December 6, 2018. *(Corresponding author: Abdul Baseer Buriro).*

The authors are with the Department of Electrical and Computer Engineering, University of Canterbury, Christchurch 8041, New Zealand, and also with the New Zealand Brain Research Institute, Christchurch 8011, New Zealand (e-mail: abdul.buriro@nzbri.org; reza.shoorangiz@nzbri.org; steve.weddell@canterbury.ac.nz; richard.jones@canterbury.ac.nz).

Digital Object Identifier 10.1109/TNSRE.2018.2878587

microsleep events [11]. Lin *et al.* [12] achieved an accuracy of ~0.78 on determining the effectiveness of providing feedback during behavioral lapses (i.e., microsleeps) in a discrete task of sustained-attention. These studies, however, are limited to class-balanced binary data sets.

Due to the multivariate property of EEG signals, it was considered that inter-channel relationships might be advantageous over independently extracted features from individual channels. Such inter-channel relationships are also known as functional and effective connectivity [13] between EEG electrodes or brain anatomical regions and may or may not be normalized [14]. Normalized EEG inter-channel relationships have been widely used to analyze brain function with different stimuli, e.g., visual oddball task [15]. Different measures of connectivity have been compared on simulated data [16] and real EEG data [17], [18]. The comparisons were to differentiate causal and non-causal connections [16], quantify the level of synchronization [17], and understand connectivity patterns [18]. Connectivity patterns of alert versus drowsy [19], [20] and connectivity patterns of microsleep onset versus offset have been reported as significantly different [21]. However, these studies have been restricted to offline characterization, i.e., statistical comparison between equal samples of two classes taken from the whole data. Unfortunately, such characterization cannot be utilized in real-time detection/prediction of microsleeps.

In emotion classification, however, connectivity patterns have been used as features. Lee and Hsieh [22] compared three EEG-based functional connectivity patterns and achieved the best accuracy of 0.82 with phase synchronization index and quadratic discriminant analysis (QDA) classifier. Similarly, with different connectivity and event-related potential (ERP) features, the LDA classifier resulted in an accuracy of 0.89 [23]. The accuracies, though impressive, are based on small and class-balanced data. Functional connectivity patterns with RBF-SVM resulted in accuracies of 0.92 and 0.97, respectively, on the classification of mental fatigue and alert states on a 1-h simulated driving task and on a psychomotor vigilance task (PVT) [24]. Although these accuracies are impressive, it is important to note that they have been obtained on class-balanced data from a discrete task (i.e., first and last 5 min). Also, although data independence in classification was maintained by leave-one-out cross-validation, amount (percentage) of data used in feature selection was unclear.

From joint entropy features between EEG channels and with an LDA classifier, we were able to predict (0.25-s ahead) microsleep states from continuous EEG with a mean phi accuracy of 0.38 [25]. This was ~15% better than individually extracted spectral features [9], [10] and 65% better than mutual information features [25].

As far as we are aware, we are the first to have investigated EEG inter-channel relationships in online classification/prediction of microsleep states. The aim of the current study was to investigate (1) the prediction performances of seven model-free, bivariate, and symmetric inter-channel relationships as feature sets to LDA and linear SVM (LSVM) classifiers and (2) present a preprocessing step in feature space to address heterogeneity inherent in EEG and reaction times over multiple subjects and sessions. In addition to simplicity and computational ease, model-free (generally symmetric) inter-channel relationships are reliable, accurate in the presence of common noise [14], [26], and robust against a common reference [27] and data decimation [16], and therefore were preferred in the investigation.

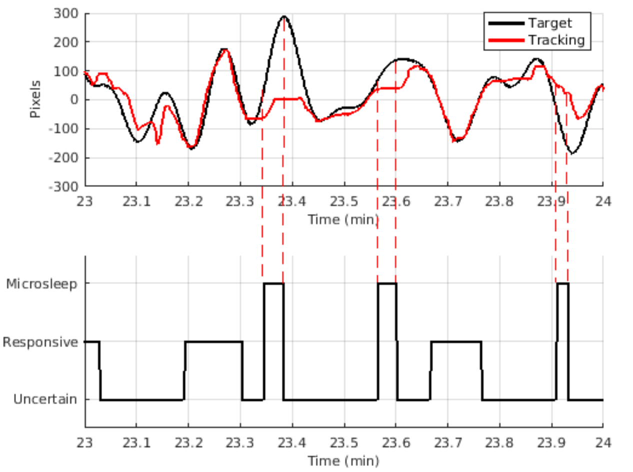

Fig. 1. Illustration of tracking performance and corresponding gold standard. The microsleeps are instances of unresponsiveness, as indicated by either essentially flat or incoherent tracking.

II. Methods

A. Data

The original data were from a previous study [2], comprising behavioral and EEG recordings over two 1-h sessions, one week apart, from 15 non-sleep-deprived healthy male subjects, aged 18-36 years, and with no neurological or sleep disorder. Data in the current study were from a subset of the 8 subjects who had had at least one definite microsleep over the two sessions.

Each subject performed a 1-D preview tracking task. The task was to track a pseudo-random target with minimum error. EEG was sampled at 256 Hz from 16 channels, namely Fp1, F3, F7, Fp2, F4, F8, C3, C4, P3, P4, T3, T5, T4, T6, O1, and O2, placed as per the international 10-20 system. Face video at 25 fps and tracking performance at 64 Hz were recorded.

B. Gold Standard

A gold standard, shown in Fig. 1, comprising 3 classes – responsive, microsleep, and uncertain – was formed from tracking performance and independent expert video ratings. The video recordings were conservatively rated by Peiris *et al.* [2], without knowledge of the corresponding tracking performances. He rated video on a 6-level scale using criteria similar to those of Wierwille and Ellsworth (1994). Levels 1–6 were marked alert, distracted, forced eye closure while alert, drowsy, deep drowsy, and sleep respectively.

Coherent tracking irrespective of the video rating was labeled as responsive. Incoherent tracking (i.e., mean absolute error $>$ 3 cm lasting for 1 s) or unresponsiveness (i.e., tracking speed $<$ 0.1*target speed) along with a video rating of deep

TABLE I
TOTAL MICROSLEEPS (BOTH SESSIONS) FOR EACH SUBJECT AND THE CORRESPONDING IMBALANCE RATIO

Sub	Microsleeps		Responsive durations (min)	Microsleep/ Responsive states
	Events	Duration (s)		
1	60	1011.3	56.6	4045/13570
2	29	121.6	74.2	486/17796
3	5	6.3	84.7	25/20335
4	18	57.3	95.5	229/22905
5	36	106.6	91.7	426/22004
6	25	103.6	68.8	414/16510
7	41	92.3	79.3	369/19043
8	33	573.3	21.7	2293/5184

Number of states = 4*Duration (temporal resolution of 0.25 s).

drowsiness was labeled as microsleep. Epochs that did not fall unequivovally into either of these classes were labeled as uncertain [9] and discarded during training and testing.

