A machine learning framework for analysis of resting-state EEG in patients

Li, Qianliang

Publication date:
2022

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

Ph.D. Thesis

Doctor of Philosophy

DTU Compute
Department of Applied Mathematics and Computer Science

A machine learning framework for analysis of resting-state EEG in patients

Qianliang Li

Kongens Lyngby 2022

DTU

DTU Compute

Department of Applied Mathematics and Computer Science

Technical University of Denmark

Richard Petersens Plads

Building 321

2800 Kongens Lyngby, Denmark

Phone +45 4525 3031

compute@compute.dtu.dk

www.compute.dtu.dk

Summary

The neural mechanisms underlying most psychiatric disorders remain unclear. Many researchers have employed neuroimaging to investigate the neural differences associated with the disorders, and electroencephalography (EEG) are often chosen due to being non-invasive, low cost, and relatively easy to implement in the clinic or laboratory. However, most of the identified potential biomarkers have yet to be translated to the clinic and many previous studies were limited due to their focus on finding group-mean differences for specific EEG features, thus whether combinations of multiple EEG features could serve as diagnostic biomarkers remain unknown.

This thesis presents a machine learning framework developed for analysis of resting-state EEG for biomarker discovery in patients. We implemented EEG signal processing, source localization, computation of an extensive set of commonly utilized EEG features, and unsupervised and supervised machine learning algorithms for dimensionality reduction, feature selection, clustering and predictive modelling. The framework was applied to data from combat-exposed veterans with post-traumatic stress disorder (PTSD) and adults with autism spectrum disorder (ASD).

In the PTSD study, we observed significant group-mean differences in some of the spectral EEG features, and the classifier was able to classify the PTSD group with up to 63% balanced test accuracy. Interestingly, clustering the PTSD group into two distinct subtypes revealed one subtype with functional connectivity relatively similar to the combat-exposed control group without PTSD and another subtype with prominent hyperconnectivity. Our classifiers trained to classify each of the subtypes against the control group did not obtain better performance on the subtype with relatively normal connectivity, but the classification of the subtype with hyperconnectivity improved up to 79% balanced test accuracy. Additionally, many of the connectivity features utilized by the classifier trained to classify the subtype with hyperconnectivity were positively

correlated with arousal severity scores, one of the central symptom clusters of PTSD, and the subtype with hyperconnectivity had greater arousal scores compared to not only the control group, but also the other subtype.

In the ASD study, we observed that the adults with ASD had EEG activity patterns within the typical range of the non-autistic comparison group, with no significant group-mean or group-variance differences for any of the EEG features and the best classifier merely obtained 56% balanced test accuracy. We also identified two ASD subtypes, but were unable to derive a clinically meaningful interpretation of the subtypes.

Taken together, the novel framework presented in this thesis has been demonstrated in two clinical EEG datasets and as it can readily be expanded to other datasets and disorders, we hope the framework can serve as a stepping-stone for future studies and may pave the way for better identification of quantifiable biomarkers in resting-state EEG.

Resumé

De underliggende neurale mekanismer bag de fleste psykiske sygdomme er stadigvæk ukendte. Mange forskere har prøvet at bruge neurovidenskabelige billeddannelsesmetoder i søgningen efter neurale ændringer associeret med psykiske sygdomme og elektroencefalografi (EEG) er ofte brugt, fordi det er en non-invasiv og billig metode som er relativt nemt at implementere i klinikken eller laboratoriet. Ikke desto mindre er de fleste potentielle biomarkører ikke blevet realiseret til praktisk brug i klinikken. Mange af de tidligere studier har været begrænset af deres fokus på at finde forskelle på gruppe niveau for bestemte EEG variabler, og har derfor ikke undersøgt om kombinationer af forskellige EEG variabler kan fungere som diagnostiske biomarkører.

I denne afhandling præsenterer vi en maskinlæringsmetode som er udviklet med henblik på at finde biomarkører i patienter. Vi implementerede EEG signalbehandling, kilde lokalisering, udregning af et stort antal ofte brugte EEG variabler og uovervåget og overvåget maskinlæringsalgoritmer for dimensionalitetsreduktion, selektion of variabler, klyngeanalyse og prædiktiv analyse. Vores analysemetode blev anvendt på data fra krigsveteraner med posttraumatisk stressforstyrrelse (PTSD) og voksne med autismespektrumforstyrrelse (ASD).

I PTSD studiet observerede vi en signifikant forskel mellem grupper i nogle af de spektrale EEG variabler og vores klassificeringsmodeller opnåede op til 63% balanceret test nøjagtighed for prædiktion af PTSD gruppen. Klyngeanalysen af PTSD gruppen fandt to tydeligt forskellige undergrupper. Den ene undergruppe havde normal funktionel konnetivitet, hvorimod den anden var karakteriseret af stærkere funktionel konnetivitet. Vores klassificeringsmodeller var ikke bedre til at prædiktere undergruppen med normal funktionel konnetivitet fra kontrol gruppen, men undergruppen med stærkere funktionel konnetivitet var lettere at prædiktere og blev klassificeret med 79% balanceret test nøjagtighed. Mange af de funktionelle konnetiviteter som blev anset for at være vigtige

for prædiktionen var også positivt korreleret med øget alarmberedskab, en af de centrale symptomklynger for PTSD, og undergruppen med stærkere funktionel konnetivitet havde også øget alarmberedskab i forhold til ikke kun kontrolgruppen, men også den anden undergruppe.

I ASD studiet observerede vi at de voksne med ASD havde normale EEG aktivitetsmønstre. Der var ikke nogen forskelle i gruppe gennemsnit eller varians for de forskellige EEG variabler og den bedste klassificeringsmodel opnåede kun 56% balanceret test nøjagtighed. Vi prøvede også klyngeanalyse på ASD gruppen, men fandt ikke nogle klinisk relevante undergrupper.

Alt i alt har vi i denne afhandling demonstreret hvorledes den nye analysemetode vi har udviklet kan anvendes på to kliniske EEG datasæt og analysemetoden kan let blive udvidet til andre datasæt eller sygdomme. Vi håber vores analysemetode kan være en trædesten for fremtidige studier og bane vejen for bedre identifikation af kvantitative EEG biomarkører.

Preface

This thesis was prepared at the Section for Cognitive Systems, Department of Applied Mathematics and Computer Science at the Technical University of Denmark in fulfillment of the requirements for acquiring a Ph.D.-degree. It was conducted under guidance of the supervisors Assoc. Prof. Tobias Andersen and Assoc. Prof. Ivana Konvalinka, from the section of Cognitive Systems at DTU Compute.

The thesis presents the machine learning framework we developed for analysis of resting-state EEG, along with an introduction to the underlying theory. The framework was applied to discover biomarkers for post-traumatic stress disorder and autism spectrum disorder and the key findings of the two studies are presented. Manuscripts of the two studies are included in the appendix. The project started in September 2019 and ended in August 2022.

Qianliang Li

Kongens Lyngby, August 31, 2022

Acknowledgements

I would like to express my deepest gratitude and thank my supervisors Tobias Andersen and Ivana Konvalinka. Embarking on the road of machine learning and computational modelling as a student with a biomedical background was a daunting task and I would definitely have gotten lost without their guidance and encouragements. Thanks for all your enthusiasm, patience and tireless support! Special thanks to Tobias for his creative ideas, in-depth mathematical explanations, and availability for great academic sparring and to Ivana for her inspiration, attention to details and great sense of humour.

I would also like to thank Klaus Linkenkaer-Hansen, for hosting me during my external stay at Integrative Neurophysiology lab (INF), Vrije Universiteit Amsterdam, and for his passionate support and inputs. A big thanks to everyone at INF who warmly welcomed me and provided an excellent research environment, which made my external stay a wonderful experience in the midst of the COVID-19 pandemic. Another big thanks goes to all my fellow researchers at the section for Cognitive Systems at DTU.

Last but not least, I would like to thank my family and friends for their moral support and for keeping me grounded in our three-dimensional reality, while I was bewilderingly exploring the complex high-dimensional space of machine learning.

List of abbreviations

AQ	Autism-spectrum quotient short questionnaire
ASD	Autism spectrum disorder
Coh	Coherence
CV	Cross-validation
DFA	Detrended fluctuation analysis
DSM	The Diagnostic and Statistical Manual of Mental Disorders
E/I	Excitation/inhibition
EEG	Electroencephalography
fMRI	Functional magnetic resonance imaging
GC	Granger causality
GEV	Global explained variance
ICA	Independent component analysis
ICD	International Statistical Classification of Diseases and Related Health Problems
Imcoh	Imaginary part of coherence
LRTC	Long-range temporal correlations
MEG	Magnetoencephalography
PCA	Principal component analysis

PCL	Post-traumatic stress disorder checklist
PEC	Power envelope correlations
PET	Positron emission tomography
PLV	Phase locking value
PTSD	Post-traumatic stress disorder
RDoC	Research domain criteria
SPQ	Sensory perception quotient short questionnaire
SVM	Support vector machine
t-SNE	t-distributed stochastic neighbor embedding
wPLI	Weighted phase lag index

Contents

CHAPTER 1

Introduction

Most psychiatric disorders are currently diagnosed based on subjective self-reports of specific symptoms. The diagnosis is given if the right combinations of specific symptoms and their severity are above clinically defined thresholds, e.g. as described in the newest version of The Diagnostic and Statistical Manual of Mental Disorders (DSM-5 [1]) or International Statistical Classification of Diseases and Related Health Problems (ICD-11 [2]). While the DSM and ICD have facilitated reliable clinical diagnosis, emerging problems have also been observed. Most noticeably, the diagnostic categories show a heavy overlap of behavioral symptoms and biological features, e.g. high genetic correlation between psychiatric disorders [3], and high heterogeneity is also observed within each disorder [4]. These problems have spurred the development of the Research Domain Criteria (RDoC), which is a dimensional framework based on classification of mental disorders based on empirical data from neuroscience and genetics [5]. The hope is that data-driven classifications would better reflect the pathophysiological mechanisms underlying the diseases and attenuate the high heterogeneity often observed in psychiatric disorders [6]. Characterizing diseases in a way that is better aligned with the disease mechanisms, would potentially increase treatment efficiency through more optimized treatment strategy choices. The trend towards moving away from self-reports and using data-driven methods to arrive at better classifications of clinical groups has also been labeled as precision medicine, which ultimately tries to tailor the diagnosis and treatment for each specific individual based on quantifiable biomarkers [6–9].

In this thesis we focus on two disorders defined by DSM: post-traumatic stress disorder (PTSD), a chronic disorder characterized by stress symptoms following a traumatic event, and autism spectrum disorder (ASD), a developmental condition characterized by problems with social interactions, restricted interests and repetitive behaviors. Previous studies have tried to discover biomarkers and investigated the neurophysiological mechanisms behind these two disorders using resting-state electroencephalography (EEG),

but the results have been i) inconsistent with no clear consensus for either condition; ii) limited due to low sample sizes; iii) limited to small sets of pre-selected EEG features (reviewed in [10]). Therefore, we set out to perform an exploratory analysis and estimated an extensive set of EEG features reflecting different aspects of brain activity patterns in relatively large sample sizes of around 200 participants for both disorders. Our overall goal was to determine if any of the EEG features or combinations of features could serve as clinical biomarkers and bring us one step closer to precision medicine.

1.1 Thesis aim and scope

The aim of this thesis was to develop a machine learning framework for discovery of EEG biomarkers for clinical purposes. Specifically, we applied the framework to a PTSD and ASD dataset, in order to identify potential biomarkers for each condition.

The thesis is organized as follows:

Chapter 2 briefly introduces background theory within EEG and machine learning.

Chapter 3 is a method section on the analysis framework we developed.

Chapter 4 describes the results on applying the framework on EEG data from PTSD patients.

Chapter 5 describes the results on applying the framework on EEG data from adults with ASD.

Chapter 6 summarizes the findings and reflects on the outlook on future research.

CHAPTER 2

Background theory

The topic of this thesis is inter-disciplinary and includes concepts and terminology from neuroimaging, health science and computer science. In order to accommodate readers from different backgrounds, this chapter will cover some basic background theory on the core concepts of EEG and machine learning.

2.1 Origin of EEG

The neurons in our brains communicate with each other through electrical and chemical signals and form complex neural networks that are crucial for the functioning of the brain. There are multiple neuroimaging modalities that tries to tap into this communication, e.g. functional magnetic resonance imaging (fMRI), positron emission tomography (PET), and EEG and magnetoencephalography (MEG). They all measure different aspects of the neuronal communication and there are no perfect method, as different methods have their pros and cons.

This thesis focuses on EEG, which is a method for recording brain activity through measuring electrical potential differences between electrodes placed along the scalp. The method has been used in humans for almost a century, with Hans Berger being known as the first person to record EEG in humans in the late 1920s [11]. Compared to the other neuroimaging methods, EEG is known to have good temporal resolution, in the order of milliseconds, and good clinical practicality, due to being non-invasive, low cost, mobile and relatively easy to implement. The recording schemes most often employed are task-specific, e.g. recording EEG during a visual or auditory task, or resting-state, i.e. individuals are instructed to either have their eyes closed or open and fixated on a cross and told to not fall asleep, but otherwise their minds are free to wander. Due to the lack of specific instructions, resting-state recordings are seen as more uncontrolled

and variable compared to task-based EEG [12], however the simple instructions and lack of a task also makes it more straightforward to apply in the clinic. The clinical datasets analyzed in this thesis are also resting-state EEG, as potential resting-state EEG biomarkers are more appealing and less resource demanding for clinical implementation.

Besides the advantages of EEG, it is also important to be aware of its limitations. The primary disadvantage of EEG is the relatively poor spatial resolution. This is partly because the electrical signal from each neuron has to pass through the skull in order to reach the electrode, which creates a smearing effect, and partly because the electrical activity from each neuron is very tiny. Therefore, what the electrode is actually measuring is the summation of electrical currents from multiple neurons, also referred to as compound potentials. Additionally, EEG is biased towards cells with the same orientation, as their electrical fields are less likely to cancel each other out, and also biased towards the cells closest to the electrodes, hence the pyramidal cortical neurons are thought to be the primary contributors for the EEG activity [13, 14].

To improve the spatial resolution of EEG, algorithms that map the electrode signals on the scalp to their sources in the brain have been developed. These source localization methods try to solve an inverse problem. The neurons inside the brain generates the electrical signal measured by the electrodes, which is modelled by a forward model, and using this forward model it is possible to obtain an inverse model that tries to unmix and estimate which sources could have generated the signal that is measured by the electrodes. This is a very challenging procedure, as there are likely many more sources than electrodes and many different combinations of sources can produce the same signal measured by the scalp electrodes, thus a priori assumptions, based on mathematical, neurophysiological, neurobiological and anatomical knowledge, are needed, and this have given rise to different source localization algorithms (reviewed in [15, 16]).

The main benefit of source localization is that the signal becomes more neurophysiologically relevant. The electrical activity measured by one single electrode is a mixture of many different underlying sources, which makes the interpretation challenging. However, by narrowing down which brain areas are responsible for the signal, the signal becomes clearer and easier to relate to previous findings within neuroscience. The unmixing of signal is also helpful for attenuating the problem of volume conduction [16], which refers to how electric fields transmit through biological tissue and thus the same source might

be captured by two neighboring electrodes. Besides the benefits, it is also important to be aware of the drawbacks of source localization. Due to the inverse problem being underdetermined, there are no perfect unique solution and the quality of the source localization is influenced by how clean the EEG data is and the number of electrodes [16].

2.2 EEG analysis

There are many ways to analyze the EEG signal, with the most straightforward being qualitative visual inspection of the time series. Figure 2.1 shows an example of how a resting-state EEG signal looks. Experts trained on looking at these EEG time series can spot patterns of brain activity associated with different behavior, e.g. muscle movements, eye movements, hyperventilation, epilepsy and different sleep stages [11].

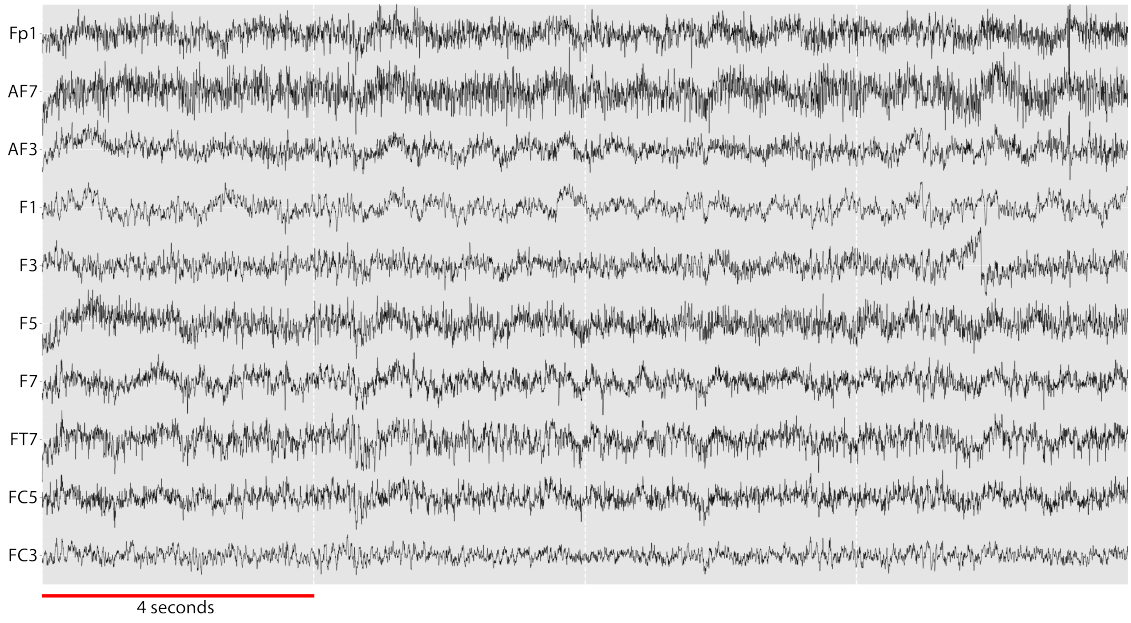

Figure 2.1: Example of an EEG signal. Snapshot of a preprocessed EEG signal measured at 10 frontal electrodes over the left brain hemisphere.

Besides qualitative analysis, it is also possible to quantify and compute different features that characterize the EEG signal. As evident from Figure 2.1, EEG signals oscillate with different rhythms and the most commonly applied quantification of EEG is spectral analysis, where the power at different frequencies are estimated. These rhythms have

been found to be important for neuronal communication, with different roles attributed to different frequency bands [17]. The most common categorization of EEG oscillations in humans are delta (<4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma oscillations (>30 Hz). Two transformations are often applied to the power features; absolute power refers to power converted to a decibels (dB) scale by taking the log base 10 and multiplying with 10, and relative power is computed as proportion of power in each frequency band normalized to the total power in all frequency bands.

The most predominant oscillation during rest is alpha, which is thought to be associated with an "idle" state, where the brain is ready to process information. Alpha activity are often seen correlated with inhibitory signaling and could serve to suppress task-irrelevant processes and play a role as an attentional mechanism in a feedback/top-down direction [17–19]. Gamma oscillations are thought to modulate excitation of post-synaptic neurons, as inputs that are synchronized with fast gamma oscillations are more likely to trigger action potentials, and have been found important for attending visual stimuli in a feedforward/bottom-up direction [17, 20]. Theta oscillations have also been found important for attention, and have been thought to play a role in resetting gamma synchronizations, thus enabling a shift in attention [17] and theta power have also been found positively correlated with the ability to encode new information [21]. Delta and beta oscillations are less studied, with delta activity being most prominent during sleep [22] and beta activity have been associated with maintaining the status quo [23]. Although the role of the different frequencies have been investigated for many years, there are still much debate and uncertainty about their roles, e.g. despite delta being commonly known as a characteristic of slow wave sleep, delta activity has also been found in association with cognitive tasks. This contradiction has spurred the proposal that the functional role of delta oscillations is brain state specific. Evidence supporting this notion has been found from PET studies measuring metabolism in parallel with EEG, which showed positive correlation between awake delta and PET metabolism [24] and negative correlations between delta of slow wave sleep and PET metabolism [25]. Delta oscillations have also been postulated to modulate attention, due to its presence in the attention networks during mental tasks requiring internal concentration [22]. These examples are just a few out of many findings suggesting new roles for different frequency bands and the investigation of the roles of brain oscillations is an active research field [26].

Besides estimating power in each frequency band, the ratios between different frequency bands have also been associated with brain functions, e.g. frontal theta/beta ratio have been negatively correlated with cognitive control of attention and greater theta/beta ratio are often observed in people with attention deficit hyperactive disorder [27]. Ratios of power between brain hemispheres have also been investigated, e.g. frontal alpha asymmetry (right−left) have been found to be negatively correlated with greater cognitive vulnerability for depression [28] and observed to be lower in depressed patients [29]. It is thought that greater frontal alpha asymmetry is correlated with positive emotional experiences and approach, whereas lower frontal alpha asymmetry is associated with negative emotional states and withdrawal [30].

The specific frequency for where the peak in the alpha frequency band occurs and the 1/f power law relation with spectral power have also been studied and both features have been shown to decrease with age [31–36]. Additionally, peak alpha frequency have been observed to increase during cognitive tasks and this increase is dependent on the intensity of the task [31, 32]. The 1/f power law exponent have also been correlated with physiology [34], e.g. negatively correlated with excitation/inhibition (E/I) ratio [37]. A recent study also employed critical brain dynamics and described a method to estimate functional E/I ratio (fEI) using EEG and found that children with ASD had elevated fEI [38].

The EEG features described up until now have all been univariate metrics, i.e. they are estimated for each electrode or brain source. Cognitive and perceptual tasks require neuronal communication and information flow through many different functionally specialized brain regions, and it is believed that synchronization of the rhythmic oscillations at different brain areas might facilitate this neuronal communication. To quantify this synchronization, metrics that characterize the bivariate correlations between two electrodes or brain sources have been developed and these are collectively called functional connectivity features (reviewed in [39, 40]). Many different connectivity features have been developed, with the most widely used being coherence (Coh; [41]). Coh can be viewed as the frequency-domain analogue to Pearson's correlation in the time domain, as both measure the linear dependency of two signals. However, it has been argued that Coh is susceptible to noise due to the cross-spectral density, that is used to estimate Coh, contains information about both the amplitudes and the phase difference between the two signals of interest. When looking at the synchronization between two signals, using

only the relative phase information and disregarding the amplitudes might be better at capturing the underlying neural activity and this idea was developed into phase-locking value (PLV; [42]).

Another noise component to take into account is volume conduction, which is particularly problematic when estimating bivariate functional connectivity metrics, that characterize the synchronization of neuronal oscillations between different areas [39]. If the same source is measured in two neighboring electrodes, then the connectivity between those two electrodes would be artificially increased, thus source localization, which tries to unmix the underlying sources, is especially beneficial for functional connectivity analysis [16]. To further reduce the effect of volume conduction, alterations have also been developed to the functional connectivity features themselves to correct for the effect of volume conduction. Imaginary coherence (Imcoh; [43]), weighted Phase Lag Index (wPLI; [44]) and orthogonalized Power Envelope Correlations (PEC; [45, 46]) are examples of connectivity features that removes the zero-phase lag correlations associated with volume conduction. Additionally, some functional connectivity measurements also estimate the direction of information flow, e.g. Granger Causality (GC; [47–49].

