Resting-state EEG functional connectivity predicts post-traumatic stress disorder subtypes in veterans
Journal of Neural Engineering
Li, Qianliang; Theodorsen, Maya Coulson; Konvalinka, Ivana; Eskelund, Kasper; Karstoft, Karen-Inge; Andersen, Søren Bo; Andersen, Tobias S.

Published in:
Journal of Neural Engineering

DOI:
10.1088/1741-2552/ac9aaf

Publication date:
2022

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
Li, Q., Theodorsen, M. C., Konvalinka, I., Eskelund, K., Karstoft, K-I., Andersen, S. B., & Andersen, T. S. (2022). Resting-state EEG functional connectivity predicts post-traumatic stress disorder subtypes in veterans: Journal of Neural Engineering. Journal of Neural Engineering, 19, [066005]. https://doi.org/10.1088/1741-2552/ac9aaf
Resting-state EEG functional connectivity predicts post-traumatic stress disorder subtypes in veterans

To cite this article: Qianliang Li et al 2022 J. Neural Eng. 19 066005

View the article online for updates and enhancements.

You may also like

- Assessing the effectiveness and pathways of planned shelters in protecting mental health of flood victims in China
  Shuang Zhong, Minghui Pang, Hung Chak Ho et al.

- A review of physiological and behavioral monitoring with digital sensors for neuropsychiatric illnesses
  Erik Reinertsen and Gari D Clifford

- The synchronous neural interactions test as a functional neuromarker for post-traumatic stress disorder (PTSD): a robust classification method based on the bootstrap
  A P Georgopoulos, H-R M Tan, S M Lewis et al.
Resting-state EEG functional connectivity predicts post-traumatic stress disorder subtypes in veterans

Qianliang Li, Maya Coulson Theodorsen, Ivana Konvalinka, Kasper Eskelund, Karen-Inge Karstoft, Søren Bo Andersen, and Tobias S Andersen

1 Section for Cognitive Systems, DTU Compute, Technical University of Denmark, Kongens Lyngby, Denmark
2 Department of Military Psychology, Danish Veteran Centre, Danish Defence, Copenhagen, Denmark
3 Research and Knowledge Centre, Danish Veteran Centre, Danish Defence, Ringsted, Denmark
4 Department of Psychology, University of Copenhagen, Copenhagen, Denmark

* Author to whom any correspondence should be addressed.
E-mail: glia@dtu.dk

Keywords: functional connectivity, machine learning, PTSD, resting-state EEG, subtypes

Abstract

Objective. Post-traumatic stress disorder (PTSD) is highly heterogeneous, and identification of quantifiable biomarkers that could pave the way for targeted treatment remains a challenge. Most previous electroencephalography (EEG) studies on PTSD have been limited to specific handpicked features, and their findings have been highly variable and inconsistent. Therefore, to disentangle the role of promising EEG biomarkers, we developed a machine learning framework to investigate a wide range of commonly used EEG biomarkers in order to identify which features or combinations of features are capable of characterizing PTSD and potential subtypes.

Approach. We recorded 5 min of eyes-closed and 5 min of eyes-open resting-state EEG from 202 combat-exposed veterans (53% with probable PTSD and 47% combat-exposed controls). Multiple spectral, temporal, and connectivity features were computed and logistic regression, random forest, and support vector machines with feature selection methods were employed to classify PTSD. To obtain robust results, we performed repeated two-layer cross-validation to test on an entirely unseen test set.

Main results. Our classifiers obtained a balanced test accuracy of up to 62.9% for predicting PTSD patients. In addition, we identified two subtypes within PTSD: one where EEG patterns were similar to those of the combat-exposed controls, and another that were characterized by increased global functional connectivity. Our classifier obtained a balanced test accuracy of 79.4% when classifying this PTSD subtype from controls, a clear improvement compared to predicting the whole PTSD group. Interestingly, alpha connectivity in the dorsal and ventral attention network was particularly important for the prediction, and these connections were positively correlated with arousal symptom scores, a central symptom cluster of PTSD.

Significance. Taken together, the novel framework presented here demonstrates how unsupervised subtyping can delineate heterogeneity and improve machine learning prediction of PTSD, and may pave the way for better identification of quantifiable biomarkers.

1. Introduction

Post-traumatic stress disorder (PTSD) is a debilitating psychiatric disorder characterized by chronic stress symptoms following exposure to trauma. Specifically, four different clusters of symptoms have to be present for the clinical diagnosis: (a) intrusion and re-experiencing of the traumatic event, (b) avoidance symptoms, (c) negative alterations of mood and cognition and (d) alterations in arousal and reactivity (American Psychiatric Association 2013, Fenster et al 2018). There are multiple symptoms within each cluster, and thus, a plethora of different combinations can fulfil the diagnostic criteria for PTSD, resulting in PTSD being a very heterogeneous disorder (Galatzer-Levy and Bryant 2013). The heterogeneous
nature might partly explain the moderate response rate of trauma-focused psychotherapies, which are the first-line treatment option for PTSD. Indeed, a meta-analysis of psychotherapeutic treatment of combat-related PTSD found that approximately two-thirds of patients retain their diagnosis following psychotherapy (Steenkamp et al. 2015). The same is true for pharmacological treatments of PTSD: while selective serotonin reuptake inhibitors have proven superior to placebo in reduction of PTSD-symptoms, effect sizes are generally small and limited to symptom relief in certain patients (DePierro et al. 2019, Hoskins et al. 2021). Overall, the marked heterogeneity of PTSD calls for identification of clinical subtypes and biomarkers that can in turn pave the way for personalized treatment of PTSD.

Previous research has investigated whether changes in brain activity in patients with PTSD, measured using electroencephalography (EEG), could serve as potential clinical biomarkers (Butt et al. 2019). Discovery of new EEG biomarkers has the potential to improve diagnosis of PTSD and open new avenues of neurophysiological targets for EEG neurofeedback and pharmacological treatment of PTSD (Woo et al. 2017). Additionally, EEG has a high temporal resolution and good clinical practicality, due to being non-invasive, low cost, and mobile. However, EEG studies of PTSD have revealed conflicting findings. For example, some studies focusing on the alpha frequency band have reported lower alpha power in PTSD patients (Jokić-Begić and Begić 2003, Veltmeyer et al. 2006, Clancy et al. 2017), while other studies have found no differences in alpha power between healthy controls and PTSD patients (Begić et al. 2001, Ehlers et al. 2006, Rabe et al. 2006, Shankman et al. 2008, Kemp et al. 2010, Wahbeh and Oken 2013, Imperatori et al. 2014, Moon et al. 2018). Albeit less investigated, the same inconsistencies and opposite results can also be found in the other canonical frequency bands (Todder et al. 2012, see supplementary table A1 for an overview). Again, such conflicting results might pertain to the heterogeneous nature of PTSD, and call for the identification of biologically founded subtypes (Zhang et al. 2021).

Besides spectral power, other types of EEG features have also been investigated in regards to PTSD. Long-range temporal correlations (Linkenkaer-Hansen et al. 2001) were found to be lower in the alpha band in PTSD patients, but returned to a normal level following neurofeedback (Ros et al. 2017). A study measuring simultaneous EEG and fMRI observed a microstate associated with the dorsal default mode network to be more present in combat-related PTSD compared to combat-exposed controls, and its occurrence was correlated positively with the PTSD checklist scores (PCL; Weathers et al. 1993). Additionally, lower occurrence of a microstate associated with the anterior salience network and higher occurrence of a microstate associated with the posterior salience network was observed in the PTSD group (Yuan et al. 2018). A recent study also utilized microstate characteristics in a machine learning model to predict PTSD (Terpou et al. 2022). Another EEG feature that has been investigated in relation to PTSD is functional connectivity. Reduced frontal to posterior right hemispheric alpha Granger causality (GC) (Clancy et al. 2017) and reduced theta orthogonalized power envelope correlations (PECs) (Toll et al. 2020) have been observed in PTSD. Machine learning modelling using coherence has also been applied to EEG data from PTSD patients to predict the effect of transcranial magnetic stimulation (Zandvakili et al. 2019). However, common for all these non-spectral features is that each feature type has only been investigated by one or a limited number of studies in relation to PTSD, thus no clear conclusions can be drawn, as these results have not yet been reproduced. Additionally, most of the EEG studies examining PTSD have only focused on one particular feature type. To our knowledge, only one study has explored combinations of different resting-state EEG features, namely spectral power, spatial covariance, and network metrics to predict PTSD (Kim et al. 2020).

In this study, we present an extensive framework (figure 1) that computes multiple EEG features that are commonly used in EEG research, which can broadly be divided into spectral features (power, asymmetry, frontal theta/beta ratio, peak alpha frequency, and $1/f$ exponent), functional connectivity features (imaginary part of coherence (Imcoh), weighted phase lag index (wPLI), PEC, and GC), and features that capture the temporal dynamics of EEG (microstates and long-range temporal correlations estimated through detrended fluctuation analysis (DFA) exponents). Feature selection and machine learning classification were applied to investigate how well well veterans with probable PTSD (defined by a score of 44 or above on the PCL; Karstoft et al. 2014) can be distinguished from combat-exposed controls. Furthermore, we examined which features were important for the classifications. Stratified repeated two-layer cross-validation (CV) was employed for training, optimizing and testing the machine learning models (figure 2). Additionally, we also applied unsupervised sparse clustering (Witten and Tibshirani 2010) to the PTSD group to discern EEG-driven subtypes in an attempt to delineate biological heterogeneity. In line with the Research Domain Criteria (RDoC), the participants in this study were recruited broadly from a trauma-affected population to address the heterogeneity within PTSD and allow for identification of relevant subtypes of trauma-related psychopathology (Insel et al. 2010, Neria 2021).
Figure 1. Overview of the EEG analysis framework. After EEG acquisition, the data were preprocessed, and commonly used resting-state EEG features were computed. Repeated 10-by-10 two-layer cross-validation was performed to split the data. Feature selection, model training and hyperparameter tuning were performed on the inner fold training and validation set, while the generalization performance was evaluated on the entirely unseen test set. Specifically, we employed a minimal-redundancy-maximum-relevance filtering algorithm, followed by training of a linear support vector machine with recursive feature elimination, logistic regression with sequential forward selection, or random forest to classify PTSD. Multiple classifiers were used to ensure our results would be less dependent on model choice.