The incidence and duration of microsleeps for all 8 subjects (the two 1-h sessions combined) are shown in Table I.

C. EEG Preprocessing

EEG signals were band-pass filtered from 1 to 45 Hz and re-referenced to common average to improve the SNR. Artifacts were removed using artifact-subspace reconstruction (ASR) [28]. ASR requires a clean data to be used as calibration/base data to remove artifacts from rest of the data. The clean data were selected based on z-score ≤ 5 of EEG. The noisy segments of the data were decomposed into principal components (PCs), which then were projected into the calibration data's space by using its covariance matrix. PCs which represented high-amplitude artifacts were removed based on a threshold derived from the calibration data. Remaining PCs were then back-projected into EEG channel space. Considering non-stationary associated with EEG, it was segmented into 2-min epochs with 50% overlap and ASR was applied to each epoch independently. Calibration data of each epoch was found and then was used to clean the same epoch. The epochs were then concatenated together to have a cleaned set of original EEG data. The overlapping parts of consecutive epochs were averaged to avoid discontinuity. Canonical correlation analysis [29] was finally used to remove muscles artifacts [9]. Artifact-free EEG signals were then decomposed into EEG sub-bands (delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–45 Hz)), decimated to 128 Hz to reduce processing time, and segmented to 5-s epochs with steps of 0.25 s. This defines the lower operational limit of microsleep events (i.e., ≥ 250 ms) and a temporal resolution of 0.25 s. In contrast to earlier work [1], [6], [8], epochs of 5 s were used to ensure smooth and reliable connectivity estimates and to compare microsleep state prediction performances based on spectral features [9], [10]. In addition, the best classification accuracy in detecting ongoing microsleep events has been reported with an epoch of 5-s EEG [7].

D. Covariance and Correlation

Covariance and correlation (Pearson's correlation coefficient) between two equal lengths EEG time series X and Y with N samples (window length) are defined as

$$C_{XY} = \frac{1}{N-1}\sum_{i=1}^{1}(X(i)-\mu_X)(Y(i)-\mu_Y), \quad (1)$$

$$r_{XY} = \frac{C_{XY}}{\sigma_X\sigma_Y}, \quad (2)$$

where μ_X and μ_Y are the means and σ_X and σ_Y are the standard deviations of time series X and Y respectively.

E. Cross-Spectral Power and Coherence

Cross-spectral power and coherence were calculated using wavelet transform due to its variable window size and better immunity to noise over Fourier transform [30], [31]. Wavelet cross-spectrum W^{YY} and wavelet coherence C^{XY} between two equal lengths EEG time series X and Y with N samples are defined as

$$W^{XY} = \langle\left|\langle W_n^X(f)W_n^{Y*}(f)\rangle_f\right|\rangle_N, \quad (3)$$

$$C^{XY} = \langle\frac{\left|\langle W_n^X(f)W_n^{Y*}(f)\rangle_f\right|}{\sqrt{\langle|W_n^X(f)|^2\rangle_f\langle|W_n^Y(f)|^2\rangle_f}}\rangle_N, \quad (4)$$

where $W_n^X(f)$ and $W_n^Y(f)$ are the wavelet auto-spectra of time series X and Y respectively. f is the frequency index and n is the time index. The numerator and denominator of Equation (4) are required to be smoothed separately, otherwise the quantity will always be unity. Smoothing can be carried out over time, scale, or both and can be simple averaging [31]–[34]. Smoothing in scale/frequency direction has been empirically found to be more effective than smoothing in time direction [35]. For continuously long-time series (e.g., EEG), performing wavelet transform prior to the segmentation avoids edge effects and is suitable for online implementation. For consistency, both Equation (3) and (4) were smoothed over individual frequency bands using expectation $\langle.\rangle_f$. Finally, time expectation $\langle.\rangle_N$ was performed on smoothed cross-spectral powers and coherences to obtain features for the corresponding window.

The Morlet wavelet was used as the mother wavelet as it has good time-frequency localization [34]. With dimensionless frequency ω_0 and dimensionless time η, it is defined as

$$\psi_0 = \pi^{0.25}e^{-i\omega_0\eta}e^{-0.5\eta^2}. \quad (5)$$

Frequency-to-scale (s) conversion was performed according to $s = \sqrt{\omega_0^2+2}/4\pi f$, where $\omega_0 = 6$.

F. Joint Entropy and Mutual Information

Joint entropy H and mutual information I between two equal lengths EEG time series X and Y with N samples are defined as

$$H(X,Y) = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{XY}(i,j)\ln P_{XY}(i,j), \tag{6}$$

$$I(X,Y) = \sum_{i=1}^{N}\sum_{j=1}^{N} P_{XY}(i,j)\ln\left(\frac{P_{XY}(i,j)}{P_X(i)P_Y(j)}\right), \tag{7}$$

where P_{XY} is the joint probability density function (pdf) of X and Y, P_X and P_Y are marginal pdfs of X and Y respectively.

Entropies and mutual information were estimated using a K-nearest neighbor (kNN). It is a commonly used entropy estimator and has been reported to outperform for a small number of data points typically of the order of 100-1000 across noise levels [36]. Entropy via kNN is estimated [37] as

$$H(X) = \frac{1}{N}\sum_{i=1}^{N} \ln \epsilon_i \psi(k) + \ln \upsilon + \ln N \tag{8}$$

where N and k are the number of samples and nearest neighbor, respectively. $\epsilon_i = \|x_i - kx_i\|$ is the k^{th} neighborhood distance and ψ is digamma function. The volume of the one-dimensional unit ball (υ) is defined as

$$\upsilon = \frac{\pi^{1/2}}{\Gamma\left(\frac{1}{2}+1\right)}, \tag{9}$$

where Γ is the gamma function.

To account for bias-variance trade-off, we selected $k = 3$ from optimal values of 2–4 [36]–[38]. kNN-based entropies were estimated using the ITE toolbox [39]. Mutual information, however, was estimated from marginal entropies of EEG time series X and Y, and their joint entropy.