So far the EEG features mentioned have primarily been static, although some of the features could be altered to also measure the dynamic changes over time, e.g. using sliding windows [50]. However, there are also EEG features that was developed specifically to measure the temporal dynamical changes of EEG time series. Long-range temporal correlations (LRTC) have been observed in humans and thought to indicate that neuronal networks are in a state of criticality, which is optimal for a network to adapt to changes [51, 52]. LRTC have been observed to increase during neurodevelopment from childhood into early adulthood [53], and attenuated in ASD [54] and PTSD [55]. Transiently stable global scalp field maps of varying durations, coined microstates, have also been observed in resting-state EEG [56] and are thought to be generated by coordinated activity of different functional neural networks. Remarkably, four consistent microstates can often be replicated across different resting-state EEG studies and they explain a large portion of the total topographic variance [57]. The characteristics of these microstates, e.g. mean duration, time coverage and occurrence, have been found to be disrupted in schizophrenia [58], ASD [59], and many other neuropsychiatric diseases [57].

All in all, a wide range of EEG features have been developed and employed over the years,

each with their own pros and cons. Most studies have focused on one specific feature type, e.g. spectral power or one connectivity metric, as including too many features and working in a high-dimensional space is, intuitively, very complicated for humans. Investigating many features also means performing many statistical tests if the goal is to look for group-mean differences, and this requires multiple testing correction, which decreases the statistical power, i.e. the probability of true positives decreases in order to prevent an inflation of false positives. Thus, many previous studies have investigated resting-state EEG with a hypothesis-driven approach and limited themselves to a select set of handpicked features. The drawback of such an approach is that many aspects of the EEG time series are lost and the potential of whether combinations of EEG features could serve as biomarkers are not considered. One approach to circumvent the problems of working with many features is to utilize machine learning models, that excel at finding patterns in complicated high-dimensional data [8, 60, 61]. However, working with an excessive amount of data is not without its own problems; the high-dimensionality of the data is a double-edged sword, often referred to as the blessings and curses of dimensionality [62, 63]. Data tend to be sparse in high dimensions, and the more variables a model have, the more complex and flexible it becomes, and the risk of overfitting to the training data increases. Dimensionality reduction methods such as feature selection that tries to select an optimal small subset of the most informative features can alleviate the problems [63, 64], and will be further elaborated in Section 3.4.

2.3 Machine learning

Machine learning is the field within computer science that works with algorithms that learn from data. Compared to conventional statistics and modelling that are based on predefined hypotheses about specific statistical relationships, e.g. that variable A is linearly correlated with variable B, machine learning models will learn the relationships by themselves directly from data [65]. This data-centric approach of machine learning is highly suitable for exploratory analysis and befitting the current technological trend, with ever-increasing amount of data being generated every day. This trend is aptly referred to as the era of Big Data, and while it started in the computer science field, the trend is also present in biomedical [60] and neuroimaging [66] data, which are the types

of data covered in this thesis.

Broadly, machine learning can be divided into three paradigms: supervised, unsupervised and reinforcement learning [67]. As reinforcement learning was not employed in this thesis, we will instead focus on supervised and unsupervised learning. In unsupervised learning, there are no labels, and the goal of the algorithms is to look for patterns inherent in the data. Examples of some of the common unsupervised models we employed are t-distributed Stochastic Neighbor Embedding (t-SNE), K-means clustering, principal component analysis (PCA), and independent component analysis (ICA) [68–71]. In supervised learning, the data contains annotated expert labels, which are used to train the models. Examples of some of the common supervised models we employed are linear and logistic regression, support vector machines (SVM; [72]) and random forest [73].

The overarching goal of supervised learning is predicting the labels of new or future observations [67, 74]. This is different from conventional statistical testing for group-mean differences, as a significant group difference does not imply the feature is necessarily a good predictor and could serve as a diagnostic biomarker. A significant difference only means that on average, one condition have different values for the feature that is being tested, but whether this feature can be used to diagnose which specific participants belongs to which condition also depend on the effect size. Thus machine learning is naturally better applicable for identifying biomarkers, as the predictive performance can directly be translated into individual discrimination capability of the discovered features [7, 8, 75, 76].

Importantly, to estimate how well the predictions worked, it is paramount to use a separate dataset, referred to as the test set, which was not observed by the machine learning model beforehand. Otherwise, the performance metric would be biased towards the dataset, referred to as overfitting to the training set, and not necessarily reflect how well the model would predict new observations. If no separate independent datasets are available, it is possible to divide the dataset into a test and training set beforehand, and only train the machine learning model on the training set. This splitting of data is commonly referred to as cross-validation (CV). Additionally, many machine learning models have hyperparameters, which often controls the complexity of the model. In order to optimize the hyperparameters for an unseen dataset, it is important to further divide the training set in two, and only train on a portion of the data, in order to

validate on the unseen portion of the data, which hyperparameter values should be chosen. This type of CV is called two-layer nested CV. The outer layer splits the data into an inner layer set and the test set. The inner layer is then further split into the training set and the validation set. The models are trained on the training set, and the hyperparameters are chosen based on how well they worked when predicting on the validation set. Finally, after the best model have been found, the model is retrained on all the inner layer data and then the test performance is computed by predicting on the test set [9, 75, 77]. Different ways to split the data can be utilized, so the test performance is less dependent on the exact splitting of the data, e.g. by repeating the whole process but using a different initial split. Figure 2.2 shows an example of 10-by-10 two-layer CV repeated 10 times with different data splits, which is the CV scheme utilized for the PTSD and ASD study described in Chapters 4 and 5.

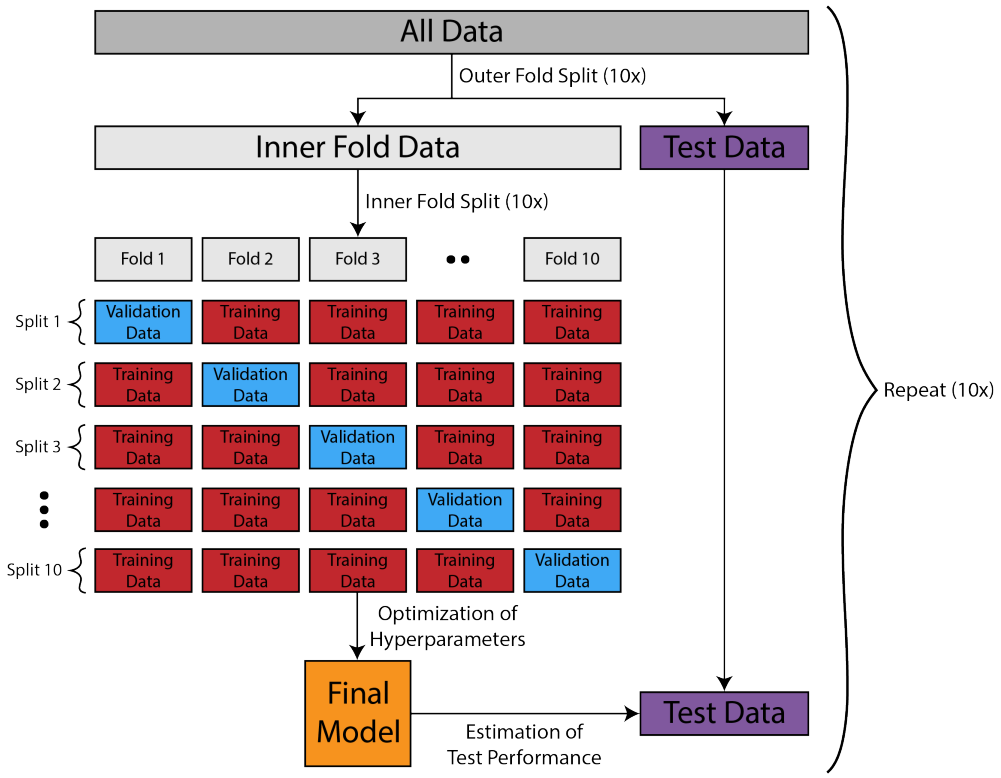

Figure 2.2: Repeated two-layer nested cross-validation. 10 repetitions of 10-by-10 two-layer cross-validation are illustrated here. The outer fold split ensured we tested on unseen data, while the inner fold split alleviates overfitting when we trained the models and optimized model hyperparameters. Figure from Appendix A.

CHAPTER 3

Method

The overall EEG analysis framework can be seen in Figure 3.1 and consists of the following steps: EEG acquisition, preprocessing, feature estimation, cross-validation, feature selection, machine learning model training, hyperparameter tuning and evaluation of the test performance. This chapter will describe the different steps and reflect on our rationale for the methodological choices. There are minor differences to how the framework was applied for the PTSD and ASD datasets, which will be further described in their respective chapters and more details can be found in the journal articles (Appendix A & B). As the EEG acquisition is specific for each dataset, this chapter will start from the preprocessing step.

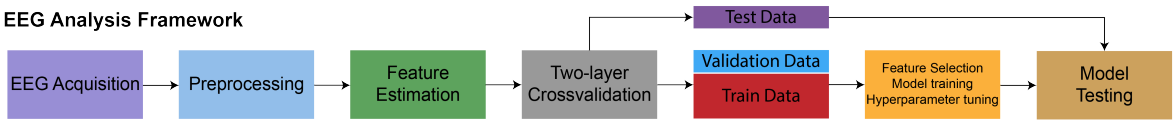

Figure 3.1: Overview of the EEG analysis framework. After EEG acquisition, the data were preprocessed, and commonly used resting-state EEG features were computed. Repeated two-layer nested cross-validation was used to split the data. Feature selection, model training and hyperparameter tuning were performed on the inner fold training and validation set, while the generalization performance was evaluated on the entirely unseen test set.

3.1 EEG preprocessing

"Garbage in, garbage out" is a popular saying within computer science and reflects the notion that if the data quality is poor, then the outputs from the models will also be poor. It highlights the importance of cleaning and quality checking the data, which is

the primary goal for the preprocessing step. Bad segments and recordings filled with artifacts or noise are removed, in order to obtain good reliable feature estimates. We performed the following preprocessing steps:

1. 1–100 Hz bandpass filtering with a Hamming window
2. 50 Hz notch filtering to remove power line noise
3. Downsampling to speed up computation time
4. Epoching into 4 second segments
5. Visual inspection to remove epochs with artifacts and bad channels
6. Re-reference to common average
7. Spherical spline interpolation of bad channels
8. ICA to identify eye blink, eye movement and heartbeat artifacts
9. AutoReject [78] guided final visual inspection

The 1 Hz highpass filtering removes slow drifts, which is beneficial for the ICA [79], while the 100 Hz lowpass filtering removes the high frequency portion of the signal we are not interested in. In Europe you can usually observe 50 Hz power line noise, which we remove with a notch filter. The data was downsampled to speed up computation time and epoched into 4 seconds non-overlapping segments. The 4 seconds were chosen based on what other resting-state EEG studies used, which primarily ranged from 1 to 10 seconds [30, 80–92]. Some of our features, e.g. Granger causality, also assumes stationarity, and while EEG signals are considered non-stationary, if the data is windowed into shorter segments, the segments themselves are more likely to be stationary [48, 49]. However we also did not want too short segments, as the features are more reliable if they are computed from longer time series, as evidenced by the finding that shorter epochs have higher connectivity in general, while with longer epochs the estimated connectivities become more stable [93].

Following epoching, we performed visual inspection and removed bad epochs and channels, however we deliberately left out the ocular and heartbeat artifacts as they would

be corrected later in the preprocessing. To get a robust average re-referencing, we ensured that we only re-referenced after the bad channels were removed [94]. The bad channels were interpolated with spherical spline interpolation [95]. The ICA algorithm tries to unmix the independent components, and in this case we want to remove the ocular artefacts and other non-neuronal sources, e.g. heartbeat artifacts, while retaining the neurophysiological signal. The primary benefit of using ICA is that we retain more data, by correcting the epochs instead of removing them entirely, and this makes ICA especially attractive if we have limited amount of data.

To ensure that ICA worked and to remove any remaining artefacts, we performed a final visual inspection. This visual inspection was guided by the AutoReject algorithm [78], which tries to automatically detect bad epochs and channels by identifying a data-driven peak-to-peak voltage rejection threshold. We did not use AutoReject to directly reject any signal, but instead looked at the places that the algorithm marked as bad and manually confirmed or rejected the suggestions by AutoReject. The manual confirmation was important, as we found the algorithm to have a high false positive rate, most likely due to our data being relatively clean from the previous artefact cleaning stages, thus the algorithm had limited amount of clear artefacts to model. Nonetheless, having this final second visual inspection step was still useful, as the ICA was not always able to unmix the eye artefacts and these situations would thus be caught by AutoReject.

3.2 Feature estimation

We estimated many commonly used EEG features in an exploratory fashion to investigate which features or combinations of features could serve as biomarkers. The most commonly computed resting-state EEG feature in the literature is spectral power density, and we computed the power in the 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). The ratios between power in different frequency bands or different brain hemispheres and the power law relationship between spectral power and frequency were also computed. Functional connectivity features with varying degrees of correction for volume conduction [39] were also estimated, as well as features that captures the temporal dynamics, e.g. microstates [56] and long-range temporal correlations [96]. We also investigated a

measure for functional excitation inhibition ratio [38]. More details on which specific features were estimated and how they are calculated can be found in the journal articles (Appendix A & B).

3.3 Cross-validation

After all the EEG features were estimated, they were combined into a data matrix with number of participants as rows and number of features as columns. Before any machine learning algorithms were applied, repeated 10-by-10 two-layer CV was performed as described in Section 2.3, to ensure the test set was always kept entirely separate from the feature selection, model training and hyperparameter tuning and only used for the final evaluation of generalization performance.

3.4 Feature selection

EEG data contain rich information, which is why we estimated so many different features, however having too many features is also a problem, sometimes referred to as the curse of dimensionality [63]. Many features mean the degrees of freedom are large and the models can be very complex. The more complex a model is, the more likely it is to overfit to noise specific for the dataset, which means its predictions might not perform well on new unseen data [69, 77]. One way to break this curse is to increase the sample size, but this is often very costly and resource demanding for clinical data.

Another way to combat this curse and attenuate overfitting is to reduce the number of features, and the goal of feature selection methods is to estimate which features would be the best to retain. Broadly, feature selection methods can be divided into three categories: filters, wrappers and embedded methods [64]. Filter methods refer to algorithms that select a subset of features prior to applying the machine learning models, e.g. determining which features have the highest correlation with the target variable and then using the top 10 features. Wrapper methods work closely with the specific machine learning models, e.g. sequential forward and backward feature selection are the two most common wrapper methods. In these two methods, the goal is to iteratively search for an optimal feature set by adding/removing features and evaluate how well the machine

learning models work on that particular feature set. Lastly, embedded methods are inherently part of the machine learning models by design. This includes for instance L1 norm regularization or the parameter that that dictates how many features each single tree should be built upon in the random forest classifier.

There are no gold standard for feature selection methods, since each have its own pros and cons. Therefore, we tried many different methods and also combinations of feature selection methods. Specifically, we tried filtering based on F-statistic, PCA and minimum Redundancy - Maximum Relevance (mRMR; [97]). mRMR is an algorithm that tries to minimize the redundancy of selected features, while maximizing the relevance. In practice, an incremental search is performed, similar to sequential forward feature selection, but instead of optimizing the performance of a model, the mRMR criterion is being optimized. At each iterative step, the relevance between the target variable and all available features are computed (e.g. using mutual information) and the redundancy is computed between all available features and the already selected features. The redundancy is then subtracted from the relevance, and the feature with the highest mRMR criterion is chosen and removed from the pool of available features. This iterative process repeats until a desired number of features are chosen. For wrapper methods, we employed sequential forward and backward selection and recursive feature elimination [72]. We also employed both L1 and L2 norm regularization, where applicable. Many of these different feature selection methods can be applied together and the exact details of which we used can be found in the journal articles (Appendix A & B).

The goal of feature selection is to find an optimal feature set, but it is important to keep in mind what it should be optimal for. If you just want to get the best performance on your dataset (i.e. low training error), then having more features will always be better. However, what we are most interested in is a model that is generalizable and able to predict on new data (i.e. low test error). To obtain a good generalization performance, the model and feature selection should also be conducted in such a way that tries to optimize for this scenario. This is why feature selection is performed in the inner fold of the nested two-layer CV.

3.5 Machine learning modelling

We implemented supervised machine learning models for prediction of the clinical labels and scores. There are many different machine learning models that have been used in EEG research (reviewed in [98]) and we employed three commonly used methods: logistic regression, SVM [72] and random forest [73] for the classification tasks and linear regression for prediction of clinical scores. By employing multiple different models, we ensured our results would be less dependent on model choice.

We also employed unsupervised sparse K-means clustering to delineate heterogeneity and identified data-driven subtypes. We adapted the subtyping methodology from Zhang et al. [99]. Briefly, sparse K-means clustering performs joint feature selection and clustering, by employing a lasso-type sparsity constraint s. The gap statistic, which compares the within-cluster dispersion with what is expected under an appropriate reference null distribution, was used to determine the number of clusters and the sparsity constraint [70, 100]. Specifically, we used grid-search and estimated the gap statistic for all combinations of K between 2 to 6 clusters and 20 values for s, equally spaced on a logarithmic scale, ranging from a few features to all features. For each s value, we found the best K using the one-standard-error criterion [70]. After determining the best number of clusters, we determined s as the lowest s within 1 standard deviation of the max gap statistic value [70]. Following the identification of the subtypes, we trained the same supervised models as previously employed, but with the goal of classifying the subtypes against the healthy controls and compared with the classification performances obtained when classifying the whole patient group.

CHAPTER 4

Post-traumatic stress disorder

PTSD is a debilitating psychiatric disorder characterized by the presence of symptoms from four symptom clusters, after exposure to trauma. The four clusters are: 1) intrusion and re-experiencing of the traumatic event, 2) avoidance symptoms, 3) negative alterations of mood and cognition, and 4) alterations in arousal and reactivity [1, 101]. Each cluster contain many different symptoms, and it was calculated that more than 600 000 different symptom combinations could fulfill the diagnostic criteria for PTSD according to DSM-5, clearly highlighting the heterogeneous nature of PTSD [102].

To better understand the neurophysiological mechanisms underlying PTSD, many studies have employed resting-state EEG to investigate if specific brain activity patterns could be used to characterize PTSD. However, the reported findings have been conflicted (Appendix A Supplementary Table A1) and no overall consensus can be drawn [10]. While most resting-state EEG studies on PTSD have focused on identifying group-mean differences, some recent studies also investigated the potential of classifying PTSD with machine learning models trained using 10-fold CV. A study [103] used a combination of source space covariance matrices, band power and network metrics and obtained around 64%, 65% and 62% test accuracies for linear discriminant analysis, SVM and random forest respectively. Interestingly, they also tried a Riemannian based classifier and obtained a noticeable increased performance of 73% test accuracy [103]. Another study [46] investigated source connectivity with PEC, and obtained around 76% balanced test accuracy using a linear relevance vector machine. EEG microstates have also been investigated and a study [104] obtained up to around 76% test accuracy using SVM.

Besides supervised models, unsupervised clustering methods have also been used to search for neurobiologically-driven subtypes to delineate the heterogeneity in PTSD. A study [99] found two subtypes based on PEC, one which was very similar to the control group, while the other group was characterized by decreased frontal and increased posterior beta PEC during the eyes open condition. Interestingly, the subtype that was most discrepant from the controls also showed less clinical improvement after either psychotherapy or antidepressants [99].

Taken together, all of these studies suggest that EEG biomarkers may be useful for PTSD. However, due to inconsistent findings and low sample sizes, there are still no clear consensus about clinically relevant EEG biomarkers for PTSD. Therefore, to disentangle the potential role of EEG biomarkers for characterization of PTSD, we developed a comprehensive machine learning analysis framework to investigate many of the previously reported promising EEG features.

4.1 Methods

The framework was applied to data acquired by the Danish Veteran Centre. To obtain a large sample, the data from three different studies were pooled and after quality checks and EEG preprocessing, the final sample size consisted of 202 combat-exposed veterans (53% with probable PTSD and 47% combat-exposed controls). The probable PTSD patients were defined using the Post-Traumatic Stress Disorder Checklist-Civilian version (PCL-C; [105]), which is a 17-item self-report questionnaire developed to capture the PTSD symptoms as described in the DSM–IV [106]. The PCL-score was computed as the sum of scores from all items and a total PCL $\geq$ 44 [107] was used to define probable PTSD cases. Severity scores for each sub-symptom cluster were also computed, i.e. intrusion score was the sum of items 1 to 5, avoidance score was the sum of items 6 to 12 and arousal score was the sum of items 13 to 17 [105].

The data was obtained in a clinical setting, with each study employing different recording systems. The number of electrodes ranged from 19 to 31 and was sampled at either 250 or 500 Hz. The participants were instructed to sit comfortably in a chair facing a computer display, which gave instructions of when the participants should have their eyes open or closed. The eye conditions were interleaved, with each condition lasting 1 min each and

repeated 5 times, resulting in 5 min of eyes open and 5 min eyes closed resting-state recordings.

The EEG was preprocessed as described in Chapter 3.1. Additionally, due to different electrode configurations for each study, a common electrode set corresponding to the union of the channels across all three studies were defined and spherical spline interpolation [95] was employed to upsample the missing electrodes. All signals were also downsampled to 200 Hz.

The features we investigated could broadly be divided into three categories: 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], power envelope correlations [PEC], and granger causality [GC]), and features that capture the temporal dynamics of EEG (microstates and long-range temporal correlations, computed using detrended fluctuation analysis [DFA exponent]). Details about each feature can be found in the supplementary material in Appendix A.

Specifically for the functional connectivity features, we source localized the EEG time series prior to computing the connectivity. Going to source space alleviates the effect of volume conduction, which functional connectivity is particularly sensitive to [16]. Inspired by [99], we manually modified the Destrieux Atlas [108] parcellation to approximate the 31 ROIs used by Zhang et al. [99] and we also performed subtyping using unsupervised sparse K-means.

4.1.1 Statistical analysis

We performed an exploratory analysis where we computed many different features, and to investigate if there was a group difference, statistical tests were performed for each feature. Thus, there was a need for multiple testing correction to avoid an inflated false positive rate, and we used the false discovery rate (FDR) correction. However, the drawback of multiple testing correction is that the statistical power decreases, i.e. the false negative rate increases. To not be too conservative and be more relatable to other resting-state EEG studies, we only corrected for the number of tests performed within each feature type separately. This does mean, however, that our family-wise error rate will not be 0.05 despite our tests employing a significance level of 0.05. Nonetheless,

this limitation was considered acceptable, as the primary aim of the study was not to identify group-mean differences, but instead to combine the many different features in multivariate models and utilize machine learning to infer which combinations of features could serve as potential biomarkers.

Where applicable, cluster-based permutation tests were employed to alleviate the decreased power when performing multiple comparisons by utilizing the spatial correlation between neighboring EEG electrodes. Instead of testing each electrode separately, clusters with similar brain activities are formed to reduce the number of tests needed [109, 110]. Besides the spatial domain, the clusters can also be formed in the temporal and spectral dimension as neighboring time-points and frequency bands also display correlations. In standard cluster-based permutation tests, a threshold have to be set for defining the clusters. As this threshold is often arbitrary, we instead employed the threshold-free cluster enhancement variant, which tries out all possible thresholds (e.g. with a step size of 0.1), to infer if a given data point belongs to a significant cluster under any of the set of cluster-thresholds [111, 112]. While cluster-based permutation tests (including threshold-free cluster enhancement) are able to control the family-wise error rate, the drawback is that the inference of the test have a lower resolution. Since single data points are not being tested, the inference of the test can only be drawn on the cluster-level, i.e. somewhere in the cluster there was a significant difference between the two groups, but we do not know which specific data points [109].