Figure 2. Schematic of the repeated two-layer cross-validation scheme. In order to obtain robust and reliable estimations of how well the classifiers would perform on unseen data, we employed ten repetitions of 10-by-10 two-layer cross-validation. The outer fold split ensured we tested on unseen data, while the inner fold split alleviates overfitting when we optimized model hyperparameters, i.e. regularization strength for logistic regression and support vector machine, number of trees and depth for random forest, and the number of features used by the models after feature selection.

2. Materials and methods

2.1. Participants
Participants for this study were pooled from three different studies. All participants were combat-exposed veterans and predominantly male (96% males). Inclusion criteria were broad in that everyone who was part of study 1 or presented at the Military Psychology Department at the Danish Defense (MPD; studies 2 and 3) was eligible for inclusion. For each study, specific inclusion criteria apply.
Study 1 is a case-control study among previously deployed soldiers from the same deployment cohort. Probable PTSD cases at 2.5 years after deployment (see Andersen et al. 2014 for overall study details) were invited to participate in an EEG-study 6.5 years after deployment, along with a control group (no PTSD) matched by age and gender. In total, 52 soldiers were included (26 individuals with probable PTSD at 2.5 years and 26 controls). At 6.5 years, 16 individuals were still considered PTSD cases (defined as PCL-score \( \geq 44 \); Karstoft et al. 2014), while 36 were not. Their mean age was 35.42 (SD = 8.24).

Study 2 is a cross-sectional study of treatment-seeking veterans aimed at studying the relation between trauma-related anhedonia and reward behaviour. Individuals referred to treatment at the MPD were included and all data were collected before treatment. To achieve a balanced sample, individuals were selected based on symptoms of anhedonia on the PCL. In total, 77 participants were included, of which 44 were probable PTSD cases (defined as PCL-score \( \geq 44 \)) while 33 were not. The mean age was 39.83 (SD = 9.53).

Study 3 is a cohort study of treatment-seeking veterans who are assessed over time from before to after treatment. For this study, we included data from the first assessment, i.e. at treatment intake. Inclusion criteria were broad and included everyone referred to treatment at the MPD. From this pool, participants were randomly selected. Overall, 61 were considered probable PTSD cases (defined as PCL-score \( \geq 44 \)) while 33 were not. The mean age was 35.11 (SD = 8.59).

Participants across all three studies gave their informed consent and the studies were approved by the Danish Data Protection Agency.

2.2. Clinical measures

The Post-Traumatic Stress Disorder Checklist-Civilian version (PCL-C; Weathers et al. 1993) is a 17-item self-report measure developed to capture the PTSD symptoms as described in the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV; American Psychiatric Association 1994). Each item is scored from 1 (not at all) to 5 (extremely). A total PCL-score is computed as the sum of the scores from all the items, ranging from 17 to 85. A criteria of total PCL-score \( \geq 44 \) (Karstoft et al. 2014) was used to define PTSD cases for the machine learning models. We also computed the sub-symptom severity scores, where the intrusion sub-symptom score was computed as the sum of items 1–5, avoidance score was the sum of items 6–12 and the arousal severity score was the sum of items 13–17 (Weathers et al. 1993).

2.3. EEG acquisition

For study 1, continuous EEG was recorded from the scalp at 22 sites (Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T7, T8, P7, P8, Fz, Cz, Pz, AFz, CPz, POz) using an EasyCap cap (EASYCAP GmbH, Woerthsee-Etterschlag, Germany) with passive sintered Ag/AgCl electrodes and an mBrainTrain SMARTING amplifier system (mBrainTrain, Beograd, Serbia). EEG was sampled at 500 Hz and 24 bit depth during the recording, with a reference to average mastoids. Impedance across all electrodes was kept below 10 kΩ.

For study 2, continuous EEG was recorded from the scalp at 31 sites (Fp1, Fpz, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FC4, FT8, FCz, T3, C3, Cz, C4, T4, CPz, TP7, CP3, CP4, TP8, T5, P3, Pz, P4, T6, O1, Oz, O2) using an ANT Waveguard cap (ANT Neuro, Hengelo, Netherlands) with passive sintered Ag/AgCl electrodes and a MITSAR EEG-202 amplifier system (Mitsar, St. Petersburg, Russia). EEG was sampled at 500 Hz and 24 bit depth with a 0.1–150 Hz bandpass filter during the recording, and a reference to linked ears. Impedance across all electrodes was kept below 10 kΩ.

For study 3, continuous EEG was recorded from the scalp at 19 sites (Fp1, Fp2, F7, F3, Fz, F4, F8, T7, C3, Cz, C4, T8, P7, P3, Pz, P4, F8, O1, O2) using an ECI cap (Electro-Cap International Inc., Eaton, Ohio, USA) with passive tin electrodes and a MITSAR EEG-201 amplifier system. EEG was sampled at 250 Hz and 16 bit depth with a 0.1–150 Hz bandpass filter during the recording, and a reference to linked ears. Impedance across all electrodes was below 10 kΩ.

For all three studies, the recordings were conducted in the following manner: participants were seated in a comfortable chair with arm and neck rests, facing a computer display. A brief breathing instruction was given to ensure that participants were sufficiently at rest. EEG was recorded in ten interleaved intervals of eyes-open and eyes-closed conditions, with each interval lasting 70 s. The first 10 s were universally rejected, leaving 60 s in each interval, resulting in 10 min of EEG data for each participant in total. During the recording, the experimenter was present, but the instructions to the participant were given via a PC screen. During eyes-open intervals, participants were required to fixate at a fixation cross at the centre of the computer display. During eyes-closed intervals, participants were required to close and open their eyes at the sounding of a signal tone. The progress of the recording intervals was controlled by the participants, being prompted by messages on the computer display.

2.4. EEG preprocessing

The EEG data were processed using MNE-python 0.22.1 (Gramfort et al. 2013). First, the data were band-pass filtered at 1–100 Hz, and a notch filter was applied at 50 Hz to remove power-line noise. After filtering, the data were divided in 4 s epochs without overlap. Bad epochs and channels with gross artefacts, with the exception of ocular and ECG artefacts, were rejected by visual inspection. The data
were re-referenced to the common average without using the bad channels in order to obtain a robust average (Bigdely-Shamlo et al 2015). The bad channels were then interpolated from adjacent channels using spherical spline interpolation (Perrin et al 1989). Ocular and ECG artefacts were removed using independent component analysis (ICA) based on FastICA, with number of components set to the number of non-interpolated channels (Bell and Sejnowski 1995, Makeig et al 1995). Remaining artefacts were removed by marking potential bad epochs using a variable peak-to-peak voltage threshold determined by Autoreject 0.2.1 (Jas et al 2017), followed by visual inspection of the marked bad epochs for the final confirmation of rejection. Participants with more than 20% bad epochs during all of the pre-processing steps were excluded from further analysis, resulting in a total sample size of 107 PTSD veterans and 95 combat-exposed controls. Finally, the EEG data were down-sampled to 200 Hz and filtered into five canonical frequency bands: delta (1.25–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–49 Hz), unless otherwise stated. Due to different electrode configurations from the three studies, the EEG data from each study were up-sampled using spherical spline interpolation to match a common electrode configuration corresponding to the union of the channels across all three studies.

2.5. EEG feature estimation

A brief description of each EEG feature type is provided, as all the estimated features are already well-established (more details can be found in appendix B). All features were separately estimated for each eye condition.

2.5.1. Spectral analysis

EEG power was computed using multitaper spectral estimation (Babadi and Brown 2014) and we estimated both absolute and relative power. EEG power asymmetry (Sutton and Davidson 1997) was estimated in frontal, central and posterior regions. Theta/beta ratio was estimated in the frontal region (Angelidis et al 2018). Peak alpha frequency and 1/f exponent were estimated using the Fitting Oscillations & One Over F algorithm (FOOOF; Donoghue et al 2020). Peak alpha frequency and 1/f exponents were calculated for each channel, and a global peak alpha frequency was also computed as the average over all channels.

2.5.2. Long-range temporal correlations

Long-range temporal correlations were computed using DFA according to the procedure described by Hardstone et al (2012). Nolds 0.5.2 (Scholzel 2019) was used to implement DFA. The DFA exponent was calculated for each channel and a global DFA exponent was also computed as the average over all channels.

2.5.3. Microstates

The modified K-means algorithm was used to find the microstates (Pascual-Marqui et al 1995). Based on global explained variance and the CV criterion (Pascual-Marqui et al 1995), we determined four microstates were the most appropriate for our data. Ratio of time covered, microstate entropy, and transition matrices were computed for the microstates. The python implementation by von Wegner and Laufs (2018) was used for microstate analysis.

2.5.4. Source functional connectivity

We adapted the source localization methodology from Zhang et al (2021), who also investigated source functional connectivity in regards to PTSD. Briefly, minimum-norm estimation was performed to transform our sensor space EEG to source time series of 20 484 vertices, followed by extraction of time series from 31 cortical regions of interest (ROIs). The Destrieux Atlas (Destrieux et al 2010) parcellation was manually modified (supplementary figure B1) to approximate the 31 ROIs used by Zhang et al (2021) (described in Toll et al 2020). The FreeSurfer average brain template from FreeSurfer 6 (Fischl 2012) was used to construct the boundary element head model and forward operator for the source modelling. Additionally, free orientations were used for the dipoles and the regularization parameter for the minimum-norm estimation was set to $\lambda^2 = 1/81$. Principal component analysis was performed to reduce the three-dimensional source signals to one-dimensional by projecting onto the dominant principal direction. Following estimation of source time series, the various functional connectivity measurements were calculated.