G. Phase Synchronization Index

An analytic time series Z for a univariate EEG time series X is defined as

$$Z = X + j\mathcal{H}[X] = Ae^{j\theta}, \tag{10}$$

where A and $\theta = \tan^{-1}\left(\frac{\mathcal{H}[X]}{X}\right)$ are the instantaneous amplitude and instantaneous phase (IP) of X respectively, and $\mathcal{H}[X]$ is the Hilbert transform of X.

The analytic time series Z is calculated in three steps [40], [41] by: (1) taking the discrete Fourier transform (DFT) Y of the time series X, (2) doubling the positive frequency bins and zeroing the negative frequency bins, and (3) taking the inverse discrete Fourier transform (IDFT) Z of Y. This results in phase shift (delay) of $\pi/2$ which does not alter the spectral distribution of the signal and, therefore, the Hilbert transform can be regarded as an all-pass filter.

The mean phase coherence (MPC) is the most commonly used and suitable PSI for analyzing EEG signals at low sampling rates [41] and is estimated as

$$\lambda = \left|\langle e^{j\pi}\rangle_N\right|, \tag{11}$$

where $\langle . \rangle_N$ is the expectation over samples (window size), and ϕ is the IP difference, also known as relative phase between the two equal lengths EEG time series X and Y, defined as $\phi = \theta_X - \theta_Y$.

H. Features Preprocessing

There can be a considerable difference in EEG signals and reaction times over different subjects and sessions. Together, these usually result in varying and subject-specific distributions and, consequently, poorer classification performances. At the start of each session, all subjects are generally alert and responsive, and, for the same task, homogeneity in EEG signals and reaction times can be assumed. Thus first 2 min of features were treated as baselines and averages thereof were subtracted from the respective feature sets in that session. The aim of this was to match data means without affecting their distributions. In contrast, the standard normalization of features changes data distributions and can, therefore, remove some of the information. Thus, features are not divided by the standard deviations of first 2 min.

I. Features Selection

M channels of EEG decomposed into B sub-bands gives $\frac{B*M(M-1)}{2}$ pairwise inter-channel relationships, i.e., 600 features per epoch in each feature set. Large numbers of features can introduce high variance in test results, and redundant and irrelevant features can substantially deteriorate classification performance. Features were therefore selected from each training set (i.e., concatenated data from 7 subjects). Initially, a subset of features was created by pruning the linearly-correlated features ($|r| > 0.9$). Finally, features were selected by using the sequential forward selection (SFS) [42] method. A Fisher-score-based filter [43] was used to rank the features of the subset as per their individual discriminatory power and the best single feature was selected. LDA-based wrapper was then used to sequentially select the ranked features. AUC_{ROC}, due to its robustness under skewed-class distributions [44] was used as the performance metric (objective function). The mean AUC_{ROC} of the 5-fold cross-validation with the top-ranked feature was calculated and saved. The successive feature was then combined with the top-ranked feature and selected if the combined mean AUC_{ROC} was improved, otherwise, it was discarded. The process was iterated until a stopping criterion was met. At every iteration, based on relative AUC_{ROC}, a feature was either selected or discarded. In this study, the stopping criterion was a logical *OR* of a maximum number of selected features (70) and allowed number of successive iterations (70) in which no performance improvement was observed.

J. Classification

Nonlinear classifiers have a computational complexity of $O(N^3)$ [45], [46], where N is the number of training instances, and propensity to over-fit [47], which can be avoided at extra computational costs involved in selecting optimal values of a regularization and kernel parameters. For large and high-dimensional data sets, linear classifiers have, however,

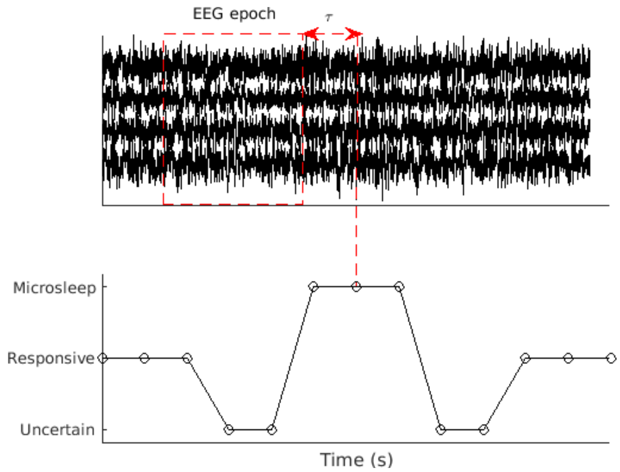

Fig. 2. Prediction of gold standard (microsleep and responsive state) from corresponding EEG epoch. Prediction time is represented by τ, which in this research is 0.25s. Events comprise a varying number of states and, hence, are of variable length.

resulted in similar performances to nonlinear classifiers and are fast [48]. The average data length in this study is $\sim$ 121,000 samples per training session and feature set, where nonlinear classifiers become infeasible to use.

Two linear classifiers – LDA and LSVM – were therefore used to validate the efficacy of a feature set in predicting microsleep states and to compare their prediction accuracies. An LDA classifier assumes each class density as multivariate Gaussian with a common covariance-matrix that leads to a linear-decision boundary. An LSVM finds an optimal linear decision boundary that maximally separates the training data into different classes and leads to good test classification performance [47].

To attain a generalized microsleep prediction system, both classifiers were trained on concatenated data from 7 training subjects and tested on the 8th subject via the leave-one-subject-out cross-validation.

An algorithm-based approach referred to as cost-sensitive learning [49] was employed by considering both classes as equally important (or equal misclassification cost) to account for class imbalance ratio in the training data sets. Equal prior probabilities (0.5) were therefore assigned to both classes in the decision.

The prediction of microsleep states was 0.25 s ahead of the gold standard, with a temporal resolution of 0.25 s as shown in Fig. 2.

K. Performance Evaluation

Performance evaluation was based on data from the 8 test subjects. Average values of widely used performance metrics of sensitivity, specificity, precision, Matthew's correlation coefficient (phi), AUC_{PR}, and AUC_{ROC} from 8 independent test subjects are reported here.

Sensitivity, precision, and specificity values alone are biased and can be misleading [50]. These performance metrics were therefore combined into a single widely-used performance metric of phi for imbalanced biomedical data [51], [52] and was subsequently used to demonstrate the effect of feature preprocessing.

TABLE II
AVERAGE NUMBER OF FEATURES SELECTED FROM EACH PROCESSED FEATURE SET

Feature set	Features selected
Corr	69
Cov	69
Wcoh	69
WCSP	64
MI	69
JE	47
PSI	70

The paired non-parametric Wilcoxon signed-rank test [53], was systematically used to compare prediction performances of all feature sets and both classifiers. The threshold-dependent performance metric of phi [52] and threshold-free metrics of AUC_{PR} and AUC_{ROC}, were used in the comparisons. AUC_{PR} is sensitive [54] while AUC_{ROC} is insensitive [44] to the class distributions.