Both peak alpha frequency and 1/f exponents had NaN values as not all subjects or electrodes have a clear alpha peak or 1/f power law relationship between spectral power and frequency. Therefore, clusters could not be formed for all data points and we performed non-parametric permutation tests instead. For source functional connectivity, we performed t-tests with FDR, due to the lack of a standardized adjacency matrix in source space, as it depends on the specific parcellation and sources are not the the same size, equidistantly separated, or necessarily continuous, making it very complicated to define which sources should be considered adjacent to each other. Additionally, because connectivity is a bivariate metric, the many combinations of connections meant we had to do over 5000 tests for each connectivity feature and more than 100 000 permutations were needed for each test, thus making permutation tests computationally infeasible.

To test the performance of the classifiers, non-parametric Wilcoxon signed-rank tests

were employed on test accuracies. Non-parametric Mann-Whitney U rank tests were used to test the unpaired differences between questionnaire scores between subtypes and group-mean connectivity comparisons. Pearson's correlation was used to test for linear correlations. The significant level was 0.05 for all hypothesis tests. Results are shown as mean with 95% confidence intervals.

4.2 Results

The result section is divided into three subsections. First, the different features are presented and the results of the statistical tests for group-mean differences are summarized. Secondly, the main machine learning modelling results on classification of PTSD are shown. Lastly, the subtyping results are presented.

4.2.1 Group-mean comparisons of PTSD and controls

4.2.1.1 Spectral power

There were no significant differences between the PTSD and control group in absolute power (Figure 4.1; cluster-based permutation test, lowest cluster $p = 0.361$).

For relative power, no significant differences were observed during eyes closed condition (cluster-based permutation test, lowest cluster $p = 0.223$), however we observed that the PTSD group had greater relative beta and gamma power in the eyes open condition. This difference was driven by a cluster spanning from frontal, central and posterior regions in the beta band, and in the gamma band the clusters were primarily in the central and posterior region. (Figure 4.2; cluster-based permutation test, lowest beta cluster $p = 0.007$, lowest gamma cluster $p = 0.006$).

4.2.1.2 Asymmetry

There were no significant differences between the PTSD and control group in the asymmetry of absolute power between right and left hemispheres (Figure 4.3; cluster-based permutation test, lowest cluster $p = 0.095$).

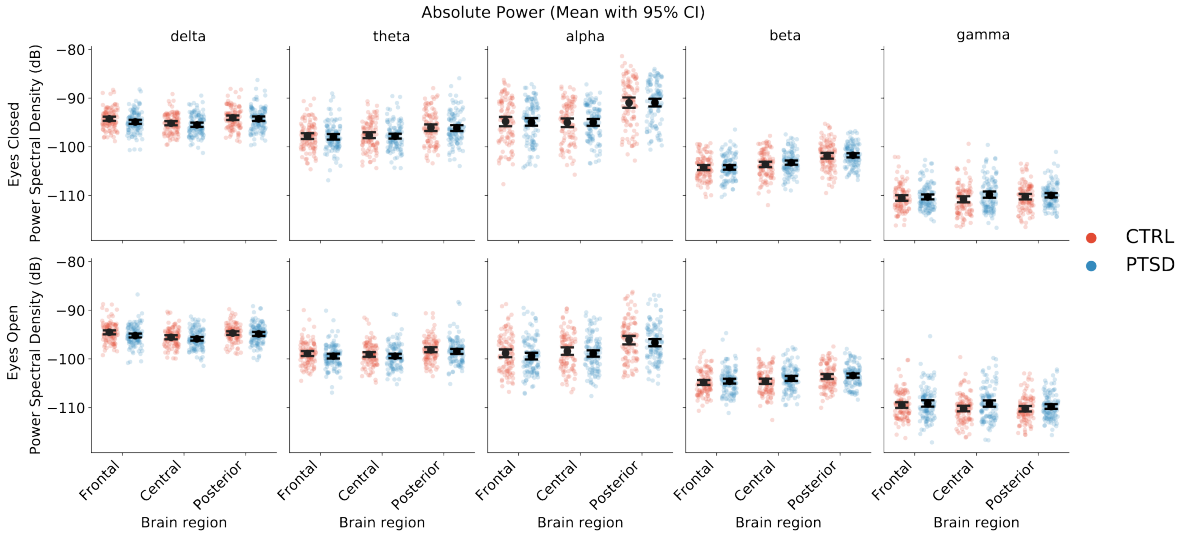

Figure 4.1: No differences in absolute power between PTSD and controls. The absolute power in different frequency bands (columns) for the eyes closed (upper row) and eyes open (lower row) condition averaged over the three brain regions. Mean ± 95% confidence intervals are shown.

4.2.1.3 Frontal theta/beta ratio

During the eyes closed condition, there was no significant difference in frontal theta/beta ratio between PTSD and controls (permutation test, $p = 0.503$), however we observed that the PTSD group had lower frontal theta/beta ratio in the eyes open condition (Figure 4.4; permutation test, $p = 0.038$).

4.2.1.4 Peak alpha frequency

There was a trend towards lower peak alpha frequency in PTSD compared to controls, but the difference was not significant (Figure 4.5; permutation test with FDR correction, lowest eyes closed $p = 0.059$, lowest eyes open $p = 0.085$).

4.2.1.5 1/f exponent

1/f exponents were significantly lower in the PTSD group compared to controls, primarily in the central and some posterior and frontal sensors in both eye conditions (Figure 4.6;

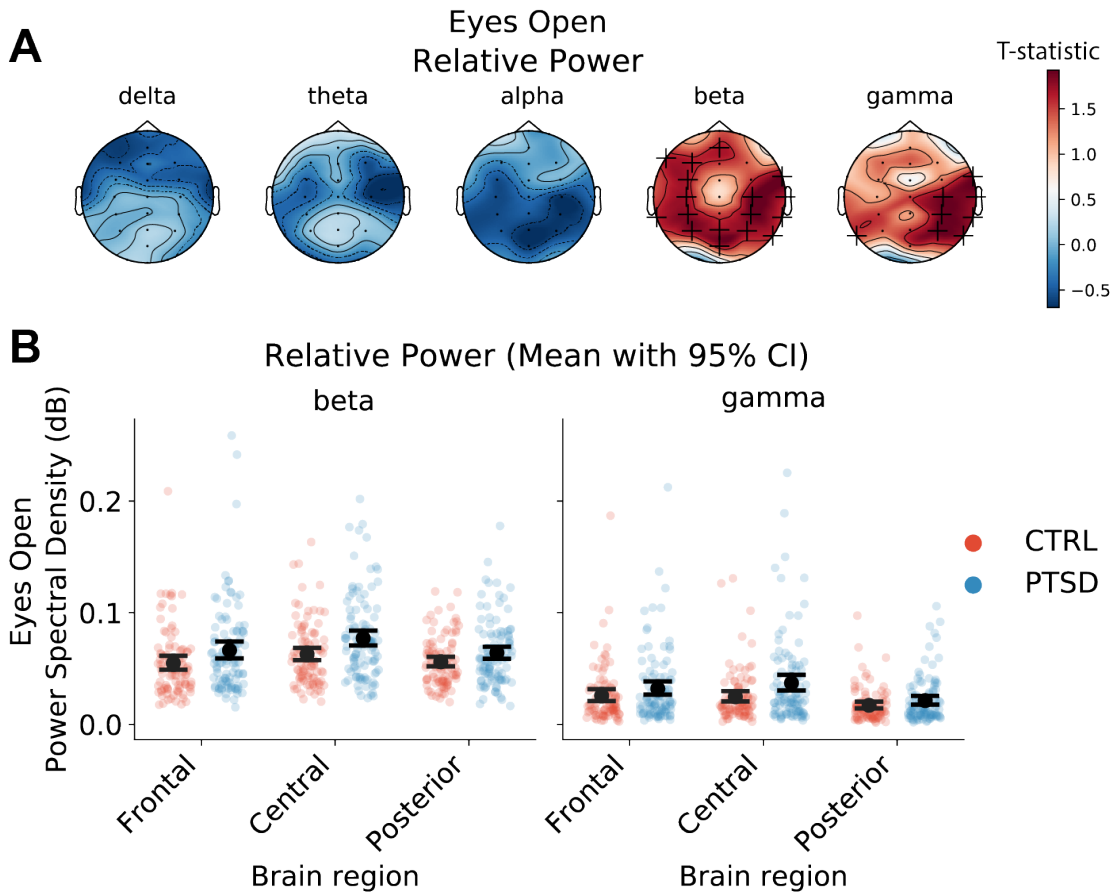

Figure 4.2: Eyes open relative power was significantly higher in the PTSD group. A) Topographic maps of the t-statistics for relative power in the eyes open condition. A plus (+) indicates the electrodes that were part of clusters where a significant effect was found. B) Relative power during eyes open condition in the beta and gamma range averaged over the three brain regions. Mean ± 95% confidence intervals are shown.

permutation test with FDR correction, lowest eyes closed $p = 0.003$, lowest eyes open $p = 0.003$).

4.2.1.6 DFA exponent

There were no significant differences between the PTSD and control group in long-range temporal correlations estimated using DFA exponents (Figure 4.7; cluster-based permutation test, lowest eyes closed cluster $p = 0.488$, lowest eyes open cluster $p = 0.751$).

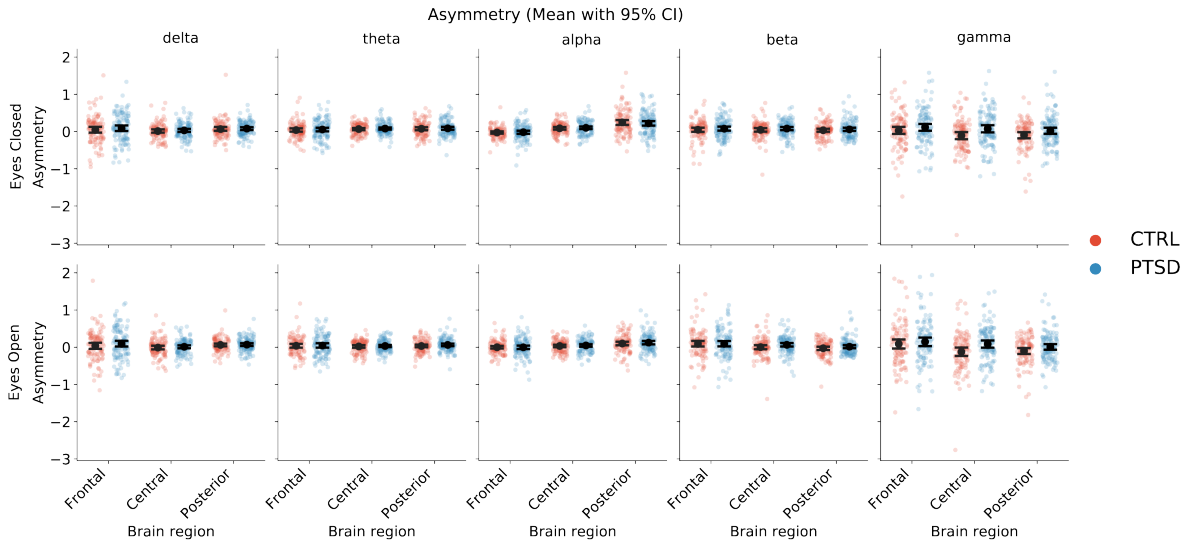

Figure 4.3: No differences in asymmetry between PTSD and controls. The asymmetry in absolute power between right and left hemispheres in different frequency bands (columns) for the eyes closed (upper row) and eyes open (lower row) condition averaged over the three brain regions. Mean ± 95% confidence intervals are shown.

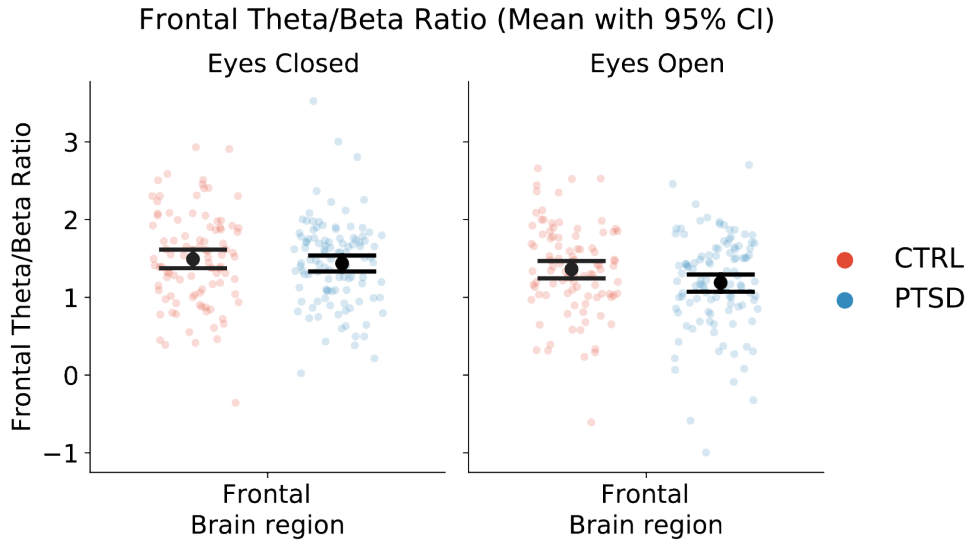

Figure 4.4: Eyes open frontal theta/beta ratio was significantly lower in the PTSD group. The frontal theta/beta ratio for eyes closed and eyes open condition. Mean ± 95% confidence intervals are shown.

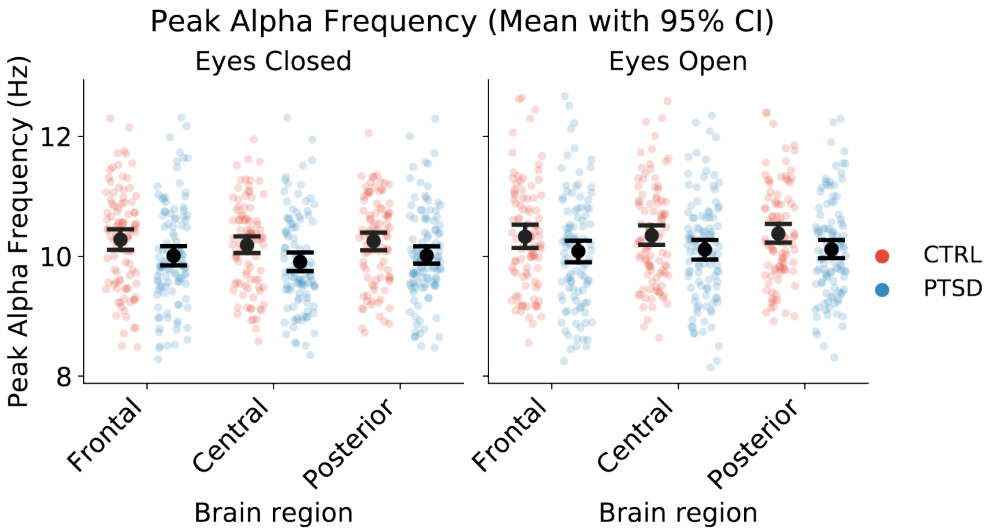

Figure 4.5: No differences in peak alpha frequency between PTSD and controls. The peak alpha frequency for eyes closed and eyes open condition averaged over the three brain regions. Mean ± 95% confidence intervals are shown.

4.2.1.7 Microstates

Four microstates were fitted and in total they explained 57% of the global variance (GEV) for the eyes closed condition (Figure 4.8A). Very similar microstates were found during the eyes open condition, which explained 47% of the global variance (Figure 4.9A). The microstates we found were very similar to the microstates that other resting-state EEG studies have observed in e.g. PTSD [104], schizophrenia [58], autism [113] and healthy participants [114]. No significant differences were observed using permutation tests with FDR corrections for transition probabilities (Figure 4.8B & 4.9B; lowest eyes closed $p = 0.298$, lowest eyes open $p = 0.144$), ratio of time covered by each microstate (Figure 4.8C & 4.9C; lowest eyes closed $p = 0.353$, lowest eyes open $p = 0.171$), and microstate entropy (Figure 4.8D & 4.9D; lowest eyes closed $p = 0.521$, lowest eyes open $p = 0.928$). The entropy was normalized to the maximum entropy obtained by a uniform distribution of microstate labels, hence the closer the entropy is to 1, the more uniformly distributed they are, while an entropy closer to 0 reflect that a single microstate label predominate the time series.

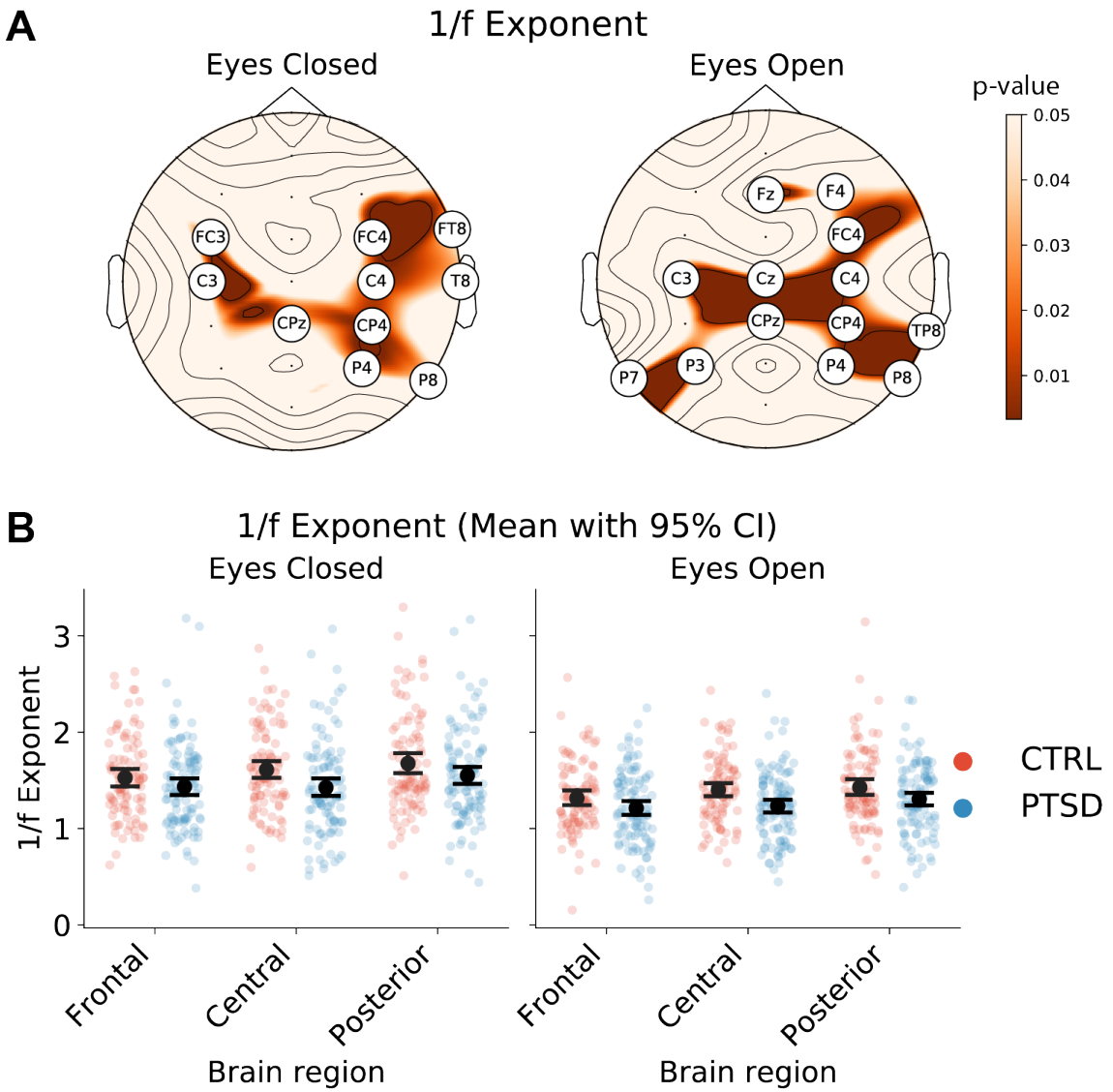

Figure 4.6: The 1/f exponents were significantly lower in the PTSD group. A) Topographic maps of the FDR-corrected permutation test p-values for 1/f exponents. Channels names of the significant sensors are shown. B) 1/f exponents during both eyes open and closed condition averaged over the three brain regions. Mean ± 95% confidence intervals are shown.

4.2.1.8 Source functional connectivity

There were no significant differences between the PTSD and control group in functional connectivity using Student's t-test with FDR correction for imaginary coherence (Figure

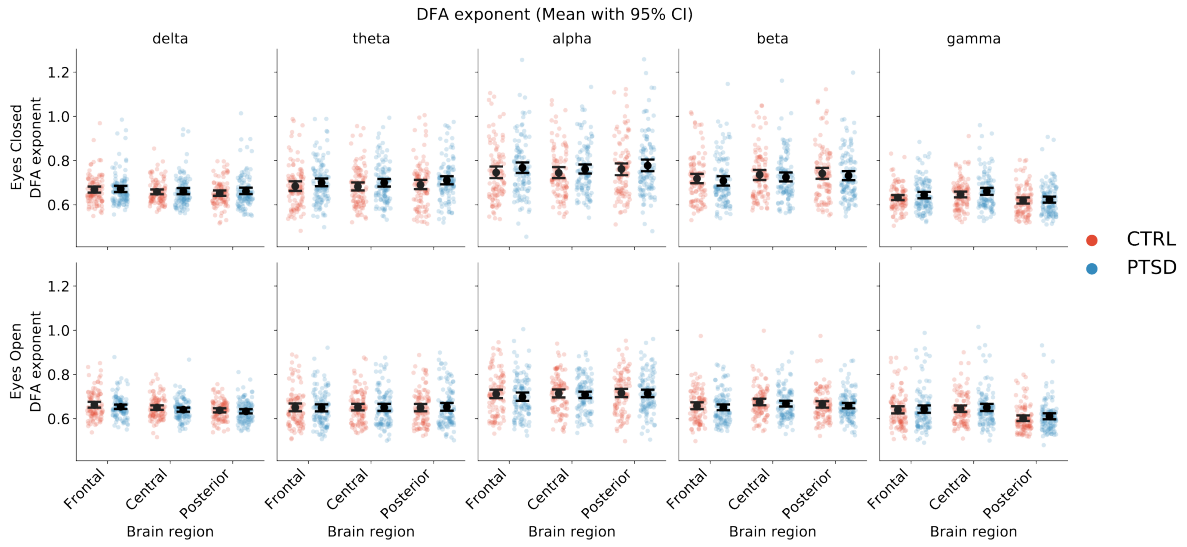

Figure 4.7: No differences in DFA exponents between PTSD and controls. The DFA exponent in different frequency bands (columns) for the eyes closed (upper row) and eyes open (lower row) condition averaged over the three brain regions. Mean ± 95% confidence intervals are shown.

4.10A; lowest eyes closed $p = 0.616$, lowest eyes open cluster $p = 0.181$), weighted phase lag index (Figure 4.10B; lowest eyes closed $p = 0.754$, lowest eyes open cluster $p = 0.186$), power envelope correlations (Figure 4.10C; lowest eyes closed $p = 0.999$, lowest eyes open cluster $p = 0.999$), and Granger causality (Figure 4.10D; lowest eyes closed $p = 0.270$, lowest eyes open cluster $p = 0.497$).