Coherence is calculated as the magnitude of the cross spectrum between two signals divided with the square root of the product of each signal’s power spectrum for normalization (Nunez et al 1997). By using only the imaginary part of coherence (Imcoh), the effect of volume conduction can be mitigated (Nolte et al 2004).

Phase lag index estimates how consistent one signal leads/lags behind the other signal. Similar to the idea behind Imcoh, the wPLI was developed to further attenuate the effect of volume conduction by weighing each phase difference with the magnitude of the lag (Vinch et al 2011, Hardmeier et al 2014).

PECs were estimated following Toll et al (2020). Briefly, the time series were bandpass filtered, Hilbert transformed, orthogonalized to each other (Hipp et al 2012), squared, and log-transformed. Finally, Pearson’s correlations were calculated and Fischer’s $r$-to-$z$ transform was applied.

GC were estimated following Ding et al (2006). GC estimation using autoregressive models assumes the data is stationary, and this assumption was confirmed using Augmented Dicker-Fuller test. Akaike information criterion (AIC) and Bayesian
information criterion (BIC) were computed to estimate the model order, but BIC failed to converge at a minimum, thus only the AIC was used and a model order of 5 was chosen. The nite 0.8.1 (NIPY 2019) python library was used for implementation of GC.

2.6. Cross-validation
A group stratified nested two-layer CV scheme with ten-fold outer CV and ten-fold inner CV (figure 2) was employed to train and evaluate the performance of our machine learning models. The group stratification took into account the differences in class imbalances for the three studies, i.e. the percentage of patients with PTSD from each study was kept consistent across folds. All hyperparameters, e.g. regularization, number of features and number of trees were determined in the inner training fold, while the generalization performance was evaluated on the outer fold test set. Additionally, to obtain more robust performance estimates, we repeated the two-layer CV ten times, to take into account the effect of random splits, resulting in 100 hyperparameter sets, one for each final model tested on the test data (ten repetitions of ten-fold outer CV).

2.7. Machine learning
All the EEG features were combined and turned into a data matrix consisting of number of participants by 24 458 features. We had many more features than samples, and in order to decrease the dimensionality and reduce overfitting (Altman and Krzywinski 2018), we applied multiple dimensionality reduction methods. All feature selection steps were applied on the inner CV fold only. First, minimal-redundancy-maximum-relevance (mRMR; Peng et al 2005) was applied to each EEG feature type separately, to attenuate the effect of imbalance in the number of features between each EEG feature type. Specifically, it primarily attenuates bias towards the connectivity feature types, which consist of many features due to being computed for all pair-wise combinations. Following the first round of mRMR, the number of features for each EEG feature type have been approximately equalized and thus a second round of mRMR can be applied with all the feature types pooled, to further reduce the dimensionality. In some of the models, we tested the performance of each feature type separately, and for those models we only used one round of mRMR. Following mRMR, three commonly employed classifiers were trained to predict PTSD (defined as PCL-score \( \geq 44 \)): (a) linear support vector machine (SVM) with recursive feature elimination (Guyon et al 2002), (b) logistic regression with sequential forward selection, and (c) random forest (Breiman 2001). Multiple classifiers were used to ensure our results would be less dependent on model choice. We primarily used scikit-learn 0.24.1 (Pedregosa et al 2011) for implementation of the machine learning algorithms, mlxtend 0.17.0 (Raschka 2018) for sequential forward selection and a python implementation of mRMR (Kiran 2017).

All model hyperparameters were tuned in the inner CV fold on the validation set with grid search. Specifically, we tested the performance of using 20, 30, 40, 50 or 60 features in mRMR, nine values on a logarithmic scale from 0.001 to 1 for the regularization parameter \( C \) for SVM and logistic regression, and 10, 100, 500 or 1000 trees with depths 1 or 2 in random forest. Balanced accuracy, computed as the mean of sensitivity and specificity, was used as the scoring metric to account for class imbalances. The range of the hyperparameters were determined during preliminary analysis of when the models started to overfit and the step sizes were designed with the computational time in mind.

To determine which of the selected features were important for classifications, we evaluated the selected features in our classifiers based on how often they were selected in the different data splits. Features selected with a high frequency across CV runs consistently provide predictive information for classification, while features selected with a low frequency might be selected due to fitting to noise specific for those particular CV data splits. For visual purposes, we chose a threshold of presence in at least 20% of all CV runs, to not visualize an overwhelming amount of features and to better focus on the most important.

2.8. Subtyping
We adapted the clustering methodology from Zhang et al (2021), who also investigated PTSD subtypes, but expanded the clustering to use all our EEG features. Briefly, sparse \( K \)-means clustering performs joint feature selection and clustering by employing a lasso-type sparsity constraint \( s \). The gap statistic was used to determine number of clusters and the sparsity constraint (Tibshirani et al 2001, Witten and Tibshirani 2010). Specifically, we employed a grid-search approach where we estimated the gap statistic for all combinations of \( K \) between 2 and 6 clusters and 20 values for \( s \), ranging from 1.2 (few features) to 141 (all features), equally spaced on a logarithmic scale. For each \( s \) value, we found the best \( K \) using the one-standard-error criterion (Witten and Tibshirani 2010), which suggested that two clusters was the best choice. After fixing the number of clusters, we determined \( s \) as the lowest \( s \) within 1 standard deviation of the max gap statistic value using two clusters (Witten and Tibshirani 2010). We did not test more than six clusters, as visualization of the gap statistic indicated more clusters would lead to more overfitting and the range of \( s \) was set according to the defaults in Pysparcl 1.4.1 (Tsumeruso 2019), which was used for performing the sparse \( K \)-means clustering.
2.9. Statistical analysis

Results are shown as mean with 95% confidence intervals. Non-parametric Wilcoxon signed-rank tests were used to test differences in balanced accuracy between classifiers, and Mann–Whitney U rank test was employed for non-parametric unpaired differences between questionnaire scores between subtypes and group mean connectivity comparisons. Pearson’s correlation was used to test for correlations. Fisher’s exact test was employed to test whether the two subtypes were homogeneously spread across the three different studies. Multiple testing correction was performed using false discovery rate (FDR). The significant level was 0.05 for all hypothesis tests.

3. Results

3.1. Resting-state functional connectivity can predict PTSD

The machine learning classifiers were trained on all features or on each feature type separately, to investigate if any feature type on its own could yield similar or better performance compared to using combinations of feature types (figure 3(A)). The best performing classifier was SVM using all features, with a balanced test accuracy of 62.9% (95% CI: [61.0, 64.7], sensitivity: 64.6%, specificity: 61.2%) and area under the receiver operating characteristic curve of 0.68 (AUC; figure 3(B)). The logistic regression and random forest using all features had similar performances with above 60% balanced test accuracy, and all three models using all features performed significantly better than chance (all three Wilcoxon signed-rank tests with \( p < 0.001 \), FDR corrected for multiple testing). While none of the single feature type classifiers performed above 60% balanced test accuracy, classifiers using 1/f exponents, GC, Imcoh and wPLI had only slightly lower balanced test accuracies than the classifiers using all features (figure 3(A)), and this pattern was consistent for the three types of machine learning classifiers (supplementary table C1).

To determine the features driving the classifications, we evaluated the selected features in our best performing classifier. Using a threshold of presence in at least 20% of all CV runs, we observed that the features selected were predominantly functional connectivity features, with Imcoh and GC as the most common features, followed by wPLI. The only consistently selected non-connectivity feature was the 1/f exponent. We employed t-distributed stochastic neighbour embedding (t-SNE; van der Maaten and Hinton 2008) on the consistently selected features, and visually investigated the inherent clustering into PTSD and control groups based on the selected features. We observed some separation with the PTSD group being more prominently together in the upper right corner, while there was a cluster with mostly controls in the lower left corner. However there was still very high overlap between the two groups, consistent with the moderate classification performances we observed (figure 3(C)). The same pattern of overlap was observed with principal component analysis (PCA) visualization (supplementary figure D1(A)).

3.2. Two neurophysiologically distinct subtypes are defined within the PTSD group

Given the modest accuracy of our best model and high overlap between the two groups, we speculated that the classifier only characterized a portion of the PTSD group and was unable to find a general pattern for the whole PTSD group, partly due to the heterogeneous nature of PTSD. Thus we applied clustering on the PTSD group to attenuate some of the heterogeneity by identifying EEG-driven subtypes. Two subtypes were found with sparse K-means clustering, which primarily differed in their functional connectivity, most notably in their wPLI, GC and PEC values during the eyes-closed rest condition (figure 4(A)). This difference was frequency specific, with the alpha band being the most prominent band (figure 4(B)), followed by theta and beta band. When we compared mean alpha wPLI in the control versus PTSD group, we did not observe any clear qualitative difference, but when we compared the mean alpha wPLI in the two subtypes with the controls, we observed that subtype 1 had higher connectivity, especially in the parietal, temporal, and visual areas, and the posterior cingulate cortex. Subtype 2 had relatively similar connectivity to the control group, albeit with slightly lower connectivity (figure 4(C)). The same connectivity patterns described for the alpha band were also observed in the theta and beta band (supplementary figures D2 and D3). To quantify the differences, we averaged the alpha wPLI across all connections to compute a global alpha wPLI for each participant. We found no difference in global alpha wPLI between control and PTSD group (\( p = 0.763 \), Cohen \( d = 0.115 \)), significantly higher global alpha wPLI in subtype 1 compared to controls (\( p < 0.001 \), Cohen \( d = 1.857 \)), and significantly lower global alpha wPLI in subtype 2 compared to controls (\( p = 0.003 \), Cohen \( d = -0.586 \), all three tests were Mann–Whitney U rank test with FDR correction for the three contrasts).