The feature sets were divided into two groups: normalized (correlation, coherence, mutual information, and phase synchronization index) and non-normalized (covariance, cross-spectral power, and joint entropy). The hierarchy of comparisons was (1) inter-group, i.e., normalized feature sets versus their non-normalized counterparts, (2) within-groups, and (3) between classifiers. Phase synchronization index was excluded from the inter-group comparison as it has no non-normalized counterpart.

The effects of class imbalances and prediction times on the phi metric were also assessed for both classifiers and feature sets from the best performing group.

L. Discriminatory Features

The numbers and durations of microsleep events for each subject over the 2 sessions (see Table I) varied considerably. Features selected from training data of concatenated subjects varied considerably in each round of leave-one-subject-out cross-validation. Drawing conclusions from the selected features is inappropriate and can lead to wrong interpretations. A greedy forward step-wise wrapper method was used to find top and consistent discriminatory features across all subjects. Initially, an LDA classifier was trained on each of joint entropy features of concatenated data from 7 subjects, tested on data of test subject, and the corresponding AUC_{ROC}s were recorded. AUC_{ROC}s of 8 iterations were averaged and the feature with the highest AUC_{ROC} was saved. The process was then repeated for each of the remaining features. At each step, the feature resulting in the highest AUC_{ROC} was added to the selected features. This procedure continued until a stopping criterion was met. In our case, a maximum of 10 features was used as the stopping criterion.

III. RESULTS

The number of features selected from each feature set are shown in Table II.

TABLE III
MICROSLEEP PREDICTION (0.25 s) PERFORMANCES (MEAN ± SE) OF UNPROCESSED FEATURE SETS WITH LDA AND LINEAR SVM CLASSIFIERS

	LDA						LSVM					
	Sensitivity	Specificity	Precision	Phi	AUC_{PR}	AUC_{ROC}	Sensitivity	Specificity	Precision	Phi	AUC_{PR}	AUC_{ROC}
Corr	0.65 ±0.02	0.87 ±0.02	0.25 ±0.11	0.28 ±0.07	0.27 ±0.11	0.86 ±0.01	0.69 ±0.11	0.86 ±0.03	0.25 ±0.03	0.30 ±0.08	0.29 ±0.11	0.87 ±0.01
Cov	0.57 ±0.11	0.95 ±0.01	0.36 ±0.12	0.35 ±0.09	0.42 ±0.12	0.93 ±0.02	0.73 ±0.09	0.85 ±0.05	0.30 ±0.12	0.34 ±0.10	0.45 ±0.13	0.93 ±0.01
Wcoh	0.64 ±0.04	0.79 ±0.03	0.18 ±0.08	0.20 ±0.04	0.19 ±0.08	0.81 ±0.01	0.73 ±0.06	0.78 ±0.03	0.19 ±0.08	0.21 ±0.04	0.21 ±0.08	0.85 ±0.02
WCSP	0.56 ±0.02	0.94 ±0.12	0.36 ±0.12	0.38 ±0.08	0.38 ±0.12	0.90 ±0.02	0.73 ±0.06	0.83 ±0.03	0.22 ±0.10	0.27 ±0.07	0.36 ±0.11	0.85 ±0.02
MI	0.61 ±0.04	0.81 ±0.02	0.19 ±0.08	0.19 ±0.04	0.21 ±0.08	0.80 ±0.01	0.65 ±0.05	0.81 ±0.03	0.20 ±0.09	0.21 ±0.04	0.23 ±0.08	0.83 ±0.01
JE	0.70 ±0.09	0.90 ±0.04	0.33 ±0.11	0.39 ±0.09	0.44 ±0.13	0.93 ±0.01	0.70 ±0.12	0.69 ±0.08	0.17 ±0.07	0.21 ±0.07	0.32 ±0.12	0.83 ±0.04
PSI	0.61 ±0.04	0.81 ±0.02	0.20 ±0.09	0.22 ±0.07	0.23 ±0.10	0.80 ±0.02	0.65 ±0.03	0.83 ±0.02	0.22 ±0.10	0.25 ±0.07	0.26 ±0.10	0.83 ±0.02
WSP	0.64 ±0.10	0.86 ±0.06	0.31 ±0.11	0.34 ±0.09	0.41 ±0.12	0.92 ±0.02	0.57 ±0.11	0.79 ±0.05	0.18 ±0.08	0.19 ±0.07	0.23 ±0.09	0.78 ±0.04

TABLE IV
MICROSLEEP PREDICTION (0.25 s) PERFORMANCES (MEAN ± SE) OF PROCESSED FEATURE SETS WITH LDA AND LINEAR SVM CLASSIFIERS

	LDA						LSVM					
	Sensitivity	Specificity	Precision	Phi	AUC_{PR}	AUC_{ROC}	Sensitivity	Specificity	Precision	Phi	AUC_{PR}	AUC_{ROC}
Corr	0.37 ±0.07	0.95 ±0.01	0.29 ±0.12	0.26 ±0.08	0.29 ±0.11	0.83 ±0.04	0.43 ±0.06	0.92 ±0.02	0.26 ±0.11	0.24 ±0.08	0.25 ±0.11	0.81 ±0.04
Cov	0.58 ±0.11	0.96 ±0.01	0.37 ±0.12	0.38 ±0.09	0.40 ±0.12	0.92 ±0.02	0.76 ±0.07	0.91 ±0.03	0.32 ±0.12	0.40 ±0.10	0.42 ±0.13	0.92 ±0.02
Wcoh	0.44 ±0.03	0.87 ±0.03	0.21 ±0.10	0.18 ±0.05	0.19 ±0.09	0.79 ±0.03	0.45 ±0.05	0.86 ±0.03	0.20 ±0.10	0.17 ±0.05	0.19 ±0.09	0.78 ±0.03
WCSP	0.61 ±0.09	0.96 ±0.01	0.43 ±0.11	0.43 ±0.09	0.42 ±0.11	0.93 ±0.01	0.76 ±0.07	0.91 ±0.02	0.33 ±0.11	0.40 ±0.09	0.45 ±0.12	0.93 ±0.01
MI	0.47 ±0.07	0.86 ±0.02	0.20 ±0.10	0.19 ±0.07	0.24 ±0.10	0.78 ±0.04	0.55 ±0.07	0.85 ±0.02	0.22 ±0.10	0.21 ±0.07	0.24 ±0.10	0.80 ±0.03
JE	0.73 ±0.07	0.96 ±0.01	0.42 ±0.13	0.47 ±0.10	0.50 ±0.12	0.95 ±0.02	0.77 ±0.06	0.93 ±0.01	0.37 ±0.13	0.44 ±0.10	0.47 ±0.12	0.94 ±0.02
PSI	0.40 ±0.07	0.89 ±0.02	0.22 ±0.11	0.18 ±0.06	0.20 ±0.10	0.76 ±0.04	0.48 ±0.06	0.88 ±0.02	0.23 ±0.11	0.19 ±0.05	0.21 ±0.10	0.80 ±0.03
WSP	0.56 ±0.11	0.96 ±0.01	0.39 ±0.12	0.37 ±0.09	0.36 ±0.12	0.92 ±0.02	0.74 ±0.08	0.89 ±0.03	0.33 ±0.11	0.39 ±0.09	0.43 ±0.13	0.92 ±0.02