The group-mean results for all the different features are summarized in Table 4.1.

4.2.2 Machine learning predictions of PTSD

Besides exploring the data by looking at group differences for each feature type, we also employed multivariate machine learning models to investigate if combinations of features could serve as EEG biomarkers for PTSD. The main finding of our analysis was that we were able to classify the PTSD group with a balanced test accuracy of 62.9% using SVM (Figure 4.11A), primarily using combinations of connectivity features and 1/f exponents (Figure 4.11B). When we estimated the capability of classifying PTSD for each feature type, we also observed that 1/f exponents, GC, Imcoh and wPLI performed better than chance-level, albeit a bit worse than using combinations of feature types. We

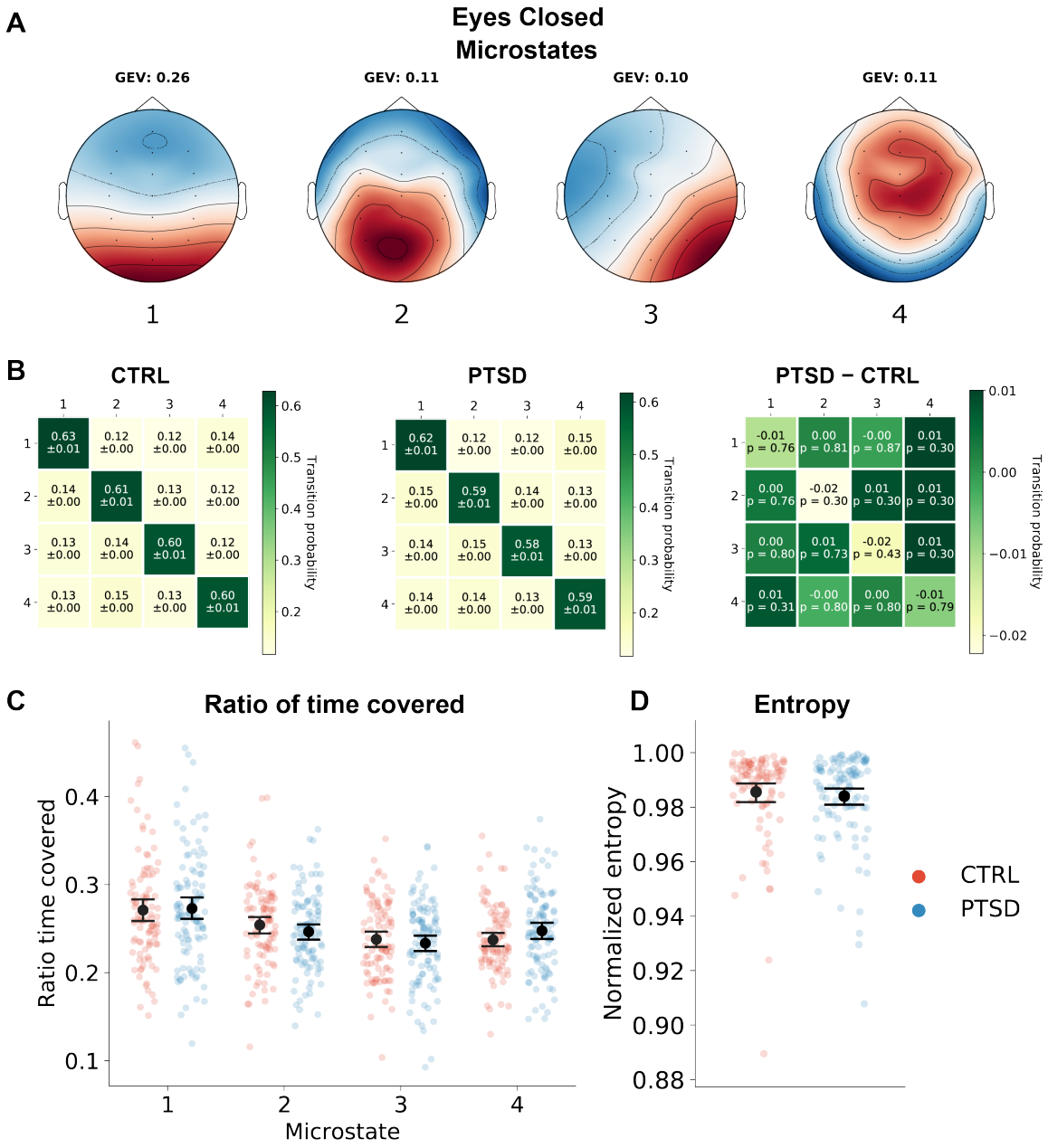

Figure 4.8: No differences in eyes closed microstates between PTSD and controls. A) Topographical maps of the microstates. B) Transition matrices for the two groups and their differences in probability of transitioning from one microstate to another. Direction should be read as row to column. C) Ratio of time covered for each microstate. D) Shannon entropy of the two groups, normalized to the maximum entropy obtained by a uniform distribution of microstate labels. Mean ± 95% confidence intervals are shown. GEV, global explained variance.

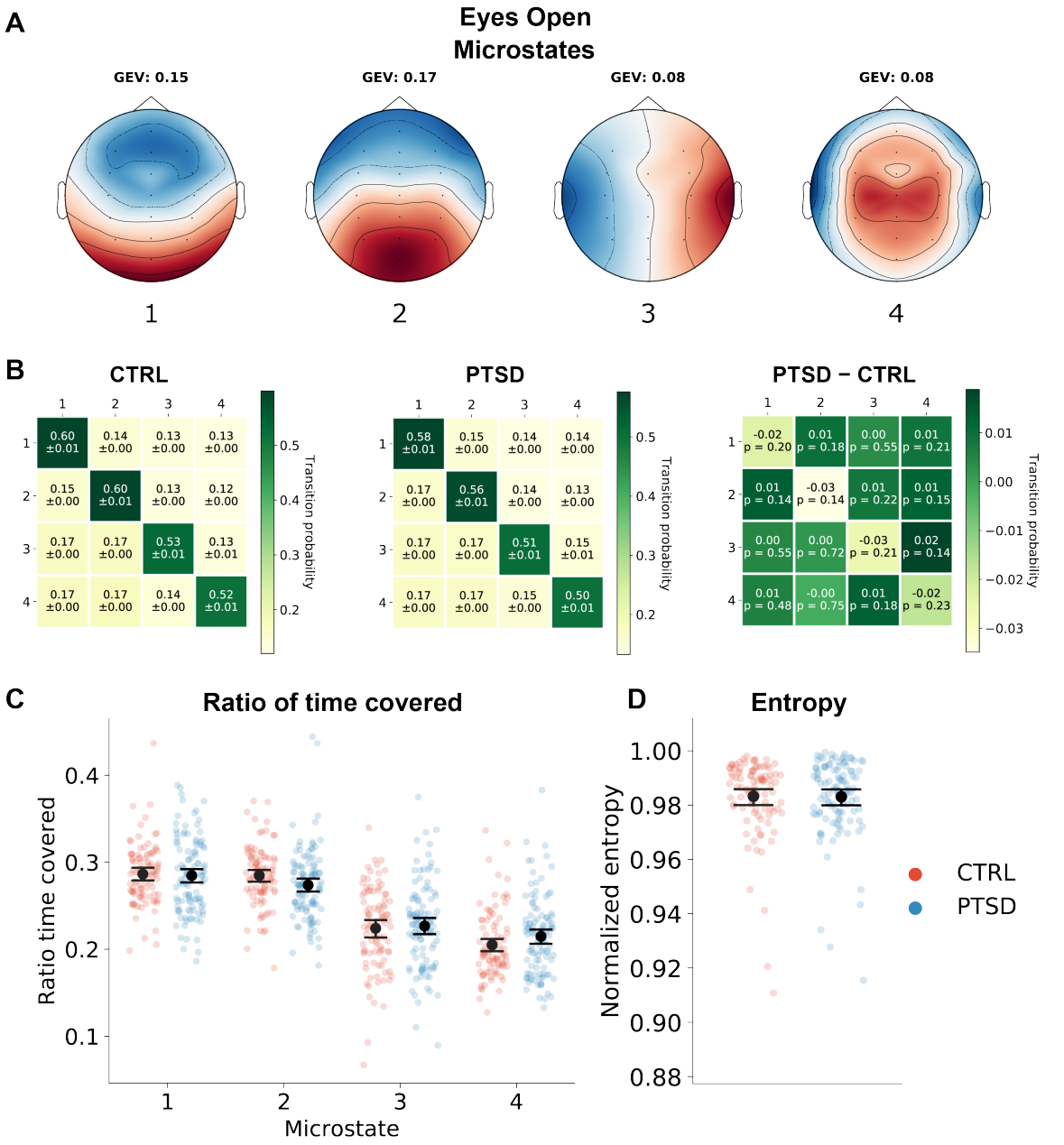

Figure 4.9: No differences in eyes open microstates between PTSD and controls. A) Topographical maps of the microstates. B) Transition matrices for the two groups and their differences in probability of transitioning from one microstate to another. Direction should be read as row to column. C) Ratio of time covered for each microstate. D) Shannon entropy of the two groups, normalized to the maximum entropy obtained by a uniform distribution of microstate labels. Mean ± 95% confidence intervals are shown. GEV, global explained variance.

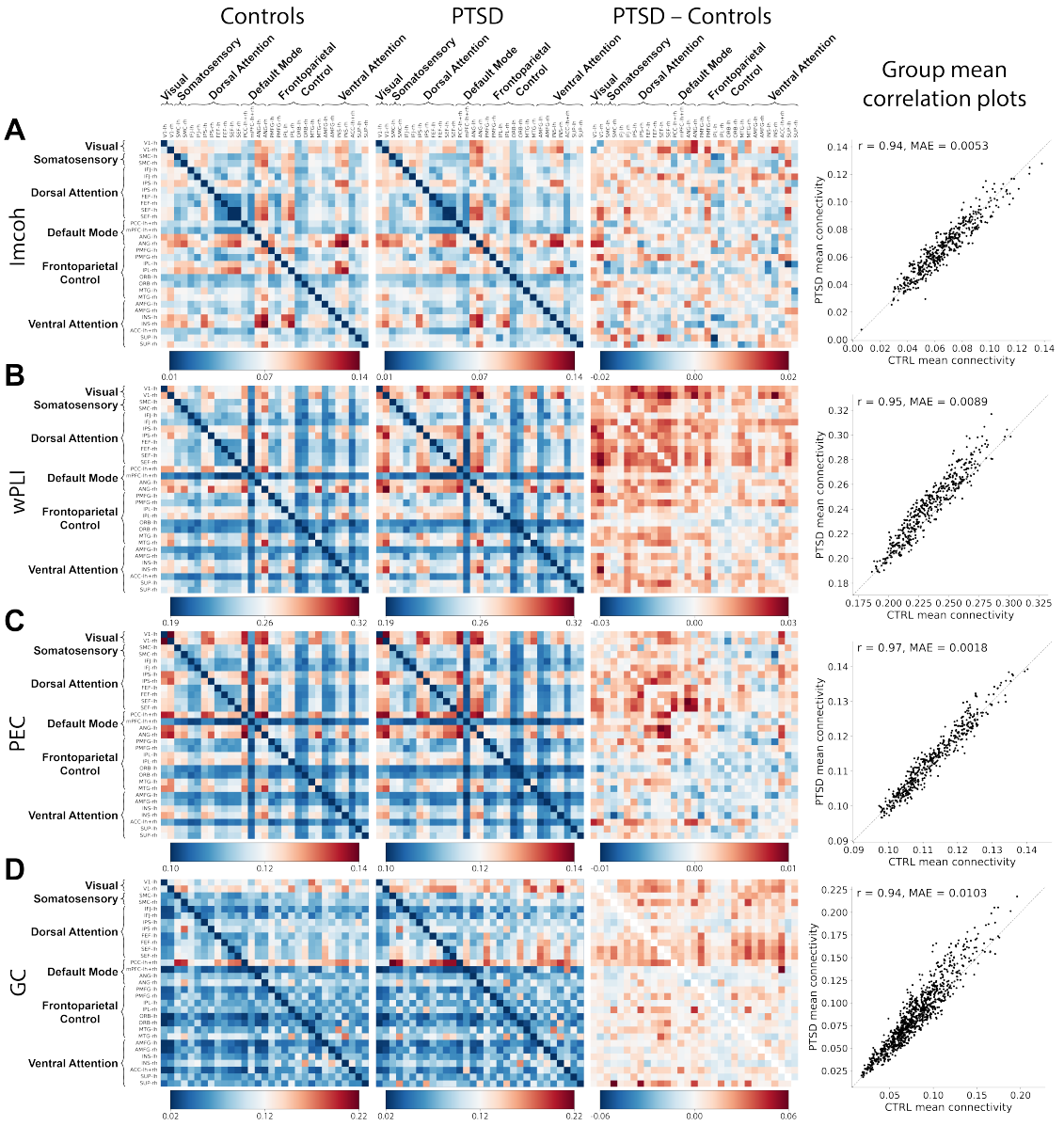

Figure 4.10: No differences in connectivity between PTSD and controls. Examples of the mean A) imaginary coherence, B) weighted phase lag index, C) power envelope correlations, and D) Granger causality for the PTSD and control group, their difference, and correlation in the alpha band during eyes closed condition. The closer the points are to the diagonal line in the correlation plots, the better the correlation between the group-mean connectivities. For the GC plots, the direction should be read as row to column. Pearson's correlation was used to compute r. MAE, mean absolute error.

EEG Feature	Statistical test results for group-mean differences
Power	Relative beta and gamma power is significantly higher in PTSD during eyes open condition
Frontal Theta/Beta Ratio	Frontal theta/beta ratio is significantly lower in PTSD during eyes open condition
Asymmetry	No significant differences observed
Peak Alpha Frequency	No significant differences observed
1/f Exponents	1/f exponents are significantly lower in PTSD in both eyes open and eyes closed condition
DFA Exponent	No significant differences observed
Microstates	No significant differences observed
Source functional connectivity	No significant differences observed

Table 4.1: Overview of group-mean findings for each EEG feature.

also used t-Distributed Stochastic Neighbor Embedding (t-SNE; [68]) on the features that were consistently selected by our best machine learning classifier and observed some separation, although there were no clear separation and a large overlap between the groups (Figure 4.11C), consistent with the moderate classification performance we observed.

4.2.3 PTSD subtyping

Inspired by Zhang et al. [99] who applied unsupervised K-means sparse clustering to identify subtypes within the PTSD group that responded differently to treatment, we adapted their method and investigated subtypes in our dataset. We found two subtypes: subtype 1 with distinctively greater connectivity and subtype 2 with relatively similar connectivity as controls, albeit slightly lower (Figure 4.12). 29 out of the 107 patients with PTSD had electrophysiological activity corresponding to subtype 1 and 78 participants to subtype 2. Interestingly, we were able to classify subtype 1 from controls with 79.4% balanced test accuracy, while the classification performance on subtype 2 remained around 63% (Figure 4.13A), indicating that the subtyping was able to delineate some of the biological heterogeneity and improved the performance of our classifiers on a subset of the patients. The decreased heterogeneity after subtyping was also qualitatively confirmed by t-SNE, where we observed clear separation of the PTSD patients with subtype 1 and the controls (Figure 4.13B).

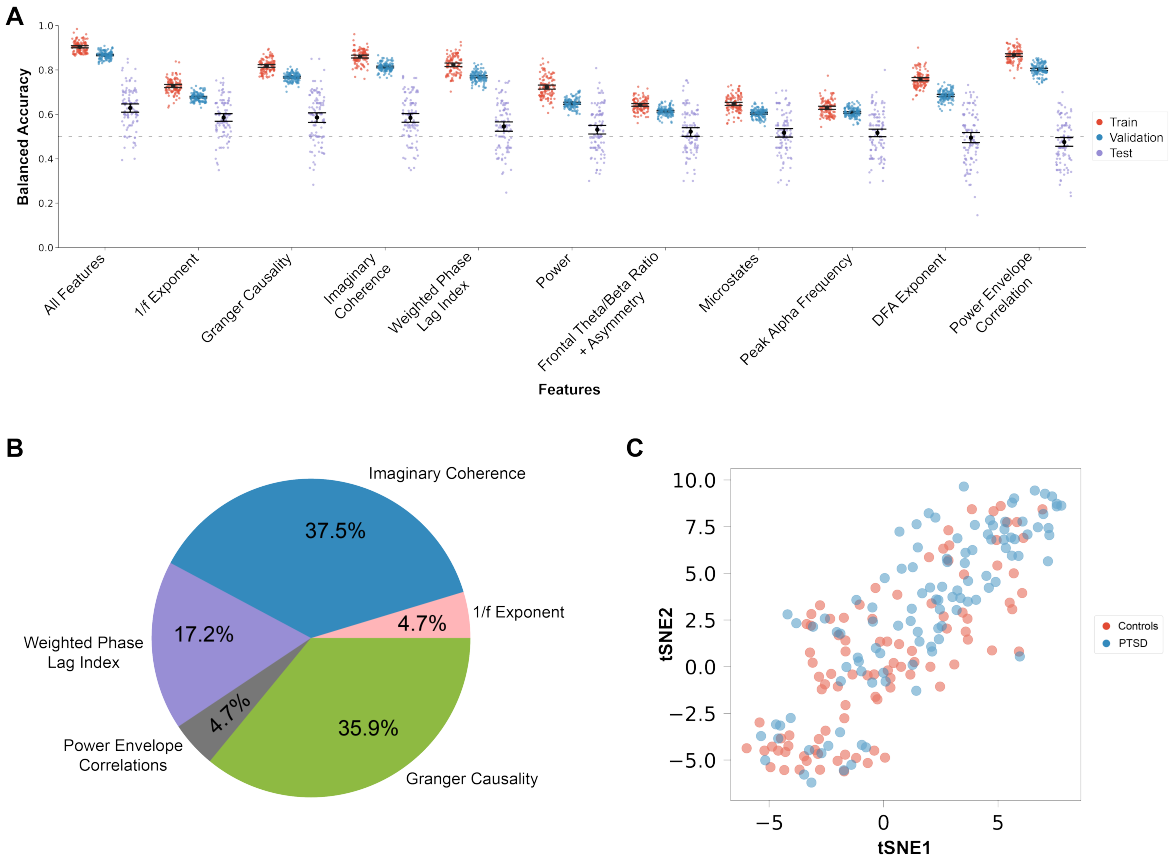

Figure 4.11: Machine learning classification of PTSD. A) Overview of SVM classification performances utilizing combinations of feature types or each feature type separately. The best model, SVM with all features, obtained a balanced accuracy of 62.9%. B) Pie chart of the features that were consistently selected by the best SVM model in at least 20% of all CV runs. It was primarily functional connectivity and 1/f exponents that were consistently selected across CV folds, indicating they consistently provide predictive information. C) t-SNE of the consistently important features selected by SVM. Some separation between the two groups can be observed with the PTSD group being more prominent in the upper right corner and controls in the lower left corner, although there is still a large overlap, reflecting the moderate classification performance we obtained using SVM with all features. Mean ± 95% confidence intervals are shown. Figure adapted from Appendix A.

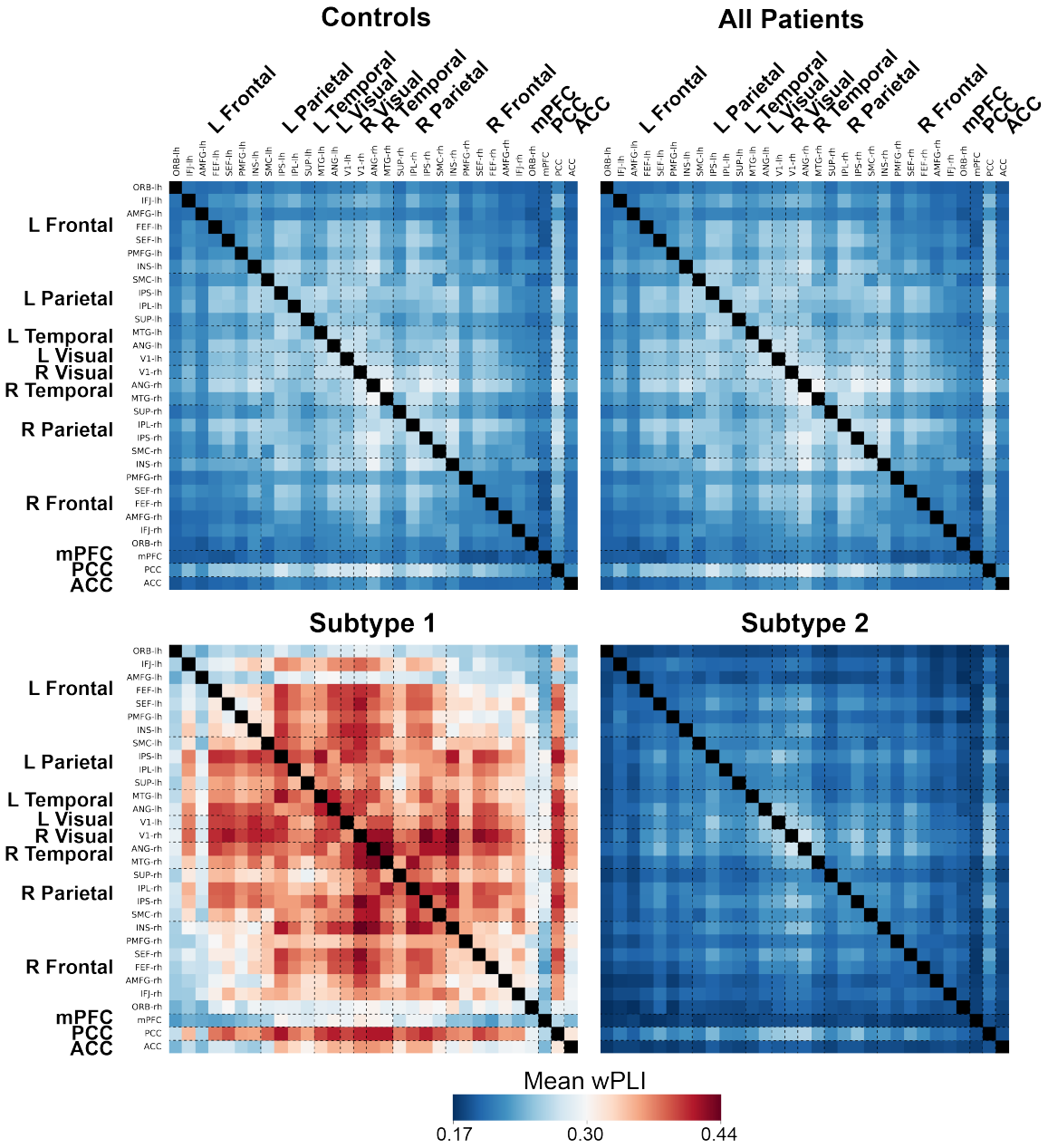

Figure 4.12: Unsupervised clustering found two distinct PTSD subtypes. Sparse K-means clustering was applied on the PTSD group and two clusters were found, primarily based on connectivity features during eyes closed condition. A) Examples of the mean alpha wPLI in the different groups. Subtype 1 had distinctively greater connectivity compared to controls, while subtype 2 was relatively similar to controls, albeit with slightly lower wPLI. Figure adapted from Appendix A.