Further evaluation revealed that the mean alpha wPLI across participants in the control group was strongly positively correlated with mean alpha wPLI across participants in both subtypes (Pearson’s \( r > 0.90 \)); however subtype 2 had a mean absolute difference across connections of 0.0282 (figure 4(D)), while subtype 1 had a considerably higher difference of 0.1032 (figure 4(E)). Interestingly, subtype 1 had significantly higher arousal symptom score compared to subtype 2 (\( p = 0.0456 \), Mann–Whitney U rank test with FDR for all three sub-symptoms), while no significant differences were observed for the intrusion and avoidance symptom scores (figure 4(F)). There
was also a difference in the number of participants clustered in each subtype, with 29 participants with electrophysiological activity corresponding to subtype 1 and 78 participants to subtype 2. We examined the distribution of the two subtypes across the three different studies to verify that the subtypes were not clustered due to study effects. Overall, subtype 1 had slightly higher amount of patients from study 3 and correspondingly fewer from study 1 compared to study 2, however these differences were not significant ($p = 0.195$, Fisher’s exact test; supplementary figure D4).

3.3. Better diagnostic predictions are obtained on PTSD subtypes

To evaluate if mitigation of heterogeneity with subtyping also translated to a better diagnostic performance, we reapplied the feature selection and classification pipeline, but instead of predicting on the whole PTSD group, we classified individuals into subtype or control group. The models were trained separately for each subtype and the features were restricted to the sparse $K$-means selected features from the eyes-closed condition. We observed high performance for classification of subtype 1 from controls, with logistic regression using wPLI obtaining 79.4% balanced test accuracy (95% CI: [77.0, 81.8], sensitivity: 66.0%, specificity: 92.8%, AUC: 0.92; figures 5(A) and (B)). The best classification of subtype 2 from controls was SVM using all features, which obtained 63.1% balanced test accuracy (95% CI: [60.7, 65.4], sensitivity: 64.8%, specificity: 61.3%, AUC: 0.66; supplementary figure D5). Table 1 provides an overview of the classification results. The results were consistent across all three machine learning classifiers (supplementary tables C2 and C3).

The best classification performance of subtype 2 was not different from the best classification of the whole patient group ($p = 0.370$, Wilcoxon signed-rank tests, FDR corrected for three comparisons). However, the best classification performance of subtype 1 was better than both classification of subtype 2 or the whole patient group ($p < 0.001$, Wilcoxon signed-rank tests, FDR corrected for three comparisons), consistent with subtype 1 having brain connectivity clearly different from controls. To better understand the wPLI connections that provided the best predictive information about subtype 1, we again employed t-SNE on the features that consistently provided important information for the classifier in
Figure 4. Data-driven clustering identified two neurophysiologically distinct subtypes in the PTSD group. (A) Sparse K-means was applied on all features and two clusters of the PTSD group were identified, primarily based on wPLI, GC and PEC. (B) Specifically, alpha wPLI, GC and PEC connectivity were important for the clustering, followed by some wPLI and GC connections in the theta and beta frequency band. (C) The mean alpha wPLI was very similar between the control and PTSD group, but after clustering, we observed a clear difference with higher wPLI in parietal, temporal and visual areas in subtype 1 compared to controls, while subtype 2 was relatively similar to controls, albeit with slightly lower wPLI. Mean alpha wPLI was strongly correlated between the two subtypes and controls (Pearson’s $r > 0.90$), although (D) the mean alpha wPLI was greater in subtype 1 for all areas compared to controls, while (E) subtype 2 had lower wPLI connectivity. (F) Subtype 1 had significantly higher arousal severity scores compared to subtype 2. Mean ± 95% confidence intervals are shown. MAE, mean absolute error.

at least 20% of all CV runs. Only 12 wPLI connections were robustly selected and using just these features, we observed a very clear separation of subtype 1 from the controls (figure 5(C)). The same pattern of clear separation was observed with PCA visualization (supplementary figure D1(B)).

To further investigate the underlying neurophysiological background for the prediction of subtype 1, we visualized the 12 wPLI connections and their importance. The feature that was most selected was beta wPLI between the anterior cingulate cortex and orbital gyrus, followed by multiple alpha wPLI connections between various areas from the dorsal attention network (DAN) and ventral attention network (VAN). Connections across the three frequency bands between the visual cortex and the two attention networks were also used by the classifier, and one connection between the somatosensory cortex and DAN in the alpha frequency range. In total one theta, eight alpha and three beta wPLI connections were consistently selected in more than 20% of all CV runs by the classifiers for prediction of subtype 1 (figure 6(A)).

Since we observed higher arousal severity symptoms in subtype 1 compared to subtype 2, we investigated whether some of the important wPLI connections were correlated with arousal severity. All eight alpha and one beta wPLI connection showed a significant positive Pearson’s correlation after FDR correction. Figure 6(B) shows examples of the four most significant correlations between the consistently selected wPLI features and arousal severity. The same significant results were obtained with Spearman’s rank correlation.
Figure 5. Subtyping improved machine learning prediction. (A) Performance of logistic regression using the sparse K-means identified eyes-closed EEG features. The best model using only wPLI features obtained 79.4% balanced test accuracy. The dashed line indicates chance level performance. (B) Receiver operating characteristic curve for the logistic regression using wPLI. (C) t-SNE of the important wPLI features selected by logistic regression. Only features that were selected in at least 20% of all cross-validation runs were included in the t-SNE. A clear separation between the control group and subtype 1 can be observed using the 12 most consistently selected wPLI features. Mean ± 95% confidence intervals are shown.

4. Discussion

In this study, we implemented a novel comprehensive framework for applying machine learning to investigate the ability of common resting-state EEG features to classify combat-related PTSD. We found that SVM with all features obtained a balanced test accuracy of 62.9%, which was significantly better than chance-level predictions. Further evaluation revealed that the functional connectivity features were the most important features driving the classifications (figure 3). The moderate performance in classifying the whole PTSD group from controls was consistent with the high heterogeneity and conflicting EEG results regarding PTSD previously found in the literature.

To mitigate some of the neurophysiological heterogeneity, we applied unsupervised sparse K-means clustering on the PTSD group. Two distinct subtypes were found, which differed mostly in their functional connectivity. Although we did not find a difference between the controls and the whole PTSD group, after subtyping, we observed that subtype 2 was relatively similar to controls, while subtype 1 had higher connectivity (figure 4). Interestingly, the best classifier obtained on average 79.4% balanced test accuracy on the test sets for classification of subtype 1, which was a clear significant improvement compared to when we predicted on the whole PTSD group, suggesting that the subtyping successfully mitigated some of the heterogeneity, and made it easier for the models to find specific EEG features related to PTSD subtypes.

Further evaluation of the specific connectivity features that were robustly utilized by our best classifier revealed 12 wPLI features from theta, alpha and beta frequency bands. The most consistently selected feature was a connection between the frontoparietal control network and VAN in the beta frequency bands.
The consistently selected features are correlated with arousal symptom severity. (A) Visualization of the 12 most consistently selected wPLI features by the logistic regression trained on classifying subtype 1 from controls. Most of the features were alpha connections between the dorsal and ventral attention networks, with some connections involving the frontoparietal control network in the beta band. One connection between the visual cortex and attention networks was also observed across all three frequency bands. (B) Example of the four most significantly correlated wPLI connections with arousal severity. All eight alpha wPLI connections were significantly positively correlated with arousal severity.

Both the DAN and VAN are key mediators for attention, and distinct attentional subprocesses have been attributed to them. The DAN is thought to mainly modulate attention towards external objects, whereas the VAN has been linked to internally directed cognition. The VAN is also sometimes referred to as the salience network or part of the salience network, due to its role in assigning salience and motivation. Thus both the DAN and VAN are important for mediating different subprocesses of attention, however the interplay and connections between both attention networks are also very important to take into account. Activity in the two attention networks has been found to correlate, but also anti-correlate depending on the task, and the frontoparietal control network is a likely candidate for governing the interplay between the two attention networks (reviewed in Menon 2011, Harmony 2013, Vossel et al 2014, Uddin et al 2019).

Previous studies have also found evidence supporting the notion of an abnormal attention network in PTSD. A one-year longitudinal EEG study found that worsened PCL scores were associated with increased delta activity in Broadman areas 13 and 44, which are part of the VAN and DAN (Jin et al 2021). Another study employed magnetoencephalography and found wPLI hyperconnectivity in the high gamma band to be associated with PTSD (Dunkley et al 2014). An fMRI study has also found an association between aberrant functional connectivity within the VAN and poor response to psychotherapy treatment (Etkin et al 2019). Thus several studies, using different neuroimaging modalities, have found evidence suggesting that attention networks in PTSD patients are abnormal.

Interestingly, among the 12 important wPLI features for prediction of subtype 1, all the alpha wPLI connections, which were primarily between the DAN and VAN, were positively correlated with arousal severity (figure 6(B)). Patients with PTSD often have impaired attentional control and display hypervigilance towards their surroundings, which could partly be explained by an impaired attention network, where threat processing is exaggerated at the cost of mental exhaustion and normal daily functioning (Shvil et al 2013). Alpha oscillations are also thought to play an active role in selective attention. During processing of stimuli, alpha oscillations have been observed to be suppressed in brain areas processing the stimuli, while increased alpha activity has been observed in brain areas not directly involved in the processing of the specific stimuli. Thus alpha oscillations are thought to mediate active suppression of irrelevant or distracting information during selective attention (Klimesch et al 2007, Rihs et al 2007, Fries et al 2008, Foxe and Snyder 2011), by having an inhibiting role on neuronal communication, in contrast to gamma oscillations, which become more prominent during processing of stimuli and are thought to enhance neuronal communication.
(Fries 2015). While the precise functions of alpha and gamma oscillations are still debated, there is consensus that they mediate attentional processes (Fries 2015). Thus if functional connectivity in the alpha frequency range is altered, as we observed in subtype 1, it may be associated with impaired attentional processes. Additionally, we also observed that subtype 1 had significantly greater arousal severity compared to subtype 2, which had more normal electrophysiological activity (figure 4(F)). This finding supports the notion of classification of clinical subgroups based on electrophysiological brain circuit dysfunctions as described in the RDoC (Insel et al 2010).