The mean prediction performance metrics for each unprocessed feature set (i.e., without demeaning with respect to the first 2 min baseline) with LDA and LSVM classifiers are presented in Table III. While the mean prediction performance metrics for each preprocessed feature set with both classifiers are presented in Table IV.

In all tables and figures, Cov, Corr, WCPS, Wcoh, JE, MI, PSI, and WSP represent features extracted using covariance, correlation, wavelet cross-spectral power, wavelet coherence, joint entropy, mutual information, phase synchronization index, and wavelet spectral power respectively.

The feature preprocessing with LDA and LSVM classifiers improved the mean phi by (20.5%, 19.4%, 8.6%) and (109.5%, 48.1%, 17.6%) respectively on (joint entropy, wavelet cross-spectral powers, covariance) features. Conversely, feature preprocessing with LDA and LSVM classifiers dropped the mean phi by (7.1%, 10.0%, 18.2%) and (20.0%, 14.2%, 24.0%) respectively on (correlation, wavelet coherence, phase synchronization index). Phi of mutual information features with both classifiers, however, remained unaffected by the feature preprocessing step.

In terms of AUC_{ROC}, AUC_{PR}, and phi, inter-group comparisons for both classifiers showed superior performances of preprocessed non-normalized features to their respective normalized counterparts (all $ps \leq 0.016$). Within-group comparisons, however, showed no significant differences, except for LDA classifier, where joint entropy was superior (all $ps \leq 0.02$) to covariance and correlation was superior (all $p \leq 0.046$) to phase synchronization index. Similarly, between-classifiers comparisons showed no significant

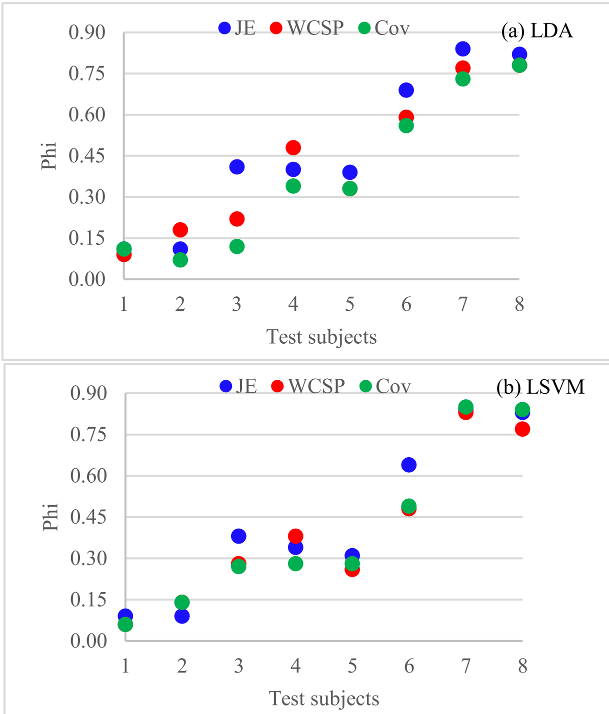

Fig. 3. Phis of non-normalized feature sets for the 8 test subjects, ordered with respect to class imbalance ratios (number of microsleep states vs number of responsive states) of 1:813.40 – 1:2.26

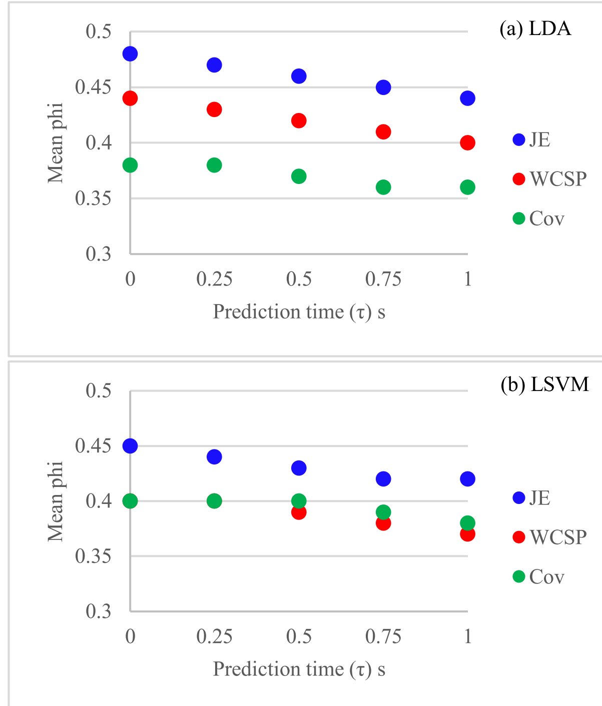

Fig. 4. Average phi of non-normalized feature sets against the prediction time (τ) s.

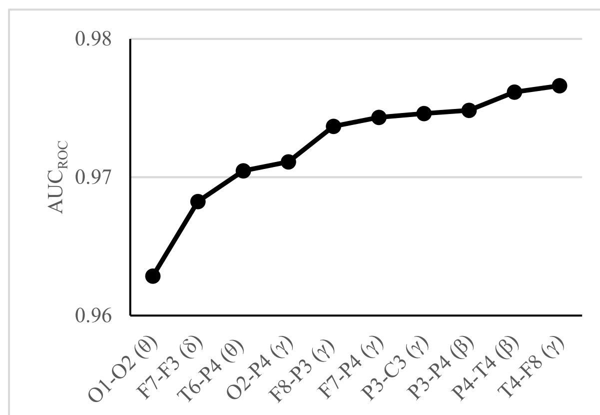

Fig. 5. Top 10 discriminatory features (frequency band) and corresponding AUC$_{ROC}$.

difference, except for joint entropy where LDA had higher phi coefficient compared to LSVM ($p = 0.031$) but had lower phi for mutual information ($p = 0.047$) and lower AUC$_{ROC}$ for phase synchronization index ($p = 0.047$).