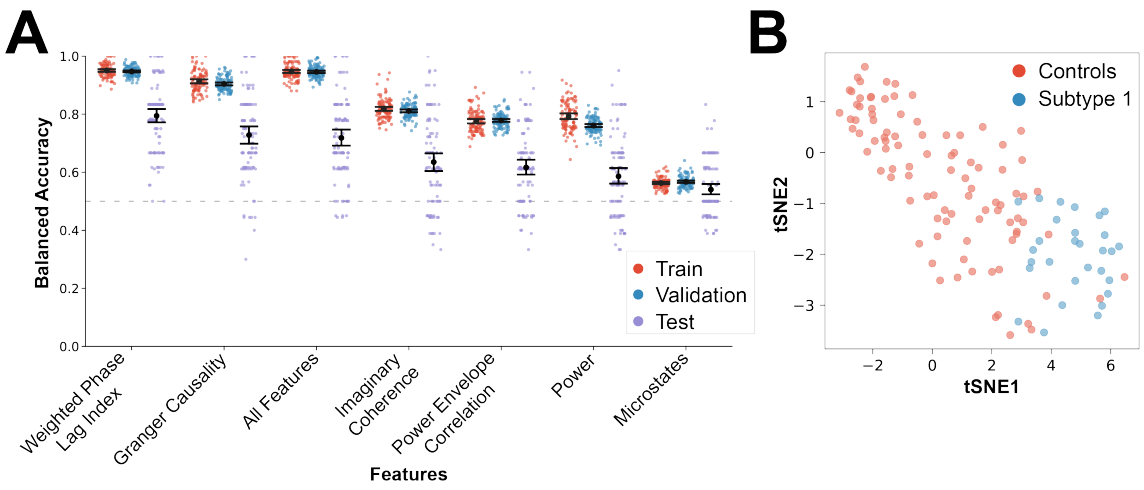

Figure 4.13: Subtyping delineated heterogeneity and improved machine learning classifications. A) Subtype 1 was classified with up to 79.4% balanced accuracy using logistic regression, a clear improvement compared to predicting on the whole PTSD group. B) t-SNE of the important wPLI features selected consistently by logistic regression (in at least 20% of all CV runs) showed a clear separation between the control group and subtype 1. Mean ± 95% confidence intervals are shown. Figure adapted from Appendix A.

4.3 Discussion

We found that relative beta and gamma power during the eyes open condition was significantly higher in the PTSD group, which is consistent with the findings from a few other studies [86, 87, 92]. Increased beta and gamma have also been found to be higher during eyes closed condition [80, 81, 85], however we did not observe this effect. As previously mentioned, there have been many conflicted findings reported in the literature regarding resting-state EEG power in PTSD (Appendix A Supplementary Table A1), thus our null-findings in delta, theta and alpha bands are consistent with some of the studies, while inconsistent with the studies that reported a significant effect.

We also observed reduced frontal theta/beta ratio during the eyes open condition. This effect seems to be driven by the increased beta power we observed, as we did not find any differences in theta power between the PTSD and control group. One other study [87] have also reported a reduced theta/beta ratio in PTSD, however after accounting

for the effect of medication this effect was no longer significant.

The last significant group-mean finding we observed was the lower 1/f exponents in the PTSD group. The 1/f relation between power and frequency in EEG power spectra have been well-known for many years and thought to be associated with self-organized criticality in neuronal networks and cognitive processing [51, 115–118]. Recently, 1/f exponents have received increased interest with the heightened focus on the notion of separating the neural power spectra into periodic (e.g. the canonical frequency bands) and aperiodic (i.e. 1/f scaling that is not associated with a specific frequency) components, due to the interactions between the two components [34, 119]. The interaction is especially prominent when looking at ratios between frequency bands [120], which theta/beta ratio is a prime example of. The reduced 1/f exponents we observed in the PTSD group, can also explain the increased beta and gamma power and reduced theta/beta ratio findings [121], as a lower 1/f exponent means the power decreases more slowly with increasing frequencies. Donoghue et al. [34] argues that in this case it is inaccurate to say the beta and gamma power is higher, because the baseline power is higher due to the reduced 1/f exponent and it is thus an aperiodic effect and not a periodic (frequency-specific) effect. To our knowledge, no other PTSD studies have investigated 1/f exponents in EEG power spectra, however differences in 1/f exponents have also been shown in development, aging and diseases [34]. Increased 1/f exponents have been observed in attention deficit hyperactivity disorder [122] and schizophrenia [123], which was normalized to the levels of the control group following medication. 1/f exponents have also been observed to be negatively correlated with excitation/inhibition (E/I) ratio, with a lower exponent being associated with higher E:I ratio [37]. Future work should further investigate the role of 1/f exponents in PTSD.

While the significant group-mean findings are interesting, they however do not tell how well the different features would serve as biomarkers for predicting PTSD individuals. They only tell us what is different on average between our PTSD and control group [76]. Visual inspection of distributions of the EEG feature values that showed a significant group-mean difference also did not invoke the idea that they would be capable of separating PTSD from control participants due to their high variability.

To identify potential biomarkers for PTSD, we employed multivariate machine learning models with two-layer CV to estimate how well our models would predict on new unseen

participants and which combinations of EEG features would be useful for such a task. We found that combinations of source functional connectivity and 1/f exponents were able to classify PTSD participants significantly better than chance-level (Wilcoxon signed-rank tests with $p < 0.001$), however only with a moderate accuracy of 62.9%. Our observed test accuracy was similar to another resting-state EEG study that employed 10 repetitions of 10-fold CV on a combination of source space covariance matrices, spectral power and network metrics to obtain around 64%, 65% and 62% test accuracies for LDA, SVM and random forest respectively [103]. Interestingly, they also tried a Riemannian classifier, which enhanced their test accuracy up to around 73% [103]. Another resting-state EEG study also employed 10 repetitions of 10-fold CV and obtained 75.6% balanced test accuracy using a linear relevance vector machine on source space PEC [46]. We also tested the performance of our three classifiers using only source PEC features, however our test accuracies were never significantly better than chance-level. We cannot provide a clear reason why PEC was distinctively worse than the other connectivity features for our dataset.

To delineate some of the high heterogeneity in the PTSD group, we applied sparse K-means clustering and discovered two subtypes, one with relatively similar connectivity to the control group, while the other had distinctively increased connectivity. We were able to classify the subtype with increased connectivity with up to 79.4% accuracy, a clear improvement compared to predicting on the whole PTSD group. This indicated that the subtyping was able to attenuate some of the heterogeneity and we found a portion of the PTSD participants with increased connectivity clearly distinguishable from the controls.

Interestingly, most of the important connections utilized consistently by the classifier was in the dorsal and ventral attention networks in the alpha frequency range and they were positively correlated with self-reported arousal severity score, a central symptom cluster in PTSD. Additionally, subtype 1 with hyperconnectivity also had higher arousal scores on average compared to subtype 2 participants (Appendix A).

Taken together, the study presented in this chapter has shown that some of the EEG features are different in the PTSD group on average, however many EEG features were also similar between the PTSD and control participants. Furthermore, we went beyond the conventional case-control group comparisons, and using multivariate machine

learning models, we were able to classify PTSD above chance-level primarily from source functional connectivity. Delineating heterogeneity in the PTSD group using data-driven unsupervised clustering further improved classification performance to around 80% accuracy and we were able to identify a subgroup of PTSD patients characterized by hyperconnectivity and higher arousal scores that were distinctively different from the control participants.

CHAPTER 5

Autism spectrum disorder

ASD is a neurodevelopmental disorder characterized by atypical sensory reactivity, persistent difficulties in social interactions, and restricted and repetitive behavior [1]. It is a highly heterogeneous disorder, since the criteria captures a broad spectrum of symptoms. Additionally, the criteria for ASD according to the DSM has changed considerably in the past decades, especially between its fourth and fifth edition where previously distinct conditions have been merged into the current ASD definition. To determine the neurobiological underlying mechanisms for ASD, many studies have employed neuroimaging [124–126], however no consensus about the physiological correlates of the disorder have been reached and it is still an actively investigated field.

Similar to the PTSD resting-state EEG literature, spectral power is also the most commonly investigated feature in the ASD resting-state EEG literature and the findings have similarly been inconsistent [124]. Many recent studies have also gone beyond the case-control group-mean investigation and applied predictive machine learning models, which aligns better with the goal of diagnostic/prognostic tools to predict new subjects or future outcomes [7]. The results have been very promising, with accuracies up to 85% [127], 95% [128], and 96.4% accuracy [129]. However, upon scrutiny it becomes apparent that these three studies did not perform any CV, thus the reported accuracies does not reflect how well they would work on new unseen subjects [9, 69, 75, 77]. Some studies did employ CV and still obtained up to 93% [130, 131], 94% [132], and 99.7% [133] accuracy. However, the models these studies employed have hyperparameters, but there were no mention of how they were tuned. If only a single layer of CV is used, then the reported accuracies also do not properly reflect how well the models would generalize to new subjects as the test data might have been used for tuning the model [69].

Taken together, many interesting EEG biomarkers have been found promising for ASD, however no clear consensus have been reached due to low sample sizes or inconsistent

findings and many of the predictive modelling studies did not estimate how well the biomarkers would work on unseen data. Thus, we set out to apply our exploratory machine learning framework with two-layer CV to determine if any of the features or combinations of features could serve as biomarkers for characterization of unseen participants with ASD.

5.1 Methods

The data was acquired in collaboration with the Netherlands Autism Register [134] and inclusion criteria for the ASD cohort was a clinical diagnosis based on DSM-IV or DSM-5 and age between 18 and 55. After quality checks and EEG preprocessing, the final sample size consisted of 186 adults (51% with ASD and 49% non-autistic participants). The autism-spectrum quotient short questionnaire (AQ; [135]), a 28-item self-report questionnaire was applied to probe the autistic traits. All items were summed to obtain a total AQ score, with higher scores indicating more autistic traits. A subset of the sensory perception quotient short questionnaire (SPQ; [136, 137]) was also applied to probe for visual (SPQvis) and auditory (SPQaud) sensitivity. Lower scores on the SPQ indicate lower sensory threshold, hence higher sensory sensitivity.

5 min of eyes closed resting-state EEG was recorded with a 64-channel BioSemi system sampled at 2048 Hz. The participants received the instructions: "Please keep your eyes closed, relax, and try not to fall asleep". The EEG was preprocessed as described in Chapter 3.1 and downsampled to 500 Hz. Source localization based on 68 cortical patches from the Desikan Killiany atlas was performed and all features were estimated in source space.

The features we investigated could broadly be divided into three categories: spectral features (power, asymmetry, theta/beta ratio, peak alpha frequency and 1/f exponent), measures of criticality (long-range temporal correlations [DFA exponents] and functional E/I ratio [fEI]) and functional connectivity (coherence, imaginary part of coherence, phase locking value, weighted phase lag index and power envelope correlations).

5.1.1 Statistical analysis

Since all features were estimated in source space, we were unable to use cluster-based permutation test due to the adjacency matrix not being well-defined. Thus, we used ordinary non-parametric permutation tests to test for group-mean differences and Levene's test to assess equality of variances. FDR was employed to correct for multiple comparisons. To test for linear correlations, we used Pearson's correlation coefficient and Student's t-distribution. Non-parametric Wilcoxon signed-rank test was used to test differences between classifiers and one-sided tests were used for baseline comparisons (chance-level for the classifiers or mean prediction for the regressors). A p-value < 0.05 was considered significant for rejection of the null hypothesis. Results are shown as mean with 95% confidence intervals.

5.2 Results

Similar to the PTSD study, we also started exploring the ASD dataset by visualizing all the different EEG features and looked for group-mean differences between the ASD and non-autistic comparison group. The initial analysis revealed that 1/f exponents were the only feature type that showed a significant difference, with the ASD participants having lower 1/f exponents (Figure 5.1A). However, further investigation of 1/f exponents indicated that 1/f exponents have previously been found to be negatively correlated with age [33–36], and in our dataset the ASD group was on average 11 years older than the non-autistic comparison group (permutation test, $p = 0.0001$). To account for the age-effect on 1/f exponents, we corrected the 1/f exponents by removing the linear trend with age. Following age-effect correction, 1/f exponents were no longer significantly different between the ASD and non-autistic comparison group (Figure 5.1B).

We also investigated if any of the other EEG features were correlated with age and found theta/beta ratio, delta PEC, absolute delta, theta and alpha power, and relative beta and gamma power to be correlated with age (Figure 5.2).

The above-mentioned features that exhibited a high degree of age dependence were corrected by removing the linear trend with age. After age-effect correction, no significant differences in group-means were observed between the ASD and non-autistic comparison group (permutation tests with FDR correction for each feature type separately,

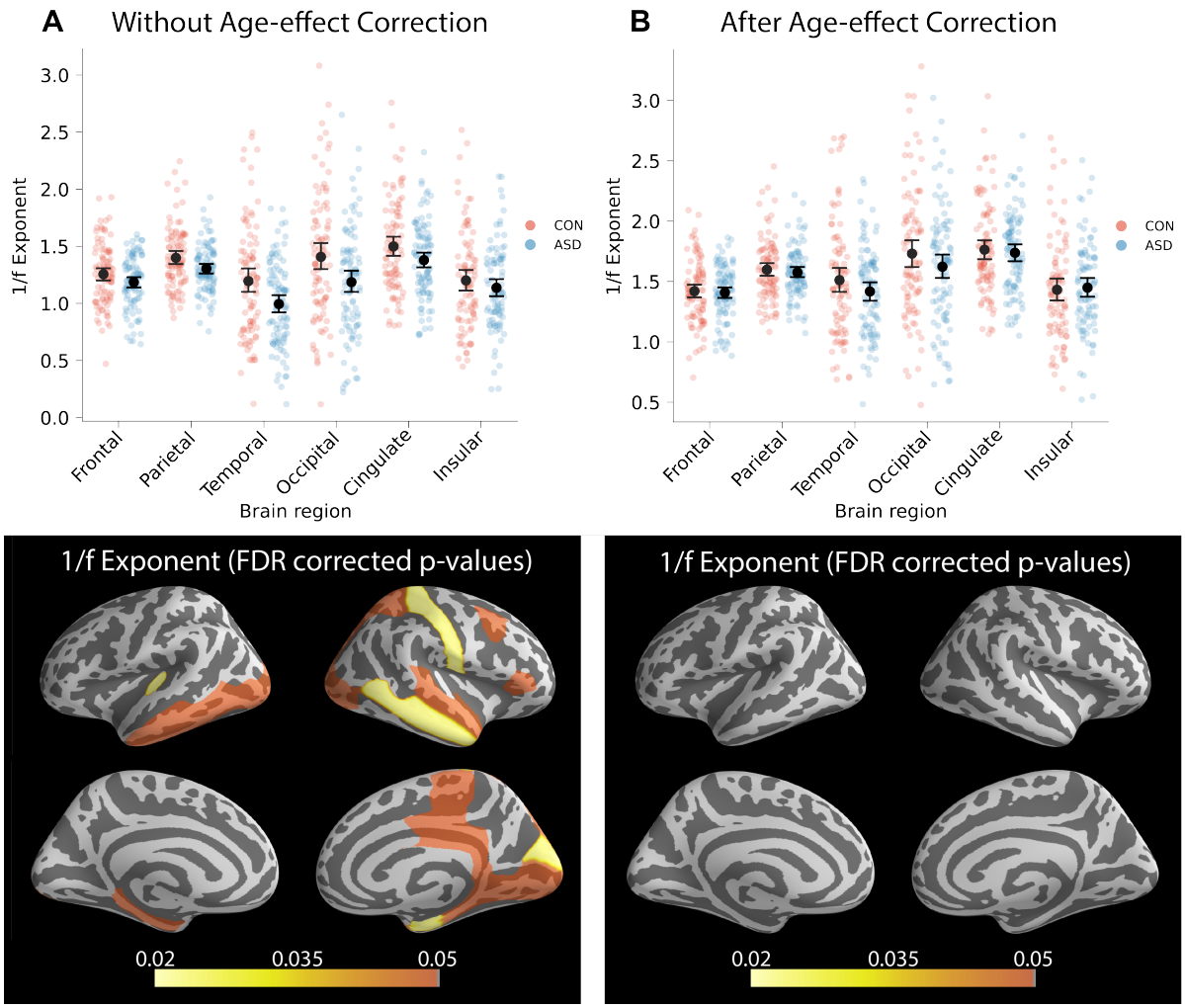

Figure 5.1: The effect of age on 1/f exponents. A) Initial analysis revealed that 1/f exponents were significantly lower in the ASD group. B) However, this effect was confounded by age differences between the two groups. After age-effect correction, the 1/f exponent differences were no longer significant. Mean ± 95% confidence intervals are shown. FDR, false discovery rate. Figure from Appendix B.

$p > 0.05$). Figure 5.3 show examples of spectral and criticality features and Figure 5.4 show examples of connectivity features. We also investigated if the ASD group showed differences in variance, but did not find any significant differences (Levene's test with FDR correction for each feature type separately, $p > 0.05$)

Based on the visual inspections and statistical tests of EEG features, we did not identify

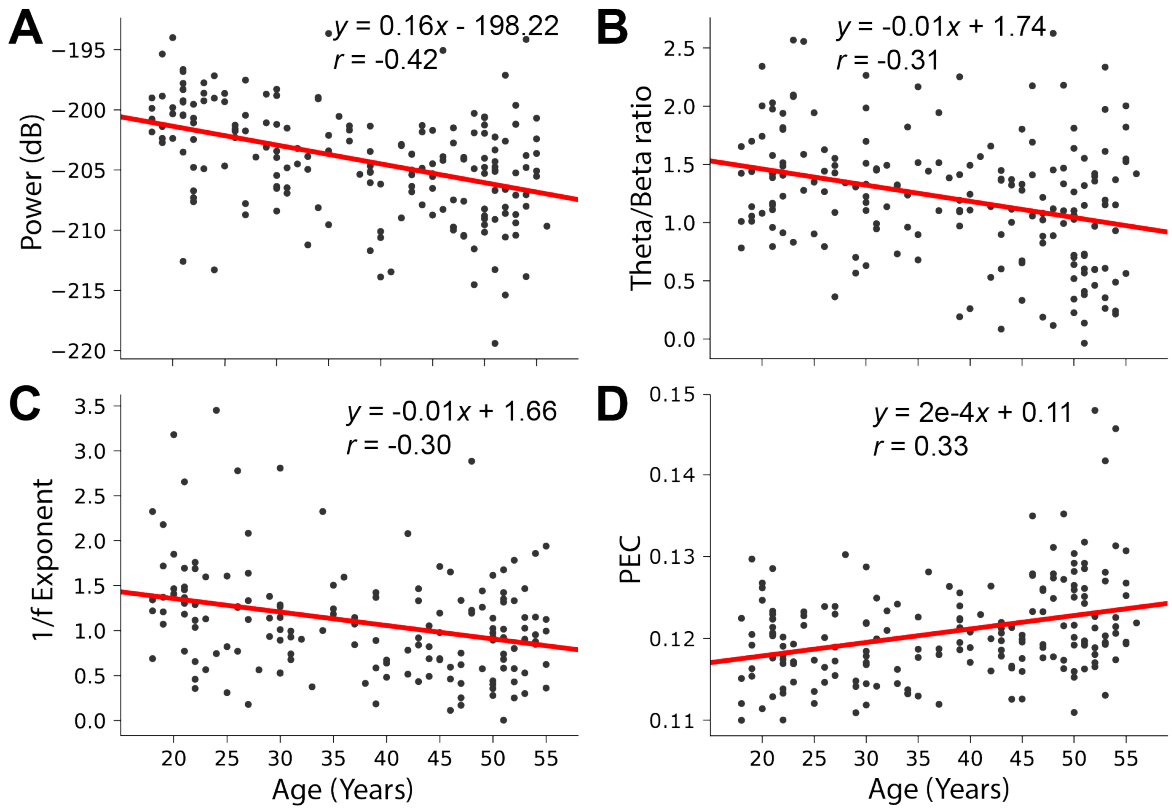

Figure 5.2: Several EEG features exhibit age dependence. Scatter plots illustrating correlations between age and A) absolute power, B) theta/beta ratio, C) 1/f exponent, and D) PEC of source-modelled signals. Figure from Appendix B.

any clear candidate biomarkers, however even if each feature have small effect sizes, they might still effectively serve as biomarkers when considered in a multivariate fashion in combination with other features [76]. Nonetheless, our machine learning models merely obtained 55.8% balanced test accuracy, indicating that they were unable to identify a brain pattern that could clearly distinguish the ASD group from the non-autistic comparison group.

5.2.1 ASD subtyping

As ASD is known to be a very heterogeneous disorder with the severity scores of the symptoms lying on a spectrum, we also performed unsupervised K-means sparse clustering on the ASD dataset to delineate heterogeneity by identifying electrophysiologically-based

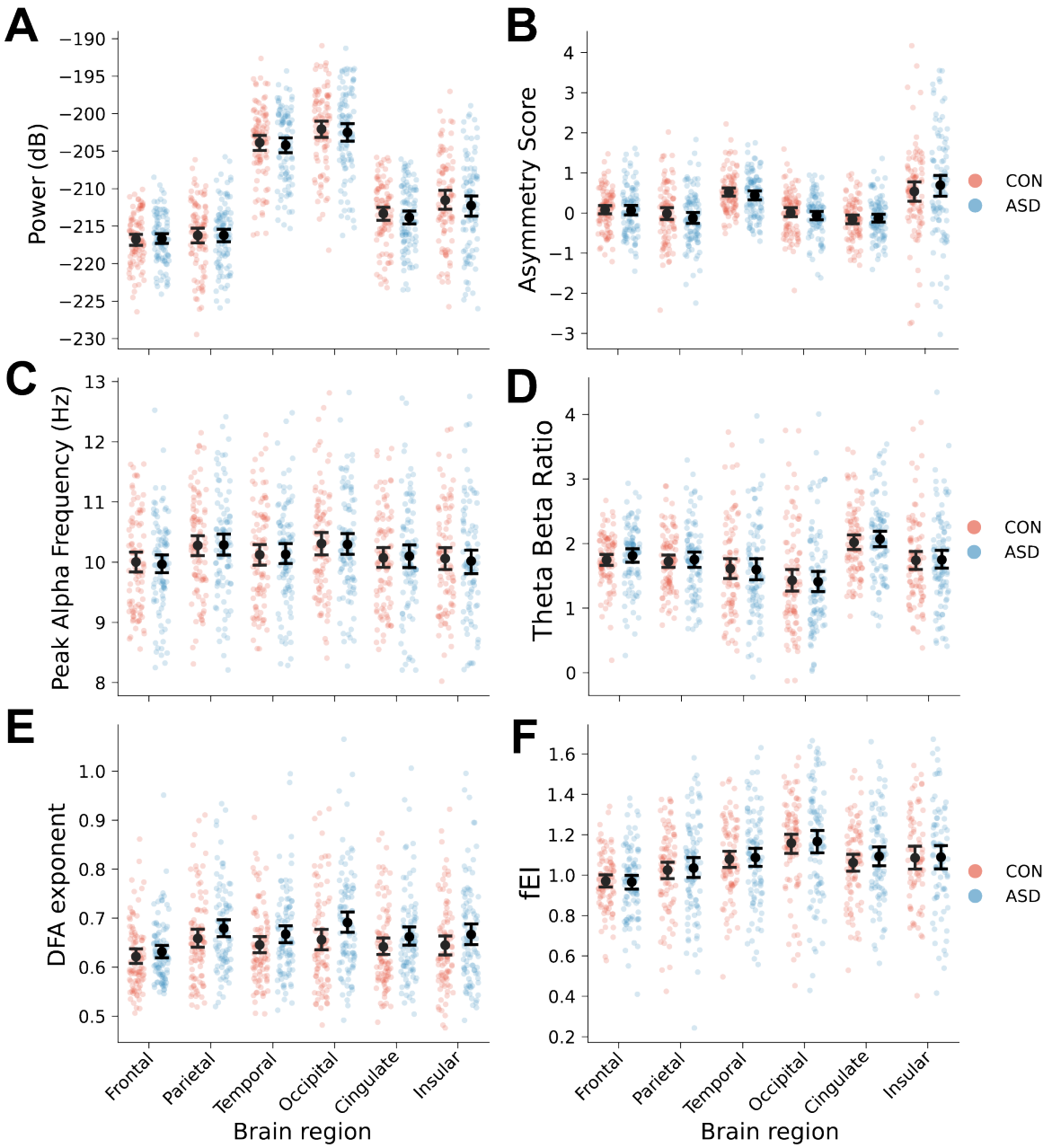

Figure 5.3: No differences in spectral and criticality features between the ASD and non-autistic comparison group. Representative examples of A) absolute alpha power, B) alpha asymmetry, C) peak alpha frequency, D) theta/beta ratio, E) alpha DFA exponents, and F) alpha fEI averaged across brain regions. Mean ± 95% confidence intervals are shown. Figure from Appendix B.