However, it is important to note that seemingly, only a portion of the PTSD patients in our dataset, i.e. subtype 1, had alpha wPLI hyperconnectivity. Additionally, this hyperconnectivity only partly explains the increased arousal in PTSD, which is evident in the weak correlations around $r \sim 0.25$ and the fact that subtype 2, which do not display these abnormal functional connectivity patterns, still have much higher arousal scores than controls. This further highlights the heterogeneous nature of PTSD, and stresses the importance of defining more homogeneous posttraumatic stress phenotypes.

Other studies have also tried finding quantifiable diagnostic biomarkers for PTSD using machine learning. One study investigated the P3-evoked response to a visual stimuli and reported around 90% sensitivity and specificity using a Fisher’s linear discriminant analysis (LDA) for classification of veterans with PTSD compared to healthy combat veterans with a similar military history (Attias et al 1996). However, the machine learning approach they employed had no formal CV scheme and they trained and tested on some of the same subjects, hence the performance values ranging from not better than chance level up to 100% accuracy on classifying subtype 1 obtained above 80% training accuracies, reflecting the inflation of the performance metric when applied to the training set. Most recent studies are now adapting CV schemes similar to the one used here, to better estimate the true generalization performance of their models. Dean et al (2020) examined data from molecular arrays of blood samples from combat veterans with and without PTSD and employed a machine learning pipeline with multiple feature selection algorithms to identify potential biomarkers before validating the biomarkers in an independent cohort. Although the type of data is different, their overall approach is similar to our pipeline. Using a random forest trained with one-layer CV for parameter tuning, they obtained a good performance with 81% accuracy on their independent test cohort. The approach of having a separate unseen test dataset for evaluation of the models is sometimes referred to as using a lockbox/hold-out dataset and circumvents the problem of training and testing on the same dataset (Hosseini et al 2020). However, while using unseen data for estimation of a generalization performance metric is definitely necessary, a single test set might not yield a reliable and robust result. We found a high variance in test accuracies, both for predicting the whole PTSD group (figure 3(A)) or the subtypes (figures 5(A) and D5(A)). The high variance in test accuracies also shows how big an impact the exact split of the data could have on our machine learning models, and reflects the general problems with heterogeneity and reproducibility. Using the exact same methodology and models, but training and testing on slightly different participants, we obtained performance values ranging from not better than chance level up to 100% accuracy on classifying subtype 1 in unseen test data. Thus when only testing on single hold-out dataset, it might have yielded a ‘lucky’ test accuracy of up to 100%, which clearly highlights the

| Diagnostic classification | BACC (%) | 95% CI |
|---------------------------|----------|--------|
| All PTSD patients         |          |        |
| Best performance: SVM with all features | 62.9 | [61.0, 64.7] |
| Median performance: SVM with wPLI | 54.6 | [52.5, 56.7] |
| Worst performance: logistic regression with PEC | 47.5 | [45.2, 49.7] |
| Subtype classification    |          |        |
| PTSD subtype 1            |          |        |
| Best performance: logistic regression with wPLI | 79.4 | [77.0, 81.8] |
| Median performance: logistic regression with Imcoh | 63.5 | [60.5, 66.5] |
| Worst performance: SVM with microstates | 50.0 | [49.9, 50.0] |
| PTSD subtype 2            |          |        |
| Best performance: SVM with all features | 63.1 | [60.7, 65.4] |
| Median performance: SVM with wPLI | 56.2 | [53.7, 58.6] |
| Worst performance: logistic regression with microstates | 50.7 | [49.0, 52.4] |

Accuracies are averaged across ten repetitions of 10-by-10 cross-validation runs. BACC, balanced accuracy.

CI, confidence interval.
importance of using a stringent robust evaluation of machine learning classifiers. Hence we recommend evaluating the performance through a two-layer CV, if hyperparameter optimization is needed, or one-layer CV for models without hyperparameter tuning. If possible, repetitions using different data splits for the CV should be conducted in order to obtain a more robust and reliable result.

The setting with ten repetitions of ten-fold CV was also employed by two recent papers investigating resting-state EEG in PTSD patients. Using a combination of source space covariance matrices, band power and network metrics, Kim et al. (2020) obtained around 64%, 65% and 62% test accuracies for LDA, SVM and random forest respectively. These performance results are consistent with the results we observed when predicting on the whole PTSD patient group. Interestingly, they also employed a Riemannian geometry-based classifier composed of a geodesic filter and minimum distance to the mean method and obtained a noticeable increased performance with around 73% test accuracy, suggesting that Riemannian based classifiers might outperform conventional classifiers on predicting PTSD.

Toll et al. (2020) also employed ten repetitions of ten fold CV to estimate the generalization performance and obtained an AUC of 0.813 (sensitivity: 76.3%, specificity: 74.9%). This result was observed with a linear kernel-based relevance vector machine using source space PEC as features in combat veterans with PTSD and combat-exposed veterans as controls. We also tested the performance of our three classifiers using only PEC features for predicting the whole PTSD group, however PEC was consistently shown to have the worst generalization performance in our dataset. We cannot provide a clear reason why PEC was distinctively worse than the other connectivity features for our dataset. We also tested the performance of our three classifiers using only PEC features for predicting the whole PTSD group, however PEC was consistently shown to have the worst generalization performance in our dataset. We cannot provide a clear reason why PEC was distinctively worse than the other connectivity features for our dataset.

Another limitation of the present study is the estimation of source level connectivity using relatively few electrodes. Ideally, more electrodes would improve the spatial resolution of the source localization. Nonetheless, the data analysed in the present study were obtained in a clinical setting, ranging from 19 to 31 electrodes, thus any result observed would be more likely to translate to the clinic. Going to source space also alleviates the effect of volume conduction, which functional connectivity is particularly sensitive to (Michel and Brunet 2019).

Only subtype 1 had an increase in predictive performance and obtained around 80% balanced accuracy, while subtype 2 still had similar predictive results around 63% balanced accuracy. Since subtype 2 corresponds to 73% of the PTSD group, it limits the clinical relevance of our findings. However, it is important to note that our control group consisted of combat-exposed veterans, who experienced the same type of traumas as our cases. Despite being considered controls in our dataset, many of the controls exhibited stress symptoms and were treatment-seeking, but did not meet the threshold for clinical PTSD. Thus the modest accuracy we observed on predicting the whole PTSD group and subtype 2 should be considered in relation to our control group, which are also veterans exposed to combat with various degrees of sub-clinical symptoms, as opposed to characterizing combat-related PTSD patients from healthy civilians unexposed to traumatic experiences. The latter classification has been shown to be easier, most likely due to greater between groups differences and less overlap in phenotypes. A resting-state fMRI study estimated functional connectivity features and obtained 92.5% accuracy with a SVM classifier in predicting PTSD from non-trauma exposed controls (Liu et al. 2014), and a structural MRI study reported 91% accuracy when comparing non-trauma exposed healthy controls with PTSD patients, but only 67% accuracy when comparing trauma exposed PTSD and controls using SVM (Gong et al. 2014).

Lastly, the sample size in our dataset is also a limiting factor. More samples are usually thought to be better, although many studies with greater sample sizes in neuroimaging have lower accuracies than small sample size studies (Arbabshirani et al. 2017, Woo et al. 2017). However, this is just the consequence of small sample size studies overfitting to their data, which would lead to a lack of reproducibility. Our sample size of more than 200 is relatively high for a clinical EEG study (Arbabshirani et al. 2017), but having more samples would not only mean the patient population would be better represented, but...
it would also make the machine learning models less likely to overfit and enable identification of more subtypes (Arbabshirani et al. 2017).

5. Conclusion

Taken together, the novel machine learning framework presented in this study has shown that the occurrence of PTSD can be classified above chance primarily from functional connectivity between EEG source regions. Delineating electrophysiological heterogeneity by identification of two distinct subtypes improved the classification performance on a portion of the PTSD patients, which was characterized by hyperconnectivity in parietal, temporal, visual areas, and posterior cingulate cortex. Further evaluation revealed alpha wPLI between DAN and VAN to be particularly important for the classification of the subtype, and these connections were correlated with arousal symptom severity. Overall, the novel framework presented here demonstrates how combining subtyping and machine learning classification enabled identification of potential quantifiable biomarkers for PTSD subtypes. This generic framework can be readily expanded to other datasets and psychiatric disorders. In the future we envision an overall similar approach, using data from more samples and multiple modalities, e.g. genetics, blood samples, neuroimaging, social behaviour etc, and with longitudinal data, will enable clinicians to administer tailored diagnosis and treatment for each subject.

Data availability statement

The code used for this study is available from https://lab.compute.dtu.dk/glia/resting-state-eeg-analysis. The data generated and/or analysed during the current study are not publicly available for legal/ethical reasons but are available from the corresponding author on reasonable request.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. No conflicts of interest are declared by the authors.
Appendix A. Overview of PTSD resting-state EEG power spectral density findings

Table A1. Overview of previous findings on power spectral density in PTSD. The results from resting-state spectral power EEG studies on PTSD were evaluated for whether they found a significant difference and the direction of the difference between a PTSD and control group.

|                          | Delta $\delta$ (1 - 4Hz) | Theta $\theta$ (4 - 8Hz) | Alpha $\alpha$ (8 - 13Hz) | Beta $\beta$ (13 - 30Hz) | Gamma $\gamma$ (30 - 50Hz) |
|--------------------------|--------------------------|-------------------------|----------------------------|--------------------------|---------------------------|
| Significantly Increased Power | 7                        | 1, 3                    | 1, 9, 12                   | 2, 7, 8                  |
| Significantly Decreased Power | 5, 9                    | 8, 9, 12                |                            |                          |
| No Significant Changes    | 1, 2, 3, 6, 9, 12        | 2, 6, 7, 12             | 1, 2, 3, 4, 6, 7, 10, 11   | 2, 3, 6, 7              |

Figure A1. Overview of previous findings on power spectral density in PTSD. The results from resting-state spectral power EEG studies on PTSD were evaluated for whether they found a significant difference and the direction of the difference between a PTSD and control group.

Appendix B. Supplementary methods

Additional information and mathematical equations for the estimated EEG features are provided here.