The mean prediction (at $\tau = 0.25$ s), AUC$_{ROC}$, AUC$_{PR}$, and phi performances of both preprocessed cross-spectral power and joint entropy features, compared to preprocessed power-spectral features extracted from individual EEG time series, were higher. However, mean prediction performances of covariance features were comparable to power-spectral features.

Overall, joint entropy feature set gave the best performance metrics followed by wavelet cross-spectral power and covariance feature sets.

Fig. 3 shows that, for both classifiers and across the non-normalized preprocessed features, the prediction accuracy phi was anti-correlated ($r = -0.48 - -0.60$) to the imbalance ratio – i.e., phi increases as the imbalance between microsleep and responsive states decreases.

Fig. 4 shows that for LDA and LSVM classifiers, the average rate of performance drop, phi, for (joint entropy, wavelet cross-spectral power, covariance) against the prediction time was (-0.024, -0.040, -0.040) and (-0.020, -0.030, -0.030) respectively. The corresponding correlations between phi and prediction time were (-1.00, -1.00, -0.94) and (-0.97, -0.97, -0.88).

Using leave-one-subject-out cross-validation to evaluate microsleep prediction performances on unforeseen data, each subject was used once as a test subject while the rest were used for the training.

Based on leave-one-subject-out cross-validation, all best discriminatory features were from the theta frequency band. Joint entropies between T6-O2 and P3-C4 were always among the top 10 discriminatory features, whereas, between P3-P4, T6-P4, F7-O2, and T5-P3 6 times. Cross-spectral power between P4-O2 was always in the top 10 discriminatory features, whereas, between F7-F8 and O1-O2 7 times, and between F7-T4 and O2-T4 6 times. Covariances between P4-O2 and P3-O1 were always in the top 10 discriminatory features, whereas, between F7-F8, O1-O2, T6-C3, and T5-O1 7 times.

The top 10 discriminatory/informative features across the study are shown in Fig. 5, where joint entropy between O1-O2 from theta band was the most discriminatory feature.

IV. Discussion

To the best of our knowledge, we are the first to have explored the use of different EEG inter-channel relationships as feature sets of a classifier in the prediction of microsleep states. Furthermore, we have obtained the highest detection and prediction performances seen for microsleep states.

Feature preprocessing (baseline correction/demeaning) led to a substantial improvement in prediction phis for non-normalized feature sets. Except for mutual information, it worsened prediction phis for normalized feature sets.

The standard normalization process requires the mean and variance of the full data and, therefore, is not practical in real-time implementations [6]. The effect of noise (outliers) on the training data is global and the test data normalized with respect to the noisy *training* data also become noisy. In contrast, demeaning the data with respect to the first 2 min of each session is local with respect to that session. Furthermore, with our technique, training and test data are independently preprocessed and similar global means are retained.

Irrespective of the classifier, non-normalized preprocessed feature sets resulted in higher prediction accuracies than spectral features [9], [10], wavelet power-spectral features, and normalized counterparts on performance metrics of phi, AUC_{PR} and AUC_{ROC}. Of these, joint entropy, on average and across both classifiers, outperformed all other feature sets. However, covariance with LSVM showed the lowest rate of performance drop against prediction time (τ). The phi performances of all non-normalized feature sets and across both classifiers had a similar adverse effect with increasing imbalance ratio.

Despite normalized feature sets being used extensively in functional connectivity analyses [17], [18], [55], their poorer prediction performances can be due to the inherent property of being scale(amplitude)-invariant, and consequently, a loss of classification-related information. Furthermore, irrespective of the task, brain regions are likely to be synchronized at times. A change in the level of such synchronization may only be time-locked to some events. Similarly, communication between brain region may be independent of synchronization between them. In such a scenario, brain regions can be considered as cognitive sources and sinks, and the communication can be asynchronous, simplex, or half-duplex. Phase-locking between the two signals may occur even if their amplitudes remain uncorrelated, and noisy signals exhibit random phase slips [56]. Similarly, because of the normalization process, spurious peaks in wavelet coherence can occur for areas of low individual wavelet powers [35]. Furthermore, changes in EEG amplitude and frequency are directly correlated with behavioral performances and circadian rhythms [57], [58]. Conversely, non-normalized feature sets are scale-variant, and consequently, have the better classification-related information.

The anti-correlation between phi and prediction time indicates the presence of microsleep-related information prior to microsleep which lessens as one gets farther ahead of the microsleep.

Both LDA and LSVM classifiers resulted in comparable mean values of AUC_{PR} and AUC_{ROC}. However, LSVM resulted in higher average values of sensitivity than LDA but at the cost of specificity and precision. Peiris *et al.* [6] reported that stacking of 7 LDA classifiers with spectral power features resulted in a similar level of performance to that of an RNN [1]. Likewise, results presented here show that the detection performance (i.e., $\tau = 0$ s in Fig. 4) of a single LDA classifier with joint entropy features is, on average, better than that of 7 stacked ESN with spectral features [8] (i.e., phi of 0.48 vs 0.44). In addition to large datasets [48], linear classifiers have often given similar [59] or better [60], [61] performances on EEG and fMRI-based small data sets, than nonlinear classifiers. This indicates that linear classifiers are more robust and less susceptible to overfitting than nonlinear classifiers. Overall, similar and/or superior performances of LDA classifiers, in these studies support the argument that simple-decision boundaries and estimates via Gaussian models are stable [47] and that this is even more evident in imbalanced datasets.

Selection of best discriminatory features from theta subband is in accordance with the findings that theta activity is associated with drowsiness/microsleeps [1], [2], [4], [6], [21], [62].

Similarly, the top 10 non-normalized features from left and right parietal, left and right frontal, and occipital regions correspond to changes in connectivity during, and prior to the onset of, microsleeps [21] and microsleep-related eye closures [62]–[64].

The mean test AUC_{ROC}(≥ 0.92) of non-normalized feature sets indicate that EEG-based microsleep state predictors can be used in applications that demand equally for sensitivity and specificity.

Our results suggest a way forward in using EEG inter-channel relationships as features in classifying microsleep states. The results boost the importance of considering non-normalized inter-channel relationships over commonly used normalized ones or spectral features. Joint entropy features gave the best mean performances across both classifiers and can be considered as an out-of-the-box choice.