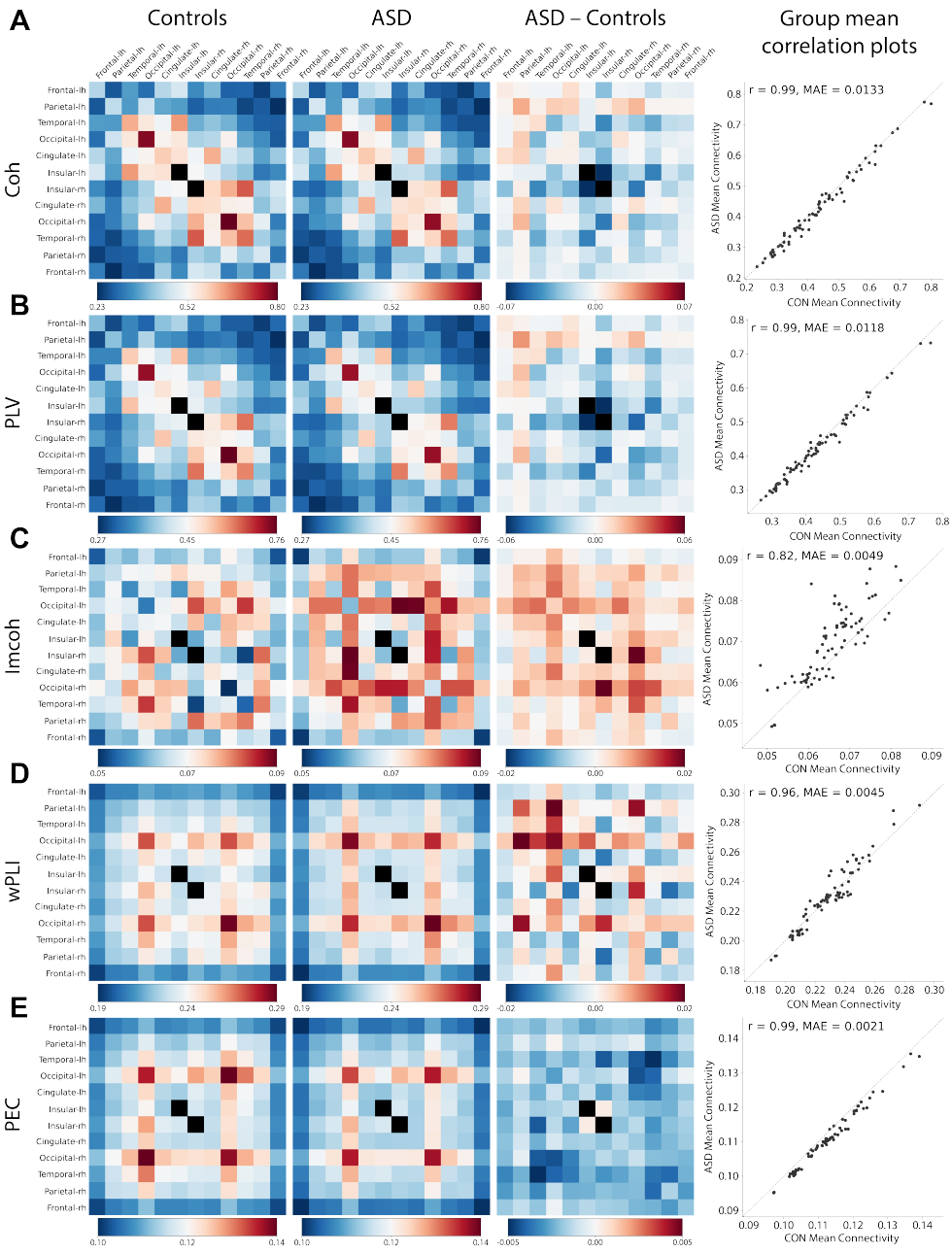

Figure 5.4: No differences in connectivity between the ASD and non-autistic comparison group. Representative examples of the mean A) coherence, B) phase locking value, C) imaginary coherence, D) weighted phase lag index, and E) power envelope correlations for the ASD and non-autistic comparison group, their difference, and correlation in the alpha band. Pearson's correlation was used to compute r. MAE, mean absolute error. Figure from Appendix B.

subtypes within adults with ASD. We found two subtypes within the ASD group, which were primarily distinct from each other based on spectral power, functional connectivity and fEI (Figure 5.5A). 38 out of the 95 ASD participants had electrophysiological activity corresponding to subtype 1, while 57 corresponded to subtype 2. Further investigation revealed it was primarily features from the alpha band that was important for distinguishing the two subtypes (Figure 5.5B). Subtype 1 was characterized by decreased relative alpha power and alpha coherence compared to the non-autistic comparison group, while subtype 2 had increased relative alpha power and alpha coherence (Figure 5.5C and 5.6).

The classifiers trained to predict the subtypes from the non-autistic comparison group also obtained slightly better performances compared to distinguishing the whole ASD group from the comparison group. We were able to classify subtype 1 with up to 64% balanced test accuracy with SVM and subtype 2 was classified with up to 68% balanced test accuracy with logistic regression with L2 norm regularization. However, although clear differences in electrophysiology were observed between the two subtypes, the two subtypes did not differ in AQ score on any of the subscores (permutation tests, $p > 0.05$), thus indicating the two subtypes were behaviorally very similar (Figure 5.7).

5.3 Discussion

After correcting for the age effect, we did not observe any group-mean differences for any of the feature types, despite only correcting for multiple tests for each feature type separately, to easier compare with previous studies that primarily focused on a single or few feature types. Our machine learning models also only performed slightly better than chance-level, with the best classification around 56% balanced test accuracy, indicating that the electrophysiological activity during resting-state was very similar between our adult ASD participants and the non-autistic adult comparison group.

This finding was inconsistent with many previous studies that reported above 90% accuracies on classifying ASD [128–133], however these studies did not employ CV, or only employed one-layer CV despite having hyperparameters. If only a single layer of CV is used, the hyperparameters might have been tuned to the test set and thus inflated the test accuracy [69]. It could also be that the EEG differences are greater during neurode-

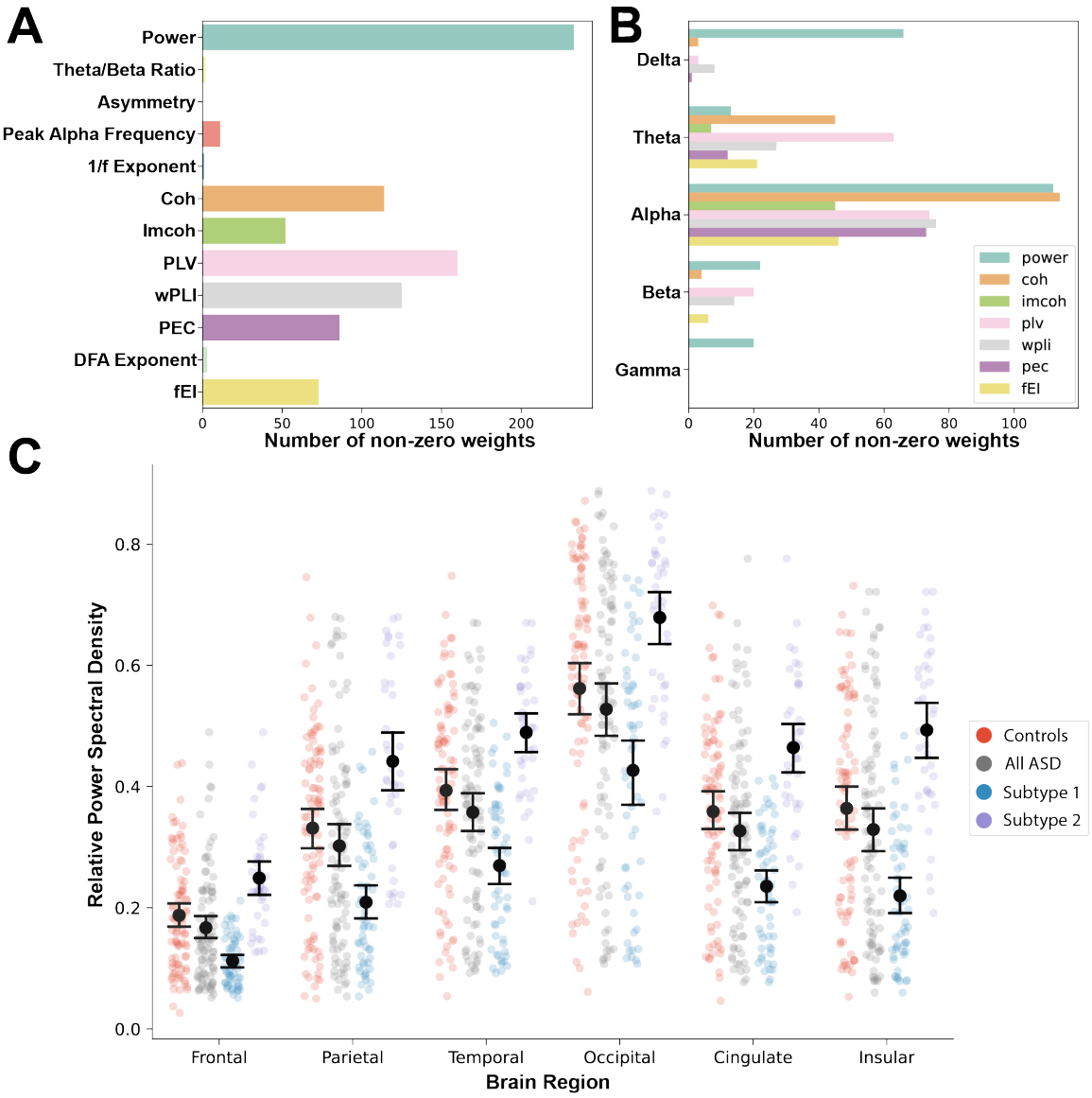

Figure 5.5: Unsupervised sparse K-means clustering identified two ASD subtypes. A) The two subtypes were primarily distinct from each other based on spectral power, functional connectivity and fEI features. B) Specifically, the features in the alpha band were mainly utilized by the sparse K-means algorithm. C) relative alpha power was lower in subtype 1, and higher in subtype 2, compared to the non-autistic comparison group. Mean ± 95% confidence intervals are shown.

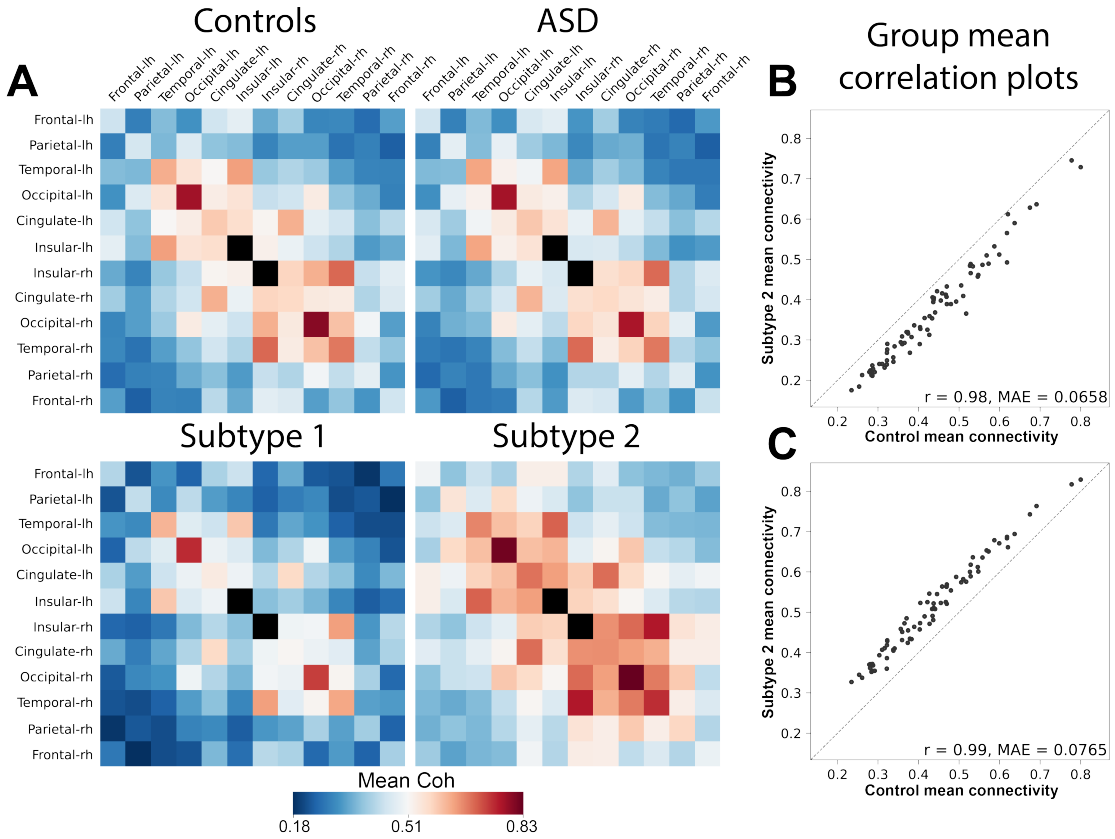

Figure 5.6: The two subtypes have different functional connectivity. Alpha coherence was lower in subtype 1, and higher in subtype 2, compared to the non-autistic comparison group. A) Connectivity matrices. B) Correlation plot between subtype 1 and the non-autistic comparison group and C) subtype 2 and the comparison group. Pearson's correlation was used to compute r. MAE, mean absolute error.

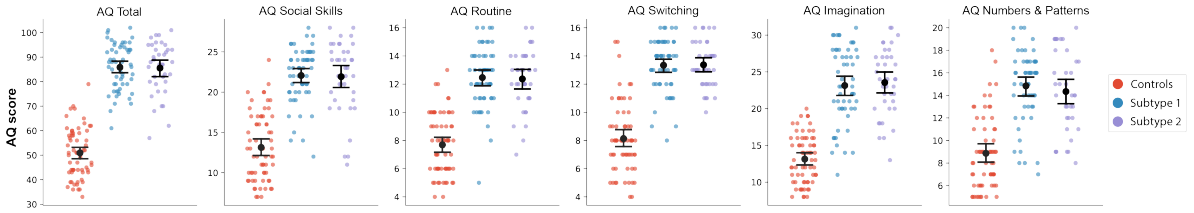

Figure 5.7: No differences in clinical severity scores between the two ASD subtypes. The two subtypes did not differ in terms of the total AQ score or in any of the AQ subscores.

velopment and childhood, as most previous ASD resting-state EEG studies investigated children.

However, a recently published resting-state EEG study investigating 411 children, adolescents and adults with and without ASD obtained around 57% accuracy on unseen data [138], which was comparable to our findings. They also tried applying machine learning on subset of the data consisting only of children, but the performance they obtained in this situation was not significantly better than chance-level [138]. To our knowledge, no other resting-state EEG studies have applied machine learning to classify adults with ASD.

Some previous studies also investigated group-mean resting-state EEG differences in adults with ASD and found significant differences in microstates [113], alpha power [139] and coherence [139–141]. However, these studies all had less than 50 samples in total, thus they are likely not representing the whole clinical population, especially in a disorder as ASD, which is known to be heterogeneous with symptoms lying on a wide spectrum [142].

To decompose some of the heterogeneity in our sample, we also applied unsupervised clustering and looked for subtypes within the ASD group. We found two subtypes, with subtype 1 having decreased alpha power and hypoconnectivity, while subtype 2 had increased alpha power and hyperconnectivity. The classifiers were better at distinguishing the subtypes from the non-autistic comparison group, although the improvement was only moderate and far from being clinically relevant. The lack of differences in clinical questionnaire scores also indicated the two subtypes had no clear behavioral differences, thus it was hard to interpret the subtypes. It is always possible to find subgroups with an unsupervised method, you just have to specify how many you want, but the importance of the subgroups are evaluated based on interpretations of their functional role, e.g. clinical relevance.

Given that the interpretation of our identified subtypes were unclear, we did not include it in the paper or work further in that direction, however it does not mean the subtyping road is a dead end. We used 2 subtypes because the gap statistic suggested that was the most appropriate number, but if we had more samples, the patient population would not only be better represented, but it would also make the machine learning models less likely to overfit and enable identification of more subtypes [7]). Besides matching

subtypes to baseline clinical scores, future work could also match potential subtypes with treatment outcomes [99], i.e. subtyping might help predict treatment outcomes and enable better tailored treatment [7, 143, 144].

To sum up, our results on the ASD resting-state EEG data indicated that adults with ASD have electrophysiological activity patterns within the typical range. This does not mean resting-state EEG contain no meaningful information about ASD, however the effects were likely too small to be picked up by our models relative to the high heterogeneity in ASD, even with a sample size of close to 200 participants. The subtyping did successfully delineate some of the heterogeneity and improved the classification performances slightly, albeit we were not able to derive a clinically meaningful interpretation of the subtypes.

CHAPTER 6

Conclusion

We set out to develop a machine learning framework for discovery of EEG biomarkers for clinical purposes, and applied the framework to a PTSD and ASD dataset. Identification of quantifiable biomarkers that can aid the transition to precision medicine is of high importance [6–8, 62, 143]. It was pioneering work, as most previous resting-state EEG studies primarily focused on single feature types or specific subsets of features in isolation, whereas we computed an extensive set of EEG features and investigated whether combinations of multiple EEG features could serve as diagnostic biomarkers. Additionally, the most common approach by previous EEG studies were investigations of group-mean differences by applying conventional statistics, however as described in Section 2.3, machine learning models are naturally better by design to identify biomarkers for diagnostic/prognostic predictions on an individual level. Some previous resting-state EEG studies did employ machine learning, but surprisingly, many studies did not test their models on an entirely unseen dataset, and thus their reported model performances were likely to be inflated due to overfitting. To attenuate the issue of overfitting, we employed a robust repeated two-layer cross-validation scheme.

Apart from the machine learning analysis part of our pipeline, we also implemented state-of-the-art EEG preprocessing to properly clean the data prior to applying the machine learning algorithms. The EEG features we computed and incorporated in our framework also ranged from classic features that have been investigated for many decades to newly developed features like fEI [38] and the FOOOF algorithm [34]. To our knowledge, the framework we have developed is the most extensive end-to-end resting-state EEG machine learning pipeline for discovery of EEG biomarkers, implementing many recently developed technologies.

We plan on contributing our EEG analysis framework in the form of code alongside the publication of the papers, so other neuroimaging researchers can utilize the framework

or draw inspiration for implementing parts of our framework. Parts of our code have also already been contributed to student projects at DTU and the development of other analysis toolboxes, e.g. MNE-Python and The Neurophysiological Biomarker Toolbox version 2 (NBT2) currently being developed in collaboration with Klaus Linkenkaer-Hansen.

The results from applying our analysis framework to the PTSD dataset revealed the PTSD group had significantly higher relative beta and gamma power and lower frontal theta/beta ratio and 1/f exponent. However, the classifiers only obtained around 63% balanced test accuracy, a moderate performance suggesting that while the EEG features were different on average between the two groups, they were not able to accurately discriminate individuals and serve as clinical biomarkers. To delineate heterogeneity within the PTSD group, we identified two subtypes, one very similar to the control group, while the other group was characterized by pronounced hyperconnectivity, most prominently in alpha wPLI. Our classifiers were able to classify the subtype with hyperconnectivity with 79% balanced test accuracy against the control group, further confirming that this subtype was markedly different from the control group. Interestingly, the subtype with hyperconnectivity displayed higher arousal than both the control group and the other subtype, and some of the alpha wPLI features deemed important by the classifiers were significantly positively correlated with arousal severity. This finding suggests that the identified alpha wPLI connectivities might serve as potential biomarkers for belonging to the subtype that is characterized by hyperconnectivity and hyperarousal, and further investigation should be considered to determine if the subtype is also associated with different treatment outcomes.

The findings from applying our analysis framework to the ASD dataset revealed that adults with ASD had EEG activity within the typical range of the non-autistic comparison group. This conclusion was further supported by the classification results, where the best classifier merely obtained around 56% accuracy. We also identified two ASD subtypes, and while we observed improved classification of the subtypes, the improvement was only moderate and we were unable to relate the subtypes to clinical behavior, thus limiting the interpretation of functional roles for the subtypes. Future work should employ higher sample sizes to better represent the whole autism spectrum, and integrating data from multiple modalities might benefit the search for biomarkers, e.g. combining genetics, neuroimaging, psychological, and social information to develop a biopsychoso-

cial model for ASD [145, 146]. Differences might also be more prominent during active tasks probing the autistic traits, e.g. interactive paradigms might reveal insights into the mechanisms underlying the interpersonal difficulties often observed in ASD [147–152].

Another avenue for future work might be on expanding the resting-state EEG feature set. Although we computed an extensive set of commonly used EEG features, many other interesting features were unfortunately not included due to time constraints. Future work should investigate if combinations of the features we computed and features from e.g. recurrence quantification analysis [153], graph theory [154, 155] or dynamic causal modelling [156–158] could serve as potential biomarkers. Most EEG studies also focus on cortical sources, due to the higher signal-to-noise ratio compared to sub-cortical areas, but interestingly, a study observed an EEG activity pattern that was correlated with fMRI amygdala activity [159]. They called the EEG activity pattern an EEG amygdala fingerprint, and it has been used as the target for a randomized double-blinded neurofeedback study in healthy soldiers and improved emotional conflict regularization and decreased alexithymia scores was observed in the group that performed neurofeedback [160]. Including the amygdala fingerprint or developing EEG fingerprints for other sub-cortical areas to be used as features could broaden the utility of EEG. Another interesting direction of recent EEG studies is the application of deep learning. The main advantage is that deep learning can be trained directly on the raw time series to learn features from the data [71, 161, 162]. Sample sizes seem to be a bottleneck for now, but this approach holds much promise as more data become available.

In conclusion, we developed an end-to-end resting-state EEG machine learning analysis framework for biomarker discovery and successfully applied it to two independent datasets. Our framework can be readily expanded to include more features and for application in other datasets and disorders. With the current fast-paced technological Big Data trends, we are optimistic about the outlook on computationally-derived biomarkers supporting clinical decisions for better tailored diagnosis and treatments for individual patients in the near future.