B.1. Spectral analysis

EEG power was computed using multitaper spectral estimation with seven discrete prolate spheroidal sequences windows for the five canonical frequency bands. Two transformations were applied to the power features; absolute power was converted into decibels (dB) by taking the log base 10 and multiplying with 10, and relative power was computed as proportion of power in each frequency band normalized to the total power in all frequency bands.

EEG power asymmetry was computed as the log transformed raw power (i.e. prior to absolute/relative transformation) in each left hemispheric channel, subtracted from the corresponding channel in the right hemisphere (Sutton and Davidson 1997):

\[ \text{Asymmetry} = \ln(\text{Power}_{rh}) - \ln(\text{Power}_{lh}). \]  

(B1)

The asymmetry values were averaged across the frontal (F3, F4, F7, F8, Fp1, Fp2), central (C3, C4, CP3, CP4, FC3, FC4, FT7, FT8, T7, T8, TP7, TP8) and posterior (O1, O2, P3, P4, P7, P8) regions respectively to yield frontal, central and posterior asymmetry.

Frontal theta/beta ratio was estimated by taking the averaged raw theta and beta power in the frontal channels (Fp1, Fpz, Fp2, AFz, Fz, F3, F4, F7, F8), followed by computing their ratio and applying the natural logarithm.

Peak alpha frequency and $1/f$ exponents were estimated using the FOOOF algorithm (Donoghue et al 2020). The FOOOF algorithm assumes the power spectral density is composed of an aperiodic ($1/f$) component and periodic components (e.g. alpha oscillations). A power law is used to fit the aperiodic component, while multiple Gaussian fits are utilized to fit the periodic components. We used the $R^2$ of the fit of the full model to evaluate how well the algorithm worked. Based on visual inspection of the fits and the $R^2$ values, we decided that for our data, the peak alpha frequency estimations of FOOOF fits with an $R^2 > 0.90$ was reliable. For the $1/f$ exponents, we set the thresholds to be $R^2 > 0.95$. If the $R^2$ was lower than the threshold, the peak alpha frequency or $1/f$ exponent values were set to NaN. Due to the presence of delta/theta peaks, which were harder to fit due to their location on the lower end of our estimated power spectra, the $R^2$ values might be below the threshold, despite the fit being good. To circumvent this specific problem, we iteratively tried fitting the FOOOF algorithm using 2, 3, 4, 5 or 6 Hz as the start frequency, until the $R^2$ was above the thresholds. The end range of the fit was set to 40 Hz. Peak alpha frequency and $1/f$ exponents were calculated for each channel, and a global peak alpha frequency was also computed as the average over all channels. Channel-wise mean imputation was applied to the NaN values for the machine learning modelling.

B.2. Long-range temporal correlations

Long range temporal correlations were computed according to the procedure described by Hardstone et al (2012). Briefly, the epochs were concatenated back into time series of up to 1 min long corresponding to their respective eye condition intervals, and band-pass filtered into the canonical frequency bands. The Hilbert transform was applied to calculate the instantaneous amplitude envelopes, and detrended fluctuation analysis (DFA) was performed to estimate the DFA exponent, which is also known as the Hurst exponent. The window sizes for the DFA were estimated for each frequency band by simulation of white-noise signals using statistics from our.
Figure B1. Source localization parcellation. The Destrieux Atlas parcellation was manually modified to approximate the 31 ROIs used by Zhang et al (2021). The areas were: visual network: 1. visual area 1 (V1); somatosensory network: 2. somatosensory cortex (SMC); dorsal attention network: 3. inferior frontal junction (IFJ), 4. intraparietal sulcus (IPS), 5. frontal eye fields (FEFs), 6. supplementary eye fields (SEFs); default mode network: 7. posterior cingulate cortex (PCC), 8. medial prefrontal cortex (mPFC), 9. angular gyrus (ANG); fronto-parietal control network: 10. posterior middle frontal gyrus (PMFG), 11. inferior parietal lobule (IPL), 12. orbital gyrus (ORB), 13. middle temporal gyrus (MTG); ventral attention network: 14. anterior middle frontal gyrus (AMFG), 15. insula (INS), 16. anterior cingulate cortex (ACC), 17. supramarginal gyrus (SUP).

B.3. Microstates

EEG microstates are characteristic global scalp potential topographical maps that remain stable for varying durations, and are thought to reflect different functional brain states (Lehmann et al 1987). No frequency decomposition was applied for microstate analysis. The modified K-means algorithm was used to find the microstates that best explain the variance in the topographical maps, i.e. find the most stable maps (Pascual-Marqui et al 1995). After the microstate maps had been determined, they were competitively back-fitted to each participant’s time-course EEG signals, and labels were estimated by assigning the microstate with the highest spatial correlation with the instantaneous topography at each time point. Finally, ratio of time covered, microstate entropy, and transition matrices were computed for the microstates. The python implementation by von Wegner and Laufs (2018) was used for microstate analysis.

B.4. Source functional connectivity

Interpolation of channels might lead to spurious connectivity, especially with neighbouring channels. To circumvent this problem and mitigate the effect of volume conduction (Michel and Brunet 2019) on the functional connectivity measurements, we dropped all interpolated channels prior to applying source localization, followed by estimation of functional connectivity measurements. The benefits of source localization were three-fold: (a) it makes it possible to compare results across the studies, (b) by giving estimates on cortical brain areas instead of the electrode locations, (c) while also alleviating volume conduction (Palva and Palva 2012, Donoghue et al 2022). We adapted the source localization methodology from Zhang et al (2021), since they also investigated source functional connectivity in regards to PTSD. Briefly, minimum-norm estimation was performed to transform our sensor space EEG to source time series of 20,484 vertices, followed by extraction of time series from 31 cortical regions of interest (ROIs). Specifically, we collapsed the vertices by finding the dominant direction of the vector orientations within each ROI and applied sign-flip to the time series at vertices that were more than 180° different from the dominant direction, before averaging across vertices within each ROI. The Destrieux Atlas (Destrieux et al 2010) parcellation was manually modified to approximate the 31 ROIs used by Zhang et al (2021) (Supplementary figure B1). The 31 ROIs were originally derived from independent component analysis of resting-state fMRI (Toll et al 2020), and consist of core cortical regions of six functional connectivity networks, ubiquitously observed in both task and resting-state fMRI studies (Uddin et al 2019). The FreeSurfer average brain template from FreeSurfer 6 (Fischl 2012) was used...
to construct the boundary element head model and forward operator for the source modelling. Additionally, free orientations were used for the dipoles and the regularization parameter for the minimum-norm estimation was set to $\lambda^2 = 1/81$. Principal component analysis was performed to reduce the three-dimensional source signals to one-dimensional by projecting onto the dominant principal direction. Following estimation of source time series, the various functional connectivity measurements were calculated.

### B.5. Imaginary coherence

Coherence can be viewed as the frequency domain analogue to Pearson’s correlation in the time domain, and both measure the linear dependency of two signals. Coherence is calculated as the magnitude of the cross spectrum between two signals divided by the square root of the product of each signal’s power spectrum for normalization (Nunez et al 1997). However, one problem with coherence is that it is sensitive to volume conduction, but by using only the imaginary part of coherence, the effect of zero-phase lag synchronization, which is the hallmark of volume conduction, can be removed (Nolte et al 2004)

$$\text{Imcoh} = \frac{\text{Im}(G_{xy})}{\sqrt{G_{xx}G_{yy}}}.$$  \hfill (B2)

Here $G_{xy}$ is the cross-spectral density between channels $x$ and $y$ and $G_{xx}$ and $G_{yy}$ is the power spectra of each signal respectively. An imaginary coherence value close to 1 indicates strong synchronization, while a value close to 0 reflects no synchronization between the two signals. Imcoh were calculated for each epoch and then averaged. The drawback of using Imcoh is that true non-volume conducted neuronal synchronization around zero-phase lag will also be attenuated, thus potentially leading to an underestimation of the true synchronization.

### B.6. Weighted phase lag index (wPLI)

The cross-spectral density contains information about both the amplitudes and the phase difference between the two signals of interest. When estimating the synchronization between two signals, using only the relative phase information and disregarding the amplitudes might be better at capturing the underlying neural activity as it becomes less sensitive towards noise. This is the idea behind the phase locking value (Lachaux et al 1999). However, a common source due to volume conduction might lead to consistent zero-phase differences and would thus inflate the phase locking value. Thus, the phase lag index was developed to circumvent this problem, by calculating how consistent one signal leads/lags behind the other signal. This is based on the assumption that if one signal consistently leads the other signal, then there is also a consistent non-zero-phase difference, whereas volume conduction would lead to a symmetrical phase difference distribution around zero-phase (Stam et al 2007)

$$\text{PLI} = \frac{1}{T} \left| \sum_{t=1}^{T} \text{sign}(|\sin(\Delta \phi_t)|) \right|$$  \hfill (B3)

where $\Delta \phi_t = \phi_{(x,t)} - \phi_{(y,t)}$ is the instantaneous phase difference in radians between the two signals of interest at timepoint $t$ and $T$ is the total number of timepoints within one epoch.

Similar to the idea behind imaginary coherence, the wPLI was developed to further attenuate the effect of volume conduction by weighing each phase difference with the magnitude of the lag, hence ensuring that phase differences around 0 will contribute minimally to the estimation of the connectivity (Vinck et al 2011, Hardmeier et al 2014)

$$\text{wPLI} = \frac{1}{T} \left| \sum_{t=1}^{T} \text{sign}(\sin(\Delta \phi_t)) \cdot \text{sign}(\sin(\Delta \phi_t)) \right|$$

$$\Leftrightarrow \text{wPLI} = \frac{\sum_{t=1}^{T} |\sin(\Delta \phi_t)| \cdot \text{sign}|\sin(\Delta \phi_t)|}{\sum_{t=1}^{T} |\sin(\Delta \phi_t)|}.$$  \hfill (B4)

The denominator normalizes the wPLI to the interval $0 \leq \text{wPLI} \leq 1$, where higher values indicate more synchronization. After estimation of wPLI in each epoch the values were averaged to obtain a more robust estimation.