Feature preprocessing with respect to the first 2 min can substantially improve the classification performances of long and imbalanced data.

Notwithstanding our achievements towards prediction of microsleeps, there is still some way to go to realize the accuracy desired for implementation in real-life applications.

References

[1] P. R. Davidson, R. D. Jones, and M. T. R. Peiris, "EEG-based lapse detection with high temporal resolution," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 832–839, May 2007.

[2] M. T. R. Peiris, R. D. Jones, P. R. Davidson, G. J. Carroll, and P. J. Bones, "Frequent lapses of responsiveness during an extended visuomotor tracking task in non-sleep-deprived subjects," *J. Sleep Res.*, vol. 15, pp. 291–300, Sep. 2006.

[3] W. Vanlaar, H. Simpson, D. Mayhew, and R. Robertson, "Fatigued and drowsy driving: A survey of attitudes, opinions and behaviors," *J. Saf. Res.*, vol. 39, pp. 303–309, Jan. 2008.

[4] G. R. Poudel, C. R. H. Innes, P. J. Bones, R. Watts, and R. D. Jones, "Losing the struggle to stay awake: Divergent thalamic and cortical activity during microsleeps," *Hum. Brain Mapping*, vol. 35, pp. 257–269, Jan. 2014.

[5] B. Sirois, U. Trutschel, D. Edwards, D. Sommer, and M. Golz, "Predicting accident probability from frequency of microsleep events," in *Proc. World Congr. Med. Phys. Biomed. Eng.*, vol. 25, Jan. 2010, pp. 2284–2286.

[6] M. T. R. Peiris, P. R. Davidson, P. J. Bones, and R. D. Jones, "Detection of lapses in responsiveness from the EEG," *J. Neural Eng.*, vol. 8, p. 016003, Feb. 2011.

[7] M. Golz, D. Sommer, M. Chen, D. Mandic, and U. Trutschel, "Feature fusion for the detection of microsleep events," *J. VLSI Signal Process. Syst. Signal Image Video Technol.*, vol. 49, no. 2, pp. 329–342, 2007.

[8] S. Ayyagari, "Reservoir computing approaches to EEG-based detection of microsleeps," Ph.D. dissertation, Dept. Elect. Comput. Eng., Univ. Canterbury, Christchurch, U.K., 2017.

[9] R. Shoorangiz, S. J. Weddell, and R. D. Jones, "Prediction of microsleeps from EEG: Preliminary results," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2016, pp. 4650–4653.

[10] R. Shoorangiz, S. J. Weddell, and R. D. Jones, "Bayesian multi-subject factor analysis to predict microsleeps from EEG power spectral features," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2017, pp. 4183–4186.

[11] M. Golz, D. Sommer, and J. Krajewski, "Prediction of immediately occurring microsleep events from brain electric signals," *Current Directions Biomed. Eng.*, vol. 2, no. 1, pp. 149–153, 2016.

[12] C.-T. Lin, K.-C. Huang, C.-H. Chuang, L.-W. Ko, and T.-P. Jung, "Can arousing feedback rectify lapses in driving? Prediction from EEG power spectra," *J. Neural Eng.*, vol. 10, no. 5, p. 056024, Oct. 2013.

[13] K. J. Friston, "Functional and effective connectivity: A review," *Brain Connectivity*, vol. 1, no. 1, pp. 13–36, 2011.

[14] K. J. Friston, A. M. Bastos, A. Oswal, B. van Wijk, C. Richter, and V. Litvak, "Granger causality revisited," *NeuroImage*, vol. 101, pp. 796–808, Nov. 2014.

[15] Y. T. Qassim, T. R. H. Cutmore, D. A. James, and D. D. Rowlands, "Wavelet coherence of EEG signals for a visual oddball task," *Comput. Biol. Med.*, vol. 43, pp. 23–31, Jan. 2013.

[16] J. S. Minna, H. Hintsala, J. Kortelainen, and T. Seppänen, "Experimental comparison of connectivity measures with simulated EEG signals," *Med. Biol. Eng. Comput.*, vol. 50, pp. 683–688, Jul. 2012.

[17] R. Q. Quiroga, A. Kraskov, T. Kreuz, and P. Grassberger, "Performance of different synchronization measures in real data: A case study on electroencephalographic signals," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 65, p. 041903, Apr. 2002.

[18] L. Astolfi *et al.*, "Estimation of effective and functional cortical connectivity from neuroelectric and hemodynamic recordings," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 17, no. 3, pp. 224–233, Jun. 2009.

[19] C. Dissanayaka *et al.*, "Comparison between human awake, meditation and drowsiness EEG activities based on directed transfer function and MVDR coherence methods," *Med. Biol. Eng. Comput.*, vol. 53, pp. 599–607, Jul. 2015.

[20] M. Awais, N. Badruddin, and M. Drieberg, "EEG brain connectivity analysis to detect driver drowsiness using coherence," in *Proc. Int. Conf. Frontiers Inf. Technol.*, Dec. 2017, pp. 110–114.

[21] J. Toppi, L. Astolfi, G. R. Poudel, C. R. H. Innes, F. Babiloni, and R. D. Jones, "Time-varying effective connectivity of the cortical neuroelectric activity associated with behavioural microsleeps," *NeuroImage*, vol. 124, pp. 421–432, Jan. 2015.

[22] Y.-Y. Lee and S. Hsieh, "Classifying different emotional states by means of EEG-based functional connectivity patterns," *PLoS ONE*, vol. 9, p. e95415, Apr. 2014.

[23] V. Bono, D. Biswas, S. Das, and K. Maharatna, "Classifying human emotional states using wireless EEG based ERP and functional connectivity measures," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Inform.*, Feb. 2016, pp. 200–203.

[24] G. N. Dimitrakopoulos *et al.*, "Functional connectivity analysis of mental fatigue reveals different network topological alterations between driving and vigilance tasks," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 4, pp. 740–749, Apr. 2018.

[25] A. Baseer, S. J. Weddell, and R. D. Jones, "Prediction of microsleeps using pairwise joint entropy and mutual information between EEG channels," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Jul. 2017, pp. 4495–4498.

[26] H. E. Wang, C. G. Benar, P. P. Quilichini, K. J. Friston, V. K. Jirsa, and C. Bernard, "A systematic framework for functional connectivity measures," *Frontiers Neurosci.*, vol. 8, p. 405, Dec. 2014.

[27] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls," *Frontiers Syst. Neurosci.*, vol. 9, p. 175, Jan. 2016.