Bibliography

[1] American Psychiatric Association. *The Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Association, 2013, pages 1–947.

[2] World Health Organization. *International Classification of Diseases, Eleventh Revision (ICD-11)*. 11th edition. World Health Organization, 2019.

[3] S. Hong Lee et al. "Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs." In: *Nature Genetics* 45.9 (2013), pages 984–994. ISSN: 10614036. DOI: 10.1038/ng.2711.

[4] Eric Feczko et al. "The Heterogeneity problem: Approaches to identify psychiatric subtypes." In: *Trends in Cognitive Sciences* 23.7 (2019), pages 584–601. DOI: 10.1016/j.tics.2019.03.009.

[5] Thomas Insel et al. "Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders." In: *American Journal of Psychiatry Online* 167 (2010), pages 748–751. ISSN: 0002-953X.

[6] Thomas R. Insel and Bruce N. Cuthbert. "Brain disorders? Precisely: Precision medicine comes to psychiatry." In: *Science* 348.6234 (2015), pages 499–500. ISSN: 10959203. DOI: 10.1126/science.aab2358.

[7] Mohammad R. Arbabshirani et al. "Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls." In: *NeuroImage* 145 (2017), pages 137–165. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2016.02.079.

[8] Danilo Bzdok and Andreas Meyer-Lindenberg. "Machine Learning for Precision Psychiatry: Opportunities and Challenges." In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 3.3 (2018), pages 223–230. ISSN: 24519030. DOI: 10.1016/j.bpsc.2017.11.007.

[9] Barnaly Rashid and Vince Calhoun. "Towards a brain-based predictome of mental illness." In: *Human Brain Mapping* 41.12 (2020), pages 3468–3535. ISSN: 10970193. DOI: 10.1002/hbm.25013.

[10] Jennifer J. Newson and Tara C. Thiagarajan. "EEG Frequency Bands in Psychiatric Disorders: A Review of Resting State Studies." In: *Frontiers in Human Neuroscience* 12 (2019), page 521. ISSN: 1662-5161. DOI: 10.3389/fnhum.2018.00521.

[11] Lara V Marcuse, Madeline C Fields, and Jiyeoun (Jenna) Yoo. "The normal adult EEG." In: *Rowan's Primer of EEG*. 2016, pages 39–66. DOI: 10.1016/b978-0-323-35387-8.00002-0.

[12] Mamona Butt et al. "The electrical aftermath: Brain signals of posttraumatic stress disorder filtered through a clinical lens." In: *Frontiers in Psychiatry* 10 (2019), page 368. ISSN: 16640640. DOI: 10.3389/fpsyt.2019.00368.

[13] Riitta Hari and Aina Puce. *MEG-EEG Primer*. Oxford University Press, 2017.

[14] Lara V Marcuse, Madeline C Fields, and Jiyeoun (Jenna) Yoo. "Origin and technical aspects of the EEG." In: *Rowan's Primer of EEG*. 2016, pages 1–37. DOI: 10.1016/b978-0-323-35387-8.00001-9.

[15] Roberta Grech et al. "Review on solving the inverse problem in EEG source analysis." In: *Journal of NeuroEngineering and Rehabilitation* 5.1 (2008), pages 1–33. ISSN: 17430003. DOI: 10.1186/1743-0003-5-25.

[16] Christoph M. Michel and Denis Brunet. "EEG source imaging: A practical review of the analysis steps." In: *Frontiers in Neurology* 10 (2019). ISSN: 16642295. DOI: 10.3389/fneur.2019.00325.

[17] Pascal Fries. "Rhythms for Cognition: Communication through Coherence." In: *Neuron* 88.1 (2015), pages 220–235. ISSN: 10974199. DOI: 10.1016/j.neuron.2015.09.034.

[18] Wolfgang Klimesch, Paul Sauseng, and Simon Hanslmayr. *EEG alpha oscillations: The inhibition-timing hypothesis*. 2007. DOI: 10.1016/j.brainresrev.2006.06.003.

[19] John J. Foxe and Adam C. Snyder. "The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention." In: *Frontiers in Psychology* 2 (2011), page 154. ISSN: 16641078. DOI: 10.3389/fpsyg.2011.00154.

[20] Pascal Fries et al. “The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4.” In: *Journal of Neuroscience* 28.18 (2008), pages 4823–4835. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.4499-07.2008.

[21] Wolfgang Klimesch. “EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis.” In: *Brain Research Reviews* 29.2-3 (1999), pages 169–195. ISSN: 0165-0173. DOI: 10.1016/S0165-0173(98)00056-3.

[22] Thalía Harmony. “The functional significance of delta oscillations in cognitive processing.” In: *Frontiers in Integrative Neuroscience* 7 (2013), pages 1–10. ISSN: 16625145. DOI: 10.3389/fnint.2013.00083.

[23] Andreas K Engel and Pascal Fries. “Beta-band oscillations — signalling the status quo?” In: *Current Opinion in Neurobiology* 20.2 (2010), pages 156–165. ISSN: 0959-4388. DOI: 10.1016/J.CONB.2010.02.015.

[24] Kenneth R. Alper et al. “Correlation of PET and qEEG in normal subjects.” In: *Psychiatry Research - Neuroimaging* 146.3 (2006), pages 271–282. ISSN: 09254927. DOI: 10.1016/j.pscychresns.2005.06.008.

[25] Pierre Maquet. “Positron emission tomography studies of sleep and sleep disorders.” In: *J Neurol* 244 (1997), pages 23–28.

[26] Christoph S. Herrmann et al. “EEG oscillations: From correlation to causality.” In: *International Journal of Psychophysiology* 103 (2016), pages 12–21. ISSN: 18727697. DOI: 10.1016/j.ijpsycho.2015.02.003.

[27] Angelos Angelidis et al. “Do not look away! Spontaneous frontal EEG theta/beta ratio as a marker for cognitive control over attention to mild and high threat.” In: *Biological Psychology* 135 (2018), pages 8–17. ISSN: 18736246. DOI: 10.1016/j.biopsycho.2018.03.002.

[28] Robin Nusslock et al. “Cognitive Vulnerability and Frontal Brain Asymmetry: Common Predictors of First Prospective Depressive Episode.” In: *Journal of Abnormal Psychology* 120.2 (2011), pages 497–503. ISSN: 19391846. DOI: 10.1037/a0022940.

[29] Stefan Debener et al. “Is resting anterior EEG alpha asymmetry a trait marker for depression? Findings for healthy adults and clinically depressed patients.” In: *Neuropsychobiology* 41.1 (2000), pages 31–37. ISSN: 0302282X. DOI: 10.1159/000026630.

[30] Linda J. Metzger et al. “PTSD Arousal and Depression Symptoms Associated with Increased Right-Sided Parietal EEG Asymmetry.” In: *Journal of Abnormal Psychology* 113.2 (2004), pages 324–329. ISSN: 0021843X. DOI: 10.1037/0021-843X.113.2.324.

[31] Andreas Mierau, Wolfgang Klimesch, and Jérémie Lefebvre. “State-dependent alpha peak frequency shifts: Experimental evidence, potential mechanisms and functional implications.” In: *Neuroscience* 360 (2017), pages 146–154. ISSN: 18737544. DOI: 10.1016/j.neuroscience.2017.07.037.

[32] C. Richard Clark et al. “Spontaneous alpha peak frequency predicts working memory performance across the age span.” In: *International Journal of Psychophysiology* 53.1 (2004), pages 1–9. ISSN: 01678760. DOI: 10.1016/j.ijpsycho.2003.12.011.

[33] Bradley Voytek et al. “Age-related changes in 1/f neural electrophysiological noise.” In: *Journal of Neuroscience* 35.38 (2015), pages 13257–13265. ISSN: 15292401. DOI: 10.1523/JNEUROSCI.2332-14.2015.

[34] Thomas Donoghue et al. “Parameterizing neural power spectra into periodic and aperiodic components.” In: *Nature Neuroscience* 23.12 (2020), pages 1655–1665. ISSN: 15461726. DOI: 10.1038/s41593-020-00744-x.

[35] Marius Tröndle et al. “Decomposing age effects in EEG alpha power.” In: *bioRxiv* (2021), page 2021.05.26.445765. DOI: 10.1101/2021.05.26.445765.

[36] Ashley Merkin et al. “Age differences in aperiodic neural activity measured with resting EEG.” In: *bioRxiv* (2021), pages 1–30.

[37] Richard Gao, Erik J. Peterson, and Bradley Voytek. “Inferring synaptic excitation/inhibition balance from field potentials.” In: *NeuroImage* 158 (2017), pages 70–78. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2017.06.078.

[38] Hilgo Bruining et al. “Measurement of excitation-inhibition ratio in autism spectrum disorder using critical brain dynamics.” In: *Scientific Reports* 10.1 (2020), pages 1–15. ISSN: 20452322. DOI: 10.1038/s41598-020-65500-4.

[39] André M. Bastos and Jan Mathijs Schoffelen. "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls." In: *Frontiers in Systems Neuroscience* 9 (2016), page 175. ISSN: 16625137. DOI: 10.3389/fnsys.2015.00175.

[40] E. van Diessen et al. "Opportunities and methodological challenges in EEG and MEG resting state functional brain network research." In: *Clinical Neurophysiology* 126.8 (2015), pages 1468–1481. ISSN: 18728952. DOI: 10.1016/j.clinph.2014.11.018.

[41] Paul L. Nunez et al. "EEG coherency I: Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales." In: *Electroencephalography and Clinical Neurophysiology* 103.5 (1997), pages 499–515. ISSN: 00134694. DOI: 10.1016/S0013-4694(97)00066-7.

[42] Jean Philippe Lachaux et al. "Measuring phase synchrony in brain signals." In: *Human Brain Mapping* 8.4 (1999), pages 194–208. ISSN: 10659471. DOI: 10.1002/(SICI)1097-0193(1999)8:4<194::AID-HBM4>3.0.CO;2-C.

[43] Guido Nolte et al. "Identifying true brain interaction from EEG data using the imaginary part of coherency." In: *Clinical Neurophysiology* 115.10 (2004), pages 2292–2307. ISSN: 13882457. DOI: 10.1016/j.clinph.2004.04.029.

[44] Cornelis J. Stam, Guido Nolte, and Andreas Daffertshofer. "Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources." In: *Human Brain Mapping* 28.11 (2007), pages 1178–1193. ISSN: 10659471. DOI: 10.1002/hbm.20346.

[45] Joerg F. Hipp et al. "Large-scale cortical correlation structure of spontaneous oscillatory activity." In: *Nature Neuroscience* 15.6 (2012), pages 884–890. ISSN: 10976256. DOI: 10.1038/nn.3101.

[46] Russell T. Toll et al. "An Electroencephalography Connectomic Profile of Posttraumatic Stress Disorder." In: *The American journal of psychiatry* 177.3 (2020), pages 233–243. ISSN: 15357228. DOI: 10.1176/appi.ajp.2019.18080911.

[47] Mingzhou Ding, Yonghong Chen, and Steven L. Bressler. "Granger Causality: Basic Theory and Application to Neuroscience." In: *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications* (2006), pages 437–460. DOI: 10.1002/9783527609970.ch17. arXiv: 0608035v1 [q-bio].

[48] Adam B. Barrett et al. “Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia.” In: *PLoS ONE* 7.1 (2012), pages 1–12. ISSN: 19326203. DOI: 10.1371/journal.pone.0029072.

[49] Anil K. Seth, Adam B. Barrett, and Lionel Barnett. “Granger causality analysis in neuroscience and neuroimaging.” In: *Journal of Neuroscience* 35.8 (2015), pages 3293–3297. ISSN: 15292401. DOI: 10.1523/JNEUROSCI.4399-14.2015.

[50] E. Damaraju et al. “Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia.” In: *NeuroImage: Clinical* 5 (2014), pages 298–308. ISSN: 22131582. DOI: 10.1016/j.nicl.2014.07.003.

[51] Klaus Linkenkaer-Hansen et al. “Long-range temporal correlations and scaling behavior in human brain oscillations.” In: *Journal of Neuroscience* 21.4 (2001), pages 1370–1377. ISSN: 02706474. DOI: 10.1523/jneurosci.21-04-01370.2001.

[52] Arthur-Ervin Avramiea et al. “Amplitude and phase coupling optimize information transfer between brain networks that function at criticality.” In: *bioRxiv* (2021), page 2021.03.15.435461. DOI: 10.1101/2021.03.15.435461.

[53] Dirk J.A. Smit et al. “Scale-free modulation of resting-state neuronal oscillations reflects prolonged brain maturation in humans.” In: *Journal of Neuroscience* 31.37 (2011), pages 13128–13136. ISSN: 02706474. DOI: 10.1523/JNEUROSCI.1678-11.2011.

[54] Huibin Jia and Dongchuan Yu. “Attenuated long-range temporal correlations of electrocortical oscillations in patients with autism spectrum disorder.” In: *Developmental Cognitive Neuroscience* 39 (2019), page 100687. ISSN: 18789307. DOI: 10.1016/j.dcn.2019.100687.

[55] Bahareh Rahmani et al. “Dynamical hurst analysis identifies EEG channel differences between PTSD and healthy controls.” In: *PLoS ONE* 13.7 (2018), e0199144. ISSN: 19326203. DOI: 10.1371/journal.pone.0199144.

[56] D. Lehmann, H. Ozaki, and I. Pal. “EEG alpha map series: brain micro-states by space-oriented adaptive segmentation.” In: *Electroencephalography and Clinical Neurophysiology* 67.3 (1987), pages 271–288. ISSN: 00134694. DOI: 10.1016/0013-4694(87)90025-3.

[57] Arjun Khanna et al. *Microstates in resting-state EEG: Current status and future directions*. 2015. DOI: 10.1016/j.neubiorev.2014.12.010.

[58] Janir Ramos da Cruz et al. "EEG microstates are a candidate endophenotype for schizophrenia." In: *Nature Communications* 11.1 (2020), pages 1–11. ISSN: 20411723. DOI: 10.1038/s41467-020-16914-1.

[59] Huibin Jia and Dongchuan Yu. "Aberrant Intrinsic Brain Activity in Patients with Autism Spectrum Disorder: Insights from EEG Microstates." In: *Brain Topography* 32.2 (2019), pages 295–303. ISSN: 15736792. DOI: 10.1007/s10548-018-0685-0.

[60] Bilal Mirza et al. "Machine learning and integrative analysis of biomedical big data." In: *Genes* 10.2 (2019). ISSN: 20734425. DOI: 10.3390/genes10020087.

[61] Gyeongcheol Cho et al. "Review of machine learning algorithms for diagnosing mental illness." In: *Psychiatry Investigation* 16.4 (2019), pages 262–269. ISSN: 19763026. DOI: 10.30773/pi.2018.12.21.2.

[62] Quentin J.M. Huys, Tiago V. Maia, and Michael J. Frank. "Computational psychiatry as a bridge from neuroscience to clinical applications." In: *Nature Neuroscience* 19.3 (2016), pages 404–413. ISSN: 15461726. DOI: 10.1038/nn.4238.

[63] Naomi Altman and Martin Krzywinski. "The curse(s) of dimensionality." In: *Nature Methods* 15.6 (2018), pages 399–400. ISSN: 15487105. DOI: 10.1038/s41592-018-0019-x.

[64] Yvan Saeys, Iñaki Inza, and Pedro Larrañaga. "A review of feature selection techniques in bioinformatics." In: *Bioinformatics* 23.19 (2007), pages 2507–2517. ISSN: 13674803. DOI: 10.1093/bioinformatics/btm344.

[65] L Breiman. "Statistical Modeling: The Two Cultures." In: *Statistical Science* 16.3 (2001), pages 199–231. ISSN: 0889-5406.

[66] Xiang Li, Ning Guo, and Quanzheng Li. "Functional Neuroimaging in the New Era of Big Data." In: *Genomics, Proteomics and Bioinformatics* 17.4 (2019), pages 393–401. ISSN: 22103244. DOI: 10.1016/j.gpb.2018.11.005.

[67] Christopher M. Bishop. *Pattern Recognition and Machine Learning*. Springer, 2006, pages 1–738.

[68] Laurens van der Maaten and Geoffrey Hinton. "Visualizing Data using t-SNE." In: *Journal of Machine Learning Research* 9 (2008), pages 2579–2605. ISSN: 15729338. DOI: 10.1007/s10479-011-0841-3.

[69] Mahan Hosseini et al. “I tried a bunch of things: The dangers of unexpected overfitting in classification of brain data.” In: *Neuroscience and Biobehavioral Reviews* 119 (2020), pages 456–467. ISSN: 18737528. DOI: 10.1016/j.neubiorev.2020.09.036.

[70] Daniela M Witten and Robert Tibshirani. “A framework for feature selection.” In: *American Statistician* 105.490 (2010), pages 713–726. ISSN: 0162-1459. DOI: 10.1198/jasa.2010.tm09415.A.

[71] F Lotte et al. “A review of classification algorithms for EEG-based brain-computer interfaces: A 10 year update.” In: *Journal of Neural Engineering* 15.3 (2018), page 031005. ISSN: 17412552. DOI: 10.1088/1741-2552/aab2f2.

[72] Isabelle Guyon, Jason Weston, and Stephen Barnhill. “Gene Selection for Cancer Classification using Support Vector Machines.” In: *Machine Learning* 46 (2002), pages 389–422.

[73] Leo Breiman. “Random Forests.” In: *Machine Learning* 45 (2001), pages 5–32.

[74] Anne Laure Boulesteix and Matthias Schmid. “Machine learning versus statistical modeling.” In: *Biometrical Journal* 56.4 (2014), pages 588–593. ISSN: 15214036. DOI: 10.1002/bimj.201300226.

[75] Russell A Poldrack, Grace Huckins, and Gael Varoquaux. *Establishment of Best Practices for Evidence for Prediction: A Review*. 2020. DOI: 10.1001/jamapsychiatry.2019.3671.

[76] Eva Loth et al. “The meaning of significant mean group differences for biomarker discovery.” In: *PLoS computational biology* 17.11 (2021), e1009477. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1009477.

[77] Steven Lemm et al. “Introduction to machine learning for brain imaging.” In: *NeuroImage* 56.2 (2011), pages 387–399. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2010.11.004.

[78] Mainak Jas et al. “Autoreject: Automated artifact rejection for MEG and EEG data.” In: *NeuroImage* 159.December 2016 (2017), pages 417–429. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2017.06.030.

[79] Irene Winkler et al. "On the influence of high-pass filtering on ICA-based artifact reduction in EEG-ERP." In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS* 2015-Novem (2015), pages 4101–4105. ISSN: 1557170X. DOI: 10.1109/EMBC.2015.7319296.

[80] Dražen Begić, Ljubomir Hotujac, and Nataša Jokić-Begić. "Electroencephalographic comparison of veterans with combat-related post-traumatic stress disorder and healthy subjects." In: *International Journal of Psychophysiology* 40.2 (2001), pages 167–172. ISSN: 0167-8760. DOI: 10.1016/S0167-8760(00)00153-7.

[81] Cindy L. Ehlers et al. "Electrophysiological responses to affective stimuli in American Indians experiencing trauma with and without PTSD." In: *Annals of the New York Academy of Sciences* 1071.1 (2006), pages 125–136. ISSN: 17496632. DOI: 10.1196/annals.1364.011.

[82] Claudio Imperatori et al. "Aberrant EEG functional connectivity and EEG power spectra in resting state post-traumatic stress disorder: A sLORETA study." In: *Biological Psychology* 102.1 (2014), pages 10–17. ISSN: 18736246. DOI: 10.1016/j.biopsycho.2014.07.011.

[83] Doran Todder et al. "The quantitative electroencephalogram and the low-resolution electrical tomographic analysis in posttraumatic stress disorder." In: *Clinical EEG and Neuroscience* 43.1 (2012), pages 48–53. ISSN: 15500594. DOI: 10.1177/1550059411428716.

[84] Helané Wahbeh and Barry S. Oken. "Peak high-frequency HRV and peak alpha frequency higher in PTSD." In: *Applied Psychophysiology Biofeedback* 38.1 (2013), pages 57–69. ISSN: 10900586. DOI: 10.1007/s10484-012-9208-z.

[85] Sun Young Moon et al. "Increased frontal gamma and posterior delta powers as potential neurophysiological correlates differentiating posttraumatic stress disorder from anxiety disorders." In: *Psychiatry Investigation* 15.11 (2018), pages 1087–1093. ISSN: 19763026. DOI: 10.30773/pi.2018.09.30.

[86] Kevin Clancy et al. "Restless 'rest': Intrinsic sensory hyperactivity and disinhibition in post-traumatic stress disorder." In: *Brain* 140.7 (2017), pages 2041–2050. ISSN: 14602156. DOI: 10.1093/brain/awx116.

[87] Melinda D Veltmeyer, Alexander C Mcfarlane, and Richard A Bryant. "Integrative assessment of brain function in PTSD: Brain stability and working memory." In: *Journal of Integrative Neuroscience* 5.1 (2006), pages 123–138.

[88] Sirko Rabe et al. "Regional brain electrical activity in posttraumatic stress disorder after motor vehicle accident." In: *Journal of Abnormal Psychology* 115.4 (2006), pages 687–698. ISSN: 0021843X. DOI: 10.1037/0021-843X.115.4.687.

[89] Stewart A. Shankman et al. "Resting electroencephalogram asymmetry and posttraumatic stress disorder." In: *Journal of Traumatic Stress* 21.2 (2008), pages 190–198. ISSN: 08949867. DOI: 10.1002/jts.20319.

[90] Evian Gordon, Donna M. Palmer, and Nicholas Cooper. "EEG alpha asymmetry in schizophrenia, depression, PTSD, panic disorder, ADHD and conduct disorder." In: *Clinical EEG and Neuroscience* 41.4 (2010), pages 178–183. ISSN: 15500594. DOI: 10.1177/155005941004100404.

[91] Erin M Falconer et al. "Developing an integrated brain, behavior and biological response profile in posttraumatic stress disorder (PTSD)." In: *Journal of Integrative Neuroscience* 7.3 (2008), pages 439–456. ISSN: 02196352. DOI: 10.1142/S0219635208001873.

[92] Nataša Jokić-Begić and Dražen Begić. "Quantitative electroencephalogram (qEEG) in combat veterans with post-traumatic stress disorder (PTSD)." In: *Nordic Journal of Psychiatry* 57.5 (2003), pages 351–355. ISSN: 08039488. DOI: 10.1080/08039480310002688.

[93] Matteo Fraschini et al. "The effect of epoch length on estimated EEG functional connectivity and brain network organisation." In: *Journal of Neural Engineering* 13.3 (2016). ISSN: 17412552. DOI: 10.1088/1741-2560/13/3/036015.

[94] Nima Bigdely-Shamlo et al. "The PREP pipeline: Standardized preprocessing for large-scale EEG analysis." In: *Frontiers in Neuroinformatics* 9.JUNE (2015), pages 1–19. ISSN: 16625196. DOI: 10.3389/fninf.2015.00016.

[95] F Perrin, J Pernier, and O Bertrand. "Spherical splines for scalp potential and current density mapping." In: *Electroencephalograp and clinical Neurophysiology* (1989), pages 184–187.

[96] Richard Hardstone et al. "Detrended fluctuation analysis: A scale-free view on neuronal oscillations." In: *Frontiers in Physiology* (2012), pages 1–13. ISSN: 1664042X. DOI: 10.3389/fphys.2012.00450.