### B.7. Power envelope correlations (PECs)

PEC were estimated following Toll et al (2020). Briefly, the time series for each signal pair were band-pass filtered for the canonical frequency bands, Hilbert transformed to obtain the analytical signals and then orthogonalized to each other. The orthogonalization ensures that the signal components which share the same phase are removed, i.e. volume conduction induced zero-phase lag correlations are removed (Hipp et al 2012). Following orthogonalization, the power envelope of the orthogonalized analytical signals are estimated by squaring the signals and log-transformed. Finally, Pearson’s correlations were calculated and Fischer’s $r$-to-$z$ transform applied to enhance normality.

### B.8. Granger causality (GC)

All previously mentioned connectivity measurements are symmetric, i.e. they do not contain information about the direction of the connectivity between the two ROIs. GC is a method that can infer the causal relationships, i.e. it provides directional information. The idea behind GC is that if $X$ ‘Granger causes’ $Y$, then it means that $X$ contains information that helps to predict the future of $Y$ better than if you only had the information of the past of $Y$. Mathematically, this can be represented with autoregressive functions and
by comparing an unrestricted model (equation (B5)) with a restricted model (equation (B6)):

\[
\begin{bmatrix}
X(t) \\
Y(t)
\end{bmatrix} = \sum_{k=1}^{p} A_k \cdot \begin{bmatrix}
X(t - kt_0) \\
Y(t - kt_0)
\end{bmatrix} + \begin{bmatrix}
E_x(t) \\
E_y(t)
\end{bmatrix}
\]

\[
Y(t) = \sum_{k=1}^{p} B_k \cdot Y(t - kt_0) + \tilde{E}_y(t).
\]

Here \( t \) denotes the time, \( t_0 \) is the time between successive observations, \( A \) and \( B \) are model coefficients, \( p \) is the model order and \( E_x, E_y \) and \( \tilde{E}_y \) are the residuals. GC can then be calculated as the log-ratio of the variance of the residuals in the restricted model and the unrestricted (Barrett et al 2012):

\[
\text{GC}_{X \rightarrow Y} = \ln \left( \frac{\text{var}(\tilde{E}_y(t))}{\text{var}(E_y(t))} \right).
\]
## Appendix C. Classification performance for all feature types and classifier combinations

### Table C1. Performance and confidence intervals of predicting the whole PTSD group from controls.

| Diagnostic classification          | BACC (%) | 95 % CI   |
|-----------------------------------|----------|-----------|
| All features                      |          |           |
| Best performance: support vector machine | 62.9    | [61.0, 64.7] |
| Median performance: random forest | 61.6    | [59.6, 63.5] |
| Worst performance: logistic regression | 60.3  | [58.2, 62.5] |
| Power                             |          |           |
| Best performance: support vector machine | 53.1    | [51.0, 55.1] |
| Median performance: random forest | 52.2    | [50.4, 54.1] |
| Worst performance: logistic regression | 51.8  | [49.6, 54.0] |
| ftBR + asymmetry                  |          |           |
| Best performance: logistic regression | 56.6    | [54.6, 58.7] |
| Median performance: support vector machine | 52.2    | [50.3, 54.2] |
| Worst performance: random forest | 51.2    | [49.3, 53.1] |
| Peak alpha frequency              |          |           |
| Best performance: random forest | 55.8    | [53.7, 58.0] |
| Median performance: logistic regression | 52.7  | [50.5, 54.9] |
| Worst performance: support vector machine | 51.6  | [49.7, 53.6] |
| 1/f Exponent                      |          |           |
| Best performance: support vector machine | 58.6    | [56.9, 60.3] |
| Median performance: random forest | 58.2    | [56.1, 60.2] |
| Worst performance: logistic regression | 57.1  | [55.0, 59.2] |
| Imaginary coherence               |          |           |
| Best performance: support vector machine | 58.4    | [56.4, 60.4] |
| Median performance: logistic regression | 57.6  | [55.4, 59.7] |
| Worst performance: random forest | 54.9    | [53.1, 56.7] |
| Weighted phase lag index          |          |           |
| Best performance: random forest | 59.0    | [57.1, 60.9] |
| Median performance: logistic regression | 57.8  | [55.7, 59.9] |
| Worst performance: support vector machine | 54.6  | [52.5, 56.6] |
| Power envelope correlation        |          |           |
| Best performance: random forest | 48.2    | [46.2, 50.2] |
| Median performance: logistic regression | 47.5  | [45.2, 49.7] |
| Worst performance: support vector machine | 47.5  | [45.5, 49.5] |
| Granger causality                 |          |           |
| Best performance: support vector machine | 58.6    | [56.4, 60.8] |
| Median performance: logistic regression | 57.1  | [55.1, 59.2] |
| Worst performance: random forest | 56.1    | [54.2, 57.9] |
| Microstates                       |          |           |
| Best performance: random forest | 54.5    | [52.6, 56.4] |
| Median performance: logistic regression | 53.7  | [51.9, 55.6] |
| Worst performance: support vector machine | 51.7  | [49.8, 53.6] |
| DFA exponent                      |          |           |
| Best performance: random forest | 52.2    | [50.2, 54.3] |
| Median performance: logistic regression | 52.0  | [49.8, 54.3] |
| Worst performance: support vector machine | 49.5  | [47.1, 51.9] |
Table C2. Performance and confidence intervals of predicting PTSD subtype 1 from controls.

| Subtype 1 classification | BACC (%) | 95 % CI       |
|--------------------------|----------|---------------|
| All features             |          |               |
| Best performance: support vector machine | 73.8     | [70.9, 76.8]  |
| Median performance: random forest   | 73.8     | [71.1, 76.4]  |
| Worst performance: logistic regression | 71.8     | [69.0, 74.7]  |
| Power                    |          |               |
| Best performance: support vector machine | 60.9     | [58.0, 63.8]  |
| Median performance: logistic regression | 58.6     | [55.8, 61.3]  |
| Worst performance: random forest   | 55.5     | [53.4, 57.6]  |
| Imaginary coherence       |          |               |
| Best performance: support vector machine | 65.1     | [62.6, 67.7]  |
| Median performance: logistic regression | 63.5     | [60.5, 66.5]  |
| Worst performance: random forest   | 61.1     | [58.6, 63.7]  |
| Weighted phase lag index   |          |               |
| Best performance: logistic regression | 79.4     | [77.0, 81.8]  |
| Median performance: support vector machine | 77.7     | [74.9, 80.5]  |
| Worst performance: random forest   | 72.2     | [69.5, 74.9]  |
| Power envelope correlation  |          |               |
| Best performance: random forest   | 62.7     | [59.8, 65.6]  |
| Median performance: support vector machine | 62.7     | [60.0, 65.4]  |
| Worst performance: logistic regression | 61.6     | [59.0, 64.3]  |
| Granger causality          |          |               |
| Best performance: logistic regression | 72.8     | [69.8, 75.8]  |
| Median performance: support vector machine | 70.1     | [67.0, 73.1]  |
| Worst performance: random forest   | 64.3     | [61.3, 67.3]  |
| Microstates                |          |               |
| Best performance: logistic regression | 54.0     | [52.1, 55.9]  |
| Median performance: random forest   | 51.7     | [49.9, 53.5]  |
| Worst performance: support vector machine | 50.0     | [49.9, 50.0]  |

Table C3. Performance and confidence intervals of predicting PTSD subtype 2 from controls.

| Subtype 2 classification | BACC (%) | 95 % CI       |
|--------------------------|----------|---------------|
| All features             |          |               |
| Best performance: support vector machine | 63.1     | [60.7, 65.4]  |
| Median performance: logistic regression | 61.6     | [59.3, 63.9]  |
| Worst performance: random forest   | 60.4     | [57.9, 62.8]  |
| Power                    |          |               |
| Best performance: logistic regression | 53.0     | [50.9, 55.1]  |
| Median performance: support vector machine | 51.1     | [48.7, 53.5]  |
| Worst performance: random forest   | 50.9     | [49.0, 52.9]  |
| Imaginary coherence       |          |               |
| Best performance: support vector machine | 55.9     | [53.9, 58.0]  |
| Median performance: logistic regression | 54.2     | [51.8, 56.7]  |
| Worst performance: random forest   | 52.5     | [50.4, 54.6]  |
| Weighted phase lag index   |          |               |
| Best performance: random forest   | 57.9     | [53.6, 60.2]  |
| Median performance: support vector machine | 56.2     | [53.7, 58.6]  |
| Worst performance: logistic regression | 55.8     | [53.5, 58.1]  |
| Power envelope correlation  |          |               |
| Best performance: random forest   | 58.3     | [56.1, 60.4]  |
| Median performance: logistic regression | 56.7     | [54.5, 58.9]  |
| Worst performance: support vector machine | 56.5     | [54.3, 58.7]  |
| Granger causality          |          |               |
| Best performance: random forest   | 62.8     | [60.6, 65.0]  |
| Median performance: support vector machine | 62.4     | [60.1, 64.6]  |
| Worst performance: logistic regression | 61.7     | [59.6, 63.8]  |
| Microstates                |          |               |
| Best performance: random forest   | 53.4     | [51.5, 55.3]  |
| Median performance: support vector machine | 52.2     | [50.5, 53.8]  |
| Worst performance: logistic regression | 50.7     | [49.0, 52.4]  |
Appendix D. Supplementary figures

Figure D1. PCA of the consistently selected features by the best classifiers. (A) PCA of the important features consistently selected by SVM trained on all features for classification of the whole PTSD group. (B) PCA of the important features consistently selected by logistic regression trained on sparse $K$-means wPLI features for classification of subtype 1. (C) PCA of the important features consistently selected by SVM trained on all sparse $K$-means features for classification of subtype 2.

Figure D2. Mean theta wPLI connectivity. No differences were observed between the PTSD and control group, but after clustering we observed higher wPLI in subtype 1, while subtype 2 was relatively similar to controls.
Figure D3. Mean beta wPLI connectivity. No differences were observed between the PTSD and control group, but after clustering we observed higher wPLI in subtype 1, while subtype 2 was relatively similar to controls.
Figure D4. Distribution of the subjects from the two PTSD subtypes across the three studies. Overall, the two subtypes contained relatively similar percentages of patients from the different studies, albeit subtype 1 had slightly higher amount of patients from study 3 and correspondingly fewer from study 1 compared to subtype 2.