[28] T. R. Mullen *et al.*, "Real-time neuroimaging and cognitive monitoring using wearable dry EEG," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 11, pp. 2553–2567, Nov. 2015.

[29] W. De Clercq, A. Vergult, B. Vanrumste, W. Van Paesschen, and S. Van Huffel, "Canonical correlation analysis applied to remove muscle artifacts from the electroencephalogram," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 12, pp. 2583–2587, Nov. 2006.

[30] A. Klein, T. Sauer, A. Jedynak, and W. Skrandies, "Conventional and wavelet coherence applied to sensory-evoked electrical brain activity," *IEEE Trans. Biomed. Eng.*, vol. 53, no. 2, pp. 266–272, Feb. 2006.

[31] J. S. Brittain, D. M. Halliday, B. A. Conway, and J. B. Nielsen, "Single-trial multiwavelet coherence in application to neurophysiological time series," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 5, pp. 854–862, May 2007.

[32] M. I. Plett, "Transient detection with cross wavelet transforms and wavelet coherence," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1605–1611, May 2007.

[33] C. Torrence and P. J. Webster, "Interdecadal changes in the ENSO–Monsoon system," *J. Climate*, vol. 12, pp. 2679–2690, Aug. 1999.

[34] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlinear Process. Geophys.*, vol. 11, nos. 5–6, pp. 561–566, 2004.

[35] D. Maraun and J. Kurths, "Cross wavelet analysis: Significance testing and pitfalls," *Nonlinear Process. Geophys.*, vol. 11, no. 4, pp. 505–514, 2004.

[36] S. Khan *et al.*, "Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 76, no. 2, p. 026209, 2007.

[37] H. Singh, N. Misra, V. Hnizdo, A. Fedorowicz, and E. Demchuk, "Nearest neighbor estimates of entropy," *Amer. J. Math. Manage. Sci.*, vol. 23, nos. 3–4, pp. 301–321, 2003.

[38] A. Kraskov, H. Stögbauer, and P. Grassberger, "Estimating mutual information," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 69, p. 066138, Jun. 2004.

[39] S. Szabó, "Information theoretical estimators toolbox," *J. Mach. Learn. Res.*, vol. 15, pp. 283–287, Jan. 2014.

[40] S. L. Marple, Jr., "Computing the discrete-time 'analytic' signal via FFT," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2600–2603, Sep. 1999.

[41] J. Sun, X. Hong, and S. Tong, "Phase synchronization analysis of EEG signals: An evaluation based on surrogate tests," *IEEE Trans. Biomed. Eng.*, vol. 59, no. 8, pp. 2254–2263, Aug. 2012.

[42] R. May, G. Dandy, and H. Maier, "Review of input variable selection methods for artificial neural networks," in *Artificial Neural Networks, Methodological Advances and Biomedical Applications*. Rijeka, Croatia: InTech, 2011.

[43] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.

[44] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Dec. 2005.

[45] O. Chapelle, "Training a support vector machine in the primal," *Neural Comput.*, vol. 19, no. 5, pp. 1155–1178, 2007.

[46] N. D. Lawrence and B. Schölkopf, "Estimating a kernel Fisher discriminant in the presence of label noise," in *Proc. ICML*, 2001, pp. 306–313.

[47] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York, NY, USA: Springer, 2008.

[48] G.-X. Yuan, C.-J. Ho, and C.-J. Lin, "Recent advances of large-scale linear classification," *Proc. IEEE*, vol. 100, no. 9, pp. 2584–2603, Sep. 2012.

[49] W.-J. Lin and J. J. Chen, "Class-imbalanced classifiers for high-dimensional data," *Briefings Bioinf.*, vol. 14, pp. 13–26, Jan. 2013.

[50] D. M. Powers, "Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation," *J. Mach. Learn. Technol.*, vol. 2, no. 1, pp. 37–63, 2011.

[51] S. Boughorbel, F. Jarray, and M. El-Anbari, "Optimal classifier for imbalanced data using Matthews correlation coefficient metric," *PLoS ONE*, vol. 12, p. e0177678, Jun. 2017.

[52] C. Parker, "An analysis of performance measures for binary classifiers," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2011, pp. 517–526.

[53] J. D. Gibbons and S. Chakraborti, *Nonparametric Statistical Inference*, 4th ed. New York, NY, USA: Marcel Dekker, 2003.

[54] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, p. e0118432, Mar. 2015.
[55] J. D. Bonita *et al.*, "Time domain measures of inter-channel EEG correlations: A comparison of linear, nonparametric and nonlinear measures," *Cognit. Neurodyn.*, vol. 8, no. 1, pp. 1–15, 2014.
[56] E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Prog. Neurobiol.*, vol. 77, pp. 1–37, Sep./Oct. 2005.
[57] U. Melia *et al.*, "Mutual information measures applied to EEG signals for sleepiness characterization," *Med. Eng. Phys.*, vol. 37, pp. 297–308, Mar. 2015.
[58] D. La Rocca *et al.*, "Human brain distinctiveness based on EEG spectral coherence connectivity," *IEEE Trans. Biomed. Eng.*, vol. 61, no. 9, pp. 2406–2412, Sep. 2014.
[59] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 11, no. 2, pp. 141–144, Jun. 2003.
[60] M. Misaki, Y. Kim, P. A. Bandettini, and N. Kriegeskorte, "Comparison of multivariate classifiers and response normalizations for pattern-information fMRI," *NeuroImage*, vol. 53, pp. 103–118, Oct. 2010.
[61] H. Jian-feng, "Comparison of different classifiers for biometric system based on EEG signals," in *Proc. Int. Conf. Inf. Technol. Comput. Sci.*, Jul. 2010, pp. 288–291.
[62] Y. Jonmohamadi, G. R. Poudel, C. C. R. H. Innes, and R. D. Jones, "Microsleeps are associated with stage-2 sleep spindles from hippocampal-temporal network," *Int. J. Neural Syst.*, vol. 26, p. 1650015, Jun. 2016.
[63] J. L. Ong, D. Kong, T. T. Y. Chia, J. Tandi, B. T. T. Yeo, and M. W. L. Chee, "Co-activated yet disconnected—Neural correlates of eye closures when trying to stay awake," *NeuroImage*, vol. 118, pp. 553–562, Sep. 2015.
[64] G. R. Poudel, R. D. Jones, C. R. H. Innes, R. Watts, P. R. Davidson, and P. J. Bones, "Measurement of BOLD changes due to cued eye-closure and stopping during a continuous visuomotor task via model-based and model-free approaches," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 18, no. 5, pp. 479–488, Oct. 2010.