Figure D5. Prediction of subtype 2 was not better than prediction of the whole PTSD group. (A) Performance of SVM using the sparse $K$-means identified eyes-closed EEG features. The best model using all the sparse $K$-means identified features obtained 63.1% balanced test accuracy. The dashed line indicates chance level performance. (B) Receiver operating characteristic curve for the best classifier. (C) t-SNE of the important features consistently selected in at least 20% of all cross-validation runs by the best classifier. Some separation between the two groups can be observed with subtype 2 being primarily prominent in the upper right corner, although there is still a large overlap, reflecting the moderate classification performance the classifier obtained. Mean $\pm$ 95% confidence intervals are shown.
dangers of unexpected overfitting in classification of brain data Neurosci. Biobehav. Rev. 119 456–67
Imperatori C, Farina B, Quintilianni M I, Onofri A, Castelli Gattinara P, Lepore M, Gnoni V, Mazzucchi E, Contardi A and Dellà Marca G 2014 Aberrant EEG functional connectivity and EEG power spectra in resting state post-traumatic stress disorder: a sLORETA study Biol. Psychol. 102 16–17
Insel T, Cuthbert B, Garvey M, Heinssen R, Pine D, Quinn K, Sanislow C and Wang P 2010 Research domain criteria (RDoC): toward a new classification framework for research on mental disorders Am. J. Psychiatry Online 167 748–51
Jas M, Engemann D A, Bekhti Y, Raimondo F and Gramfort A 2017 Autoreject: automated artifact rejection for MEG and EEG data NeuroImage 159 417–29
Jin S, Shin C, Han C, Kim Y-K, Lee J, Jeon S W, Lee S-H and Ko Y-H 2021 Changes in brain electrical activity according to post-traumatic stress symptoms in survivors of the Sewol ferry disaster: a 1 year longitudinal study Clin. Psychopharmacol. Neurosci. 19 537–44
Jokić-Begić N and Begić D 2003 Quantitative electroencephalogram (qEEG) in combat veterans with post-traumatic stress disorder (PTSD) Nordic J. Psychiatry 57 351–5
Karstoft L I, Andersen S B, Bertelsen M and Madsen T 2014 Diagnostic accuracy of the posttraumatic stress disorder checklist-civilian version in a representative military sample Psychol. Assess. 26 321–5
Kemp A H, Griffiths K, Felmingham K L, Shankman S A, Drinkenburg W, Arns M, Clark C R and Bryant R A 2010 Disorder specificity despite comorbidity: resting EEG alpha asymmetry in major depressive disorder and post-traumatic stress disorder Biol. Psychol. 85 390–4
Keyman J N et al 2019 Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience Nat. Human Behav. 3 63–73
Kim Y-w, Kim S, Shim M, Jin J and Jeon H 2020 Riemannian potential neurophysiological correlates differentiating posttraumatic stress disorder from anxiety disorders Psychiatry Invest. 15 1087–93
Neria Y 2021 Functional neuroimaging in PTSD: from discovery of underlying mechanisms to addressing diagnostic heterogeneity Am. J. Psychiatry 178 128–35
NIPY 2019 Nitime—time-series analysis of neuroscience data (available at: http://nipy.org/nitime/index.html)
Nolte G, Bir O, Wheaton L, Mari Z, Vorbach S and Hallett M 2004 Identifying true brain interaction from EEG data using the imaginary part of coherency Clin. Neurophysiol. 115 2292–307
Nunez P L, Srinivasan R, Westdorp A F, Wijesinghe R S, Tucker D M, Silverstein R B and Cadusch P J 1997 EEG coherency I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging and interpretation at multiple scales Electroencephalogr. Clin. Neurophysiol. 103 499–515
Palva S and Palva J M 2012 Discovering oscillatory interaction networks with M/EEG: challenges and breakthroughs Trends Cogn. Sci. 16 219–30
Pascual-Marqui R D, Michel C M and Lehmann D 1995 Segmentation of brain electrical activity into microstates; model estimation and validation IEEE Trans. Biomed. Eng. 42 658–65
Pedregosa F et al 2011 Scikit-learn: machine learning in python J. Mach. Learn. Res. 12 2825–30 (available at: www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf)
Peng H, Long F and Ding C 2005 Feature selection based on mutual information: criteria of max-dependency, max-relevance and min-redundancy IEEE Trans. Pattern Anal. Mach. Intell. 27 1226–38
Perrin F, Pernier J, Bertrand O and Echallier J F 1989 Spherical splines for scalp potential and current density mapping Electroencephalogr. Clin. Neurophysiol. 72 184–7
Poldrack R A, Huckins G and Varoquaux G 2020 Establishment of best practices for evidence for prediction: a review JAMA Psychiatry 77 534–40
Rabe S, Beauducel A, Zöllner T, Maercker A and Karl A 2006 Regional brain electrical activity in posttraumatic stress disorder after motor vehicle accident J. Abnormal Psychol. 115 667–80
Raschka S 2018 MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack J. Open Source Softw. 3 638
Rashid B and Calhoun V 2020 Towards a brain-based predictome of mental illness Human Brain Mapp. 41 5468–535
Rihs T A, Michel C M and Thut G 2007 Mechanisms of selective attention in healthy volunteers: evidence from source localization of human brain oscillations using an auditory oddball task J. Neurosci. 27 499–515
Ros T et al 9 2017 Neurofeedback tunes scale-free dynamics in spontaneous brain activity Cereb. Cortex 27 4911–22
Scholz C 2019 Nonlinear measures for dynamical systems (Nolds) (available at: https://doi.org/10.5281/zenodo.3814723)
Shankman S A, Silverstein S M, Williams L M, Hopkinson P J, Kemp A H, Felmingham K L, Bryant R A McFarlane A and Clark C R 2008 Resting electroencephalogram asymmetry and posttraumatic stress disorder J. Traumatic Stress 21 190–8
Shvil E, Rusch H L, Sullivan G M and Neria Y 2013 Neural, psychophysiological and behavioral markers of fear processing in PTSD: a review of the literature Curr. Psychiatry Rep. 15 358
Stam C J, Nolte G and Daffertshofer A 2007 Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources Human Brain Mapp. 28 1178–93
Steenkamp M M, Litz B T, Hoge C W and Marmar C R 2015 Psychotherapy for military–related PTSD: a review of randomized clinical trials JAMA 314 489–500
Sutton S K and Davidson R J 1997 Prefrontal brain asymmetry: a biological substrate of the behavioral approach and inhibition systems Psychol. Sci. 8 204–10
Terpou B A, Shaw S B, Thieberg J, Victor F, Michel C M, McKinnon M C, Lanius R A and Ros T 2022 Spectral decomposition of EEG microstates in post-traumatic stress disorder Neuroimage Clin. 35 103135
Tibshirani R, Walther G and Hastie T 2001 Estimating the number of clusters in a data set via the gap statistic J. R. Stat. Soc. B 63 411–23
Todder D, Levine J, Abujamah A, Mater M, Cohen H and Kaplan Z 2012 The quantitative electroencephalogram and the low-resolution electrical tomographic analysis in posttraumatic stress disorder Clin. EEG Neurosci. 43 48–53
Toll R T et al 2020 An electroencephalography connectomic profile of posttraumatic stress disorder Am. J. Psychiatry 177 233–43
Tsimeruso 2019 Pysparcl (available at: https://github.com/tsurumeso/pysparcl)
Uddin L Q, Yeo B T and Spreng R N 2019 Towards a universal taxonomy of macro-scale functional human brain networks Brain Topogr. 32 926–42
van der Maaten L and Hinton G 2008 Visualizing data using t-SNE J. Mach. Learn. Res. 9 2579–605 (available at: www.jmlr.org/papers/volume9/vandermaten08a/vandermaten08a.pdf)
Veltmeyer M D, McFarlane A C and Bryant R A 2006 Integrative assessment of brain function in PTSD: brain stability and working memory J. Integ. Neurosci. 5 123–38
Vinck M, Oostenveld R, Van Wingerden M, Battaglia F and Pennartz C M 2011 An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias NeuroImage 55 1548–65
von Wegner F and Lauß H 2018 Information-theoretical analysis of EEG microstate sequences in python Front. Neurinf. 12 1–10
Vossel S, Geng J J and Fink G R 7 2014 Dorsal and ventral attention systems: distinct neural circuits but collaborative roles Neuroscientist 20 150–9
Wahlbøe H and Oken B S 5 2013 Peak high-frequency HRV and peak alpha frequency higher in PTSD Appl. Psychophysiol. Biofeedback 38 57–69
Weathers F W, Litz B T, Herman D S, Huska J A and Keane T M 1993 The PTSD checklist (PCL): reliability, validity and diagnostic utility Annual Convention of the Int. Society for Traumatic Stress Studies (San Antonio, TX) vol 2 pp 90–92
Witten D M and Tibshirani R 2010 A framework for feature selection Ann. Stat. 105 713–26
Woo C W, Chang I J, Lindquist M A and Wager T D 2017 Building better biomarkers: brain models in translational neuroimaging Nat. Neurosci. 20 365–77
Yuan H, Phillips R, Wong C K, Zotev V, Misaki M, Wurfel B, Krueger F, Feldner M and Bodurka J 2018 Tracking resting state connectivity dynamics in veterans with PTSD NeuroImage Clin. 19 260–70
Zandvakili A, Philip N S, Jones S R, Tyrka A R, Greenberg B D and Carpenter L L 6 2019 Use of machine learning in predicting clinical response to transcranial magnetic stimulation in comorbid posttraumatic stress disorder and major depression: a resting state electrophenoccephalography study J. Affective Disorders 252 47–54
Zhang Y et al 2021 Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography Nat. Biomed. Eng. 5 309–23