[97] Hanchuan Peng, Fuhui Long, and Chris Ding. "Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy." In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (2005), pages 1226–1238. ISSN: 01628828. DOI: 10.1109/TPAMI.2005.159.

[98] Mohammad-Parsa Hosseini, Amin Hosseini, and Kiarash Ahi. "A Review on Machine Learning for EEG Signal Processing in Bioengineering." In: *IEEE Reviews in Biomedical Engineering* (2020). ISSN: 1937-3333. DOI: 10.1109/rbme.2020.2969915.

[99] Yu Zhang et al. "Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography." In: *Nature Biomedical Engineering* (2020). ISSN: 2157846X. DOI: 10.1038/s41551-020-00614-8.

[100] Robert Tibshirani, Guenther Walther, and Trevor Hastie. "Estimating the number of clusters in a data set via the gap statistic." In: *J. R. Statist. Soc. B* 63 (2001), pages 411–423. ISSN: 01676423. DOI: 10.1111/1467-9868.00293.

[101] Robert J. Fenster et al. "Brain circuit dysfunction in post-traumatic stress disorder: from mouse to man." In: *Nature Reviews Neuroscience* 19.9 (2018), pages 535–551. ISSN: 14710048. DOI: 10.1038/s41583-018-0039-7.

[102] Isaac R. Galatzer-Levy and Richard A. Bryant. "636,120 Ways to Have Posttraumatic Stress Disorder." In: *Perspectives on Psychological Science* 8.6 (2013), pages 651–662. ISSN: 17456916. DOI: 10.1177/1745691613504115.

[103] Yong-wook Kim et al. "Riemannian classifier enhances the accuracy of machine-learning-based diagnosis of PTSD using resting EEG." In: *Progress in Neuropsychopharmacology & Biological Psychiatry* 102 (2020), page 109960. ISSN: 0278-5846. DOI: 10.1016/j.pnpbp.2020.109960.

[104] Braeden A. Terpou et al. "EEG microstates in PTSD: Using machine learning to identify neuromarkers." In: *SSRN* (2022). DOI: http://dx.doi.org/10.2139/ssrn.4061516.

[105] Frank W. Weathers et al. “The PTSD Checklist (PCL): Reliability, Validity, and Diagnostic Utility.” In: *Paper presented at the Annual Convention of the International Society for Traumatic Stress Studies, San Antonio, TX* 2 (1993), pages 90–92.

[106] American Psychiatric Association. *The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. American Psychiatric Association, 1994, pages 1–358.

[107] Karen Inge Karstoft et al. “Diagnostic accuracy of the posttraumatic stress disorder checklist: Civilian version in a representative military sample.” In: *Psychological Assessment* 26.1 (2014), pages 321–325. ISSN: 10403590. DOI: 10.1037/a0034889.

[108] Christophe Destrieux et al. “Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature.” In: *Neuroimage* 15.53 (2010), pages 1–15. ISSN: 0022-5061. DOI: 10.1002/jhbs.20162.

[109] Jona Sassenhagen and Dejan Draschkow. “Cluster-based permutation tests of MEG/EEG data do not establish significance of effect latency or location.” In: *Psychophysiology* 56.6 (2019), pages 1–8. ISSN: 14698986. DOI: 10.1111/psyp.13335.

[110] Stephanie Noble, Dustin Scheinost, and R. Todd Constable. “Cluster failure or power failure? Evaluating sensitivity in cluster-level inference.” In: *NeuroImage* 209 (2020). ISSN: 10959572. DOI: 10.1016/j.neuroimage.2019.116468.

[111] Stephen M. Smith and Thomas E. Nichols. “Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference.” In: *NeuroImage* 44.1 (2009), pages 83–98. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2008.03.061.

[112] C. R. Pernet et al. “Cluster-based computational methods for mass univariate analyses of event-related brain potentials/fields: A simulation study.” In: *Journal of Neuroscience Methods* 250 (2015), pages 85–93. ISSN: 1872678X. DOI: 10.1016/j.jneumeth.2014.08.003.

[113] D’Croz-Baron DF et al. “EEG Microstates Analysis in Young Adults With Autism Spectrum Disorder During Resting-State.” In: *Frontiers in human neuroscience* 13 (2019). ISSN: 1662-5161. DOI: 10.3389/FNHUM.2019.00173.

[114] F. von Wegner, E. Tagliazucchi, and H. Laufs. "Information-theoretical analysis of resting state EEG microstate sequences - non-Markovianity, non-stationarity and periodicities." In: *NeuroImage* 158 (2017), pages 99–111. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2017.06.062.

[115] Guy C. Van Orden, John G. Holden, and Michael T. Turvey. "Self-organization of cognitive performance." In: *Journal of Experimental Psychology: General* 132.3 (2003), pages 331–350. ISSN: 00963445. DOI: 10.1037/0096-3445.132.3.331.

[116] Guy C. Van Orden, John G. Holden, and Michael T. Turvey. "Human cognition and 1/f scaling." In: *Journal of Experimental Psychology: General* 134.1 (2005), pages 117–123. ISSN: 00963445. DOI: 10.1037/0096-3445.134.1.117.

[117] Guy C. Van Orden, Heidi Kloos, and Sebastian Wallot. *Living in the Pink. Intentionality, Wellbeing, and Complexity*. Volume 10. Elsevier B.V., 2011, pages 629–672. ISBN: 9780444520760. DOI: 10.1016/B978-0-444-52076-0.50022-5.

[118] Christopher T. Kello et al. *Scaling laws in cognitive sciences*. 2010. DOI: 10.1016/j.tics.2010.02.005.

[119] Brendan Ostlund et al. "Spectral parameterization for studying neurodevelopment: How and why." In: *Developmental Cognitive Neuroscience* 54 (2022), page 101073. ISSN: 18789307. DOI: 10.1016/j.dcn.2022.101073.

[120] Thomas Donoghue, Julio Dominguez, and Bradley Voytek. "Electrophysiological Frequency Band Ratio Measures Conflate Periodic and Aperiodic Neural Activity." In: *eNeuro* 7.6 (2020). DOI: 10.1523/ENEURO.0192-20.2020.

[121] Anna J Finley et al. "Periodic and aperiodic contributions to theta-beta ratios across adulthood." In: *Psychophysiology* (2022), pages 1–20. DOI: 10.1111/psyp.14113.

[122] Madeline M. Robertson et al. "EEG power spectral slope differs by ADHD status and stimulant medication exposure in early childhood." In: *Journal of Neurophysiology* 122.6 (2019), pages 2427–2437. ISSN: 15221598. DOI: 10.1152/jn.00388.2019.

[123] Juan L. Molina et al. "Memantine Effects on Electroencephalographic Measures of Putative Excitatory/Inhibitory Balance in Schizophrenia." In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 5.6 (2020), pages 562–568. ISSN: 24519030. DOI: 10.1016/j.bpsc.2020.02.004.

[124] Jun Wang et al. “Resting state EEG abnormalities in autism spectrum disorders.” In: *Journal of Neurodevelopmental Disorders* 5.1 (2013), page 1. ISSN: 1866-1947. DOI: 10.1186/1866-1955-5-24.

[125] Christian O’Reilly, John D. Lewis, and Mayada Elsabbagh. “Is functional brain connectivity atypical in autism? A systematic review of EEG and MEG studies.” In: *PLoS ONE* 12.5 (2017), e0175870. ISSN: 19326203. DOI: 10.1371/journal.pone.0175870.

[126] Stefan Holiga et al. “Patients with autism spectrum disorders display reproducible functional connectivity alterations.” In: *Science Translational Medicine* 11.481 (2019). ISSN: 19466242. DOI: 10.1126/scitranslmed.aat9223.

[127] Jiannan Kang et al. “The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data.” In: *Computers in Biology and Medicine* 120 (2020). ISSN: 18790534. DOI: 10.1016/j.compbiomed.2020.103722.

[128] Shasha Zhang et al. “Children ASD Evaluation Through Joint Analysis of EEG and Eye-Tracking Recordings With Graph Convolution Network.” In: *Frontiers in Human Neuroscience* 15 (2021), page 651349. ISSN: 16625161. DOI: 10.3389/fnhum.2021.651349.

[129] Ali Sheikhani et al. “Detection of abnormalities for diagnosing of children with autism disorders using of quantitative electroencephalography analysis.” In: *Journal of Medical Systems* 36.2 (2012), pages 957–963. ISSN: 1573689X. DOI: 10.1007/s10916-010-9560-6.

[130] Enzo Grossi, Chiara Olivieri, and Massimo Buscema. “Diagnosis of autism through EEG processed by advanced computational algorithms: A pilot study.” In: *Computer Methods and Programs in Biomedicine* 142 (2017), pages 73–79. ISSN: 18727565. DOI: 10.1016/j.cmpb.2017.02.002.

[131] Jie Zhao et al. “A study on EEG feature extraction and classification in autistic children based on singular spectrum analysis method.” In: *Brain and Behavior* 10.12 (2020), e01721. ISSN: 21623279. DOI: 10.1002/brb3.1721.

[132] Shixin Peng et al. "Early Screening of Children With Autism Spectrum Disorder Based on Electroencephalogram Signal Feature Selection With L1-Norm Regularization." In: *Frontiers in Human Neuroscience* 15 (2021), page 656578. DOI: 10.3389/fnhum.2021.656578.

[133] Ridha Djemal et al. "EEG-Based computer aided diagnosis of autism spectrum disorder using wavelet, entropy, and ANN." In: *BioMed Research International* 2017 (2017). DOI: 10.1155/2017/9816591.

[134] Netherlands Autism Register. *Netherlands Autism Register*. 2022.

[135] Rosa A. Hoekstra et al. "The construction and validation of an abridged version of the autism-spectrum quotient (AQ-short)." In: *Journal of Autism and Developmental Disorders* 41.5 (2011), pages 589–596. ISSN: 01623257. DOI: 10.1007/s10803-010-1073-0.

[136] Teresa Tavassoli, Rosa A. Hoekstra, and Simon Baron-Cohen. "The Sensory Perception Quotient (SPQ): Development and validation of a new sensory questionnaire for adults with and without autism." In: *Molecular Autism* 5.1 (2014), pages 1–10. ISSN: 20402392. DOI: 10.1186/2040-2392-5-29.

[137] Ricarda F. Weiland et al. "The Dutch Sensory Perception Quotient-Short in adults with and without autism." In: *Autism* 24.8 (2020), pages 2071–2080. ISSN: 14617005. DOI: 10.1177/1362361320942085.

[138] Pilar Garcés et al. "Resting state EEG power spectrum and functional connectivity in autism: a cross-sectional analysis." In: *Molecular Autism* 13.1 (2022), pages 1–16. DOI: 10.1186/s13229-022-00500-x.

[139] Karen J. Mathewson et al. "Regional EEG alpha power, coherence, and behavioral symptomatology in autism spectrum disorder." In: *Clinical Neurophysiology* 123.9 (2012), pages 1798–1809. ISSN: 13882457. DOI: 10.1016/j.clinph.2012.02.061.

[140] Michael Murias et al. "Resting State Cortical Connectivity Reflected in EEG Coherence in Individuals With Autism." In: *Biological Psychiatry* 62.3 (2007), pages 270–273. ISSN: 00063223. DOI: 10.1016/j.biopsych.2006.11.012.

[141] A. Saunders, I. J. Kirk, and K. E. Waldie. "Hemispheric Coherence in ASD with and without Comorbid ADHD and Anxiety." In: *BioMed Research International* 2016 (2016). ISSN: 23146141. DOI: 10.1155/2016/4267842.

[142] Michael V. Lombardo, Meng Chuan Lai, and Simon Baron-Cohen. "Big data approaches to decomposing heterogeneity across the autism spectrum." In: *Molecular Psychiatry* 24.10 (2019), pages 1435–1450. ISSN: 14765578. DOI: 10.1038/s41380-018-0321-0.

[143] Choong Wan Woo et al. "Building better biomarkers: Brain models in translational neuroimaging." In: *Nature Neuroscience* 20.3 (2017), pages 365–377. ISSN: 15461726. DOI: 10.1038/nn.4478.

[144] Seok Jun Hong et al. *Toward Neurosubtypes in Autism.* 2020. DOI: 10.1016/j.biopsych.2020.03.022.

[145] Nancy Doyle. "Neurodiversity at work: a biopsychosocial model and the impact on working adults." In: *British Medical Bulletin* 135.1 (2020), page 108. ISSN: 14718391. DOI: 10.1093/BMB/LDAA021.

[146] Cristina Panisi and Marina Marini. "Dynamic and Systemic Perspective in Autism Spectrum Disorders: A Change of Gaze in Research Opens to A New Landscape of Needs and Solutions." In: *Brain Sciences* 12.2 (2022). ISSN: 20763425. DOI: 10.3390/BRAINSCI12020250.

[147] Hiroki C. Tanabe et al. "Hard to "tune in": Neural mechanisms of live face-to-face interaction with high-functioning autistic spectrum disorder." In: *Frontiers in Human Neuroscience* 6 (2012), page 268. ISSN: 16625161. DOI: 10.3389/fnhum.2012.00268.

[148] Leonhard Schilbach et al. "Toward a second-person neuroscience." In: *Behavioral and Brain Sciences* 36.4 (2013), pages 393–414. ISSN: 14691825. DOI: 10.1017/S0140525X12000660.

[149] Victoria Leong and Leonhard Schilbach. "The promise of two-person neuroscience for developmental psychiatry: using interaction-based sociometrics to identify disorders of social interaction." In: *The British Journal of Psychiatry* 215.5 (2019), pages 636–638. ISSN: 0007-1250. DOI: 10.1192/BJP.2019.73.

[150] Yafeng Pan and Xiaojun Cheng. "Two-Person Approaches to Studying Social Interaction in Psychiatry: Uses and Clinical Relevance." In: *Frontiers in Psychiatry* 11 (2020), page 301. ISSN: 16640640. DOI: 10.3389/FPSYT.2020.00301/BIBTEX.

[151] Jana A. Kruppa et al. “Brain and motor synchrony in children and adolescents with ASD - A fNIRS hyperscanning study.” In: *Social Cognitive and Affective Neuroscience* 16.1-2 (2021), pages 103–116. ISSN: 17495024. DOI: 10.1093/scan/nsaa092.

[152] Leonhard Schilbach. “Autism and other disorders of social interaction: where we are and where to go from here.” In: *European Archives of Psychiatry and Clinical Neuroscience* 272.2 (2022), pages 173–175. ISSN: 1433-8491. DOI: 10.1007/S00406-022-01391-Y.

[153] T. Heunis et al. “Recurrence quantification analysis of resting state EEG signals in autism spectrum disorder – a systematic methodological exploration of technical and demographic confounders in the search for biomarkers.” In: *BMC Medicine* 16.1 (2018), pages 1–17. ISSN: 17417015. DOI: 10.1186/s12916-018-1086-7.

[154] Martin Hardmeier et al. “Reproducibility of functional connectivity and graph measures based on the phase lag index (PLI) and weighted phase lag index (wPLI) derived from high resolution EEG.” In: *PLoS ONE* 9.10 (2014). ISSN: 19326203. DOI: 10.1371/journal.pone.0108648.

[155] M Shim, C. H. Im, and S. H. Lee. “Disrupted cortical brain network in post-traumatic stress disorder patients: a resting-state electroencephalographic study.” In: *Translational psychiatry* 7.9 (2017), e1231. ISSN: 21583188. DOI: 10.1038/tp.2017.200.

[156] K. E. Stephan et al. “Ten simple rules for dynamic causal modeling.” In: *NeuroImage* 49.4 (2010), pages 3099–3109. ISSN: 10538119. DOI: 10.1016/j.neuroimage.2009.11.015.

[157] Karl Friston, Rosalyn Moran, and Anil K. Seth. “Analysing connectivity with Granger causality and dynamic causal modelling.” In: *Current Opinion in Neurobiology* 23.2 (2013), pages 172–178. ISSN: 09594388. DOI: 10.1016/j.conb.2012.11.010.

[158] Frederik Van de Steen et al. “Dynamic causal modelling of fluctuating connectivity in resting-state EEG.” In: *NeuroImage* 189.May 2018 (2019), pages 476–484. ISSN: 10959572. DOI: 10.1016/j.neuroimage.2019.01.055.

[159] Yehudit Meir-Hasson et al. “An EEG Finger-Print of fMRI deep regional activation.” In: *NeuroImage* 102 (2014), pages 128–141. ISSN: 1053-8119. DOI: 10.1016/J.NEUROIMAGE.2013.11.004.

[160] Jackob N. Keynan et al. “Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience.” In: *Nature Human Behaviour* 3.1 (2019), pages 63–73. ISSN: 23973374. DOI: 10.1038/s41562-018-0484-3.

[161] Mathias Perslev, Michael Hejselbak Jensen, and Christian Igel. “U-Time : A Fully Convolutional Network for Time Series Segmentation Applied to Sleep Staging.” In: *Neural Information Processing Systems Conference* 32 (2019), pages 1–19.

[162] Demetres Kostas, Stéphane Aroca-Ouellette, and Frank Rudzicz. “BENDR: Using Transformers and a Contrastive Self-Supervised Learning Task to Learn From Massive Amounts of EEG Data.” In: *Frontiers in Human Neuroscience* 15 (2021), page 653659. ISSN: 16625161. DOI: 10.3389/fnhum.2021.653659.

APPENDIX A

PTSD Paper

Title: Resting-state EEG functional connectivity predicts post-traumatic stress disorder subtypes in veterans

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

Journal: Journal of Neural Engineering

Status: Under review

RESTING-STATE EEG FUNCTIONAL CONNECTIVITY PREDICTS POST-TRAUMATIC STRESS DISORDER SUBTYPES IN VETERANS

Qianliang Li[1*], Maya Coulson Theodorsen[1,2,3], Ivana Konvalinka[1], Kasper Eskelund[2,3], Karen-Inge Karstoft[3,4], Søren Bo Andersen[2,3] and Tobias S. Andersen[1]

[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,

[*]Corresponding author. Email: glia@dtu.dk

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 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 five minutes of eyes-closed and five minutes 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.

***Keywords*:** Functional Connectivity, Machine Learning, PTSD, Resting-state EEG, Subtypes.

1 Introduction

Post-traumatic stress disorder (PTSD) is a debilitating psychiatric disorder characterized by chronic stress symptoms following exposure to trauma. Specifically, the presence of four different clusters of symptoms have to be present for the clinical diagnosis: 1) intrusion and re-experiencing of the traumatic event, 2) avoidance symptoms, 3) negative alterations of mood and cognition and 4) 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 fulfill 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 (SSRI) have proven superior to placebo in reduction of PTSD-symptoms, effect sizes are generally small, and hence, the positive effect is modest (Hoskins et al. 2021; DePierro et al. 2019). 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; Imperatori et al. 2014; Kemp et al. 2010; Wahbeh and Oken 2013; Moon et al. 2018; Rabe et al. 2006; Shankman et al. 2008). 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. 2020).

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 (Clancy et al. 2017) and reduced theta orthogonalized power envelope correlations (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], power envelope correlations [PEC], and granger causality [GC]), and features that capture the temporal dynamics of EEG (microstates and long-range temporal correlations [DFA exponent]). Feature selection and machine learning classification were applied to investigate how well veterans with probable PTSD (defined by a score of 44 or above on the PCL for DSM-IV;

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 crossvalidation (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).

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 the original cohort (Study 1) or presented at the Military Psychology Department at the Danish Defense (MPD; Studies 2 and 3) were eligible for inclusion. For each study, specific inclusion criteria apply, see below.

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, while 36 were not (defined as PCL-score $\geq$ 44; Karstoft et al. 2014). 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 behavior. 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 over a x-year period. Overall, 61 were considered probable PTSD cases (defined as PCL-score $\geq$ 44) while 30 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 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 to 5, avoidance score was the sum of items 6 to 12 and the arousal severity score was the sum of items 13 to 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,

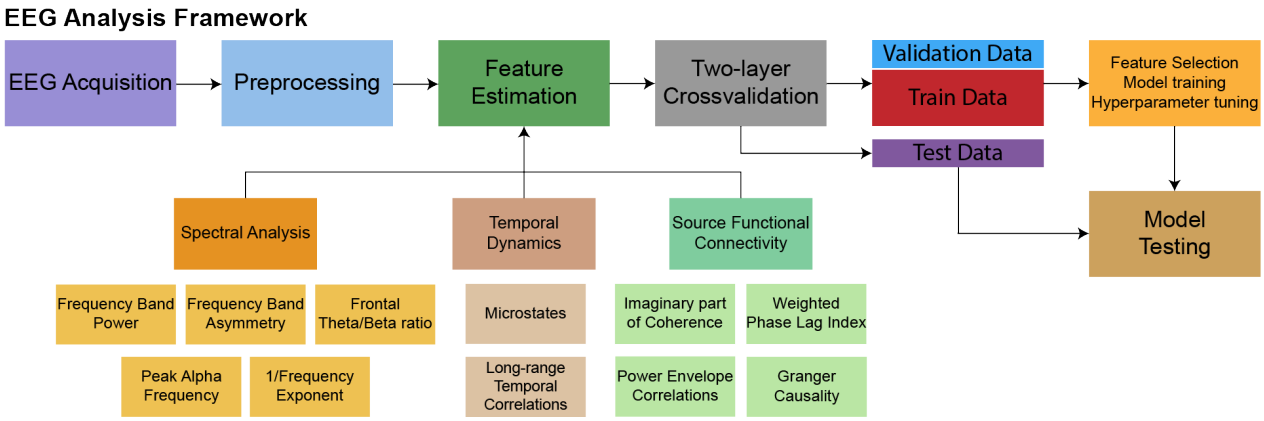

Fig. 1. **Overview of the EEG analysis framework.** After EEG acquisition, the data were preprocessed, and commonly used resting-state EEG features were computed. Repeated 10-by-10 two-layer cross-validation was performed to split the data. Feature selection, model training and hyperparamter tuning were 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.

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 band-pass 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, P8, 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 band-pass 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 10 interleaved intervals of eyes-open and eyes-closed conditions, with each interval lasting 70 seconds. The first 10 seconds were universally rejected, leaving 60 seconds 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 center 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 to 100 Hz, and a notch filter was applied at 50 Hz to remove power-line noise. After filtering, the data were divided in 4 second 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 reference (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 compo-

nents set to the number of non-interpolated channels (Bell and Sejnowski 1995; Makeig et al. 1996). 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 preprocessing 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 FOOOF algorithm (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 detrended fluctuation analysis (DFA) according to the procedure described by Hardstone et al. 2012. Nolds 0.5.2 (Schölzel 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 cross-validation criterion (Pascual-Marqui et al. 1995), we determined 4 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. 2020, 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 20484 vertices, followed by extraction of time series from 31 cortical regions of interest (ROI). The Destrieux Atlas (Destrieux et al. 2010) parcellation was manually modified to approximate the 31 ROIs used by Zhang 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 lead/lag behind the other signal. Similar to the idea behind Imcoh, the weighted phase lag index (wPLI) was developed to further attenuate the effect of volume conduction by weighing each phase difference with the magnitude of the lag (Vinck et al. 2011; Hardmeier et al. 2014).

Power envelope correlations (PEC) were estimated fol-

lowing 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 applied.

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

2.6 Cross-validation

A group stratified nested two-layer cross-validation (CV) scheme with 10-fold outer CV and 10-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 10 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 (10 repetitions of 10-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 24458 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, minimum 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): 1) linear SVM with recursive feature elimination (Guyon et al. 2002), 2) logistic regression with sequential forward selection, and 3) 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 (Pedrogosa et al. 2011) for implementation of the 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, 9 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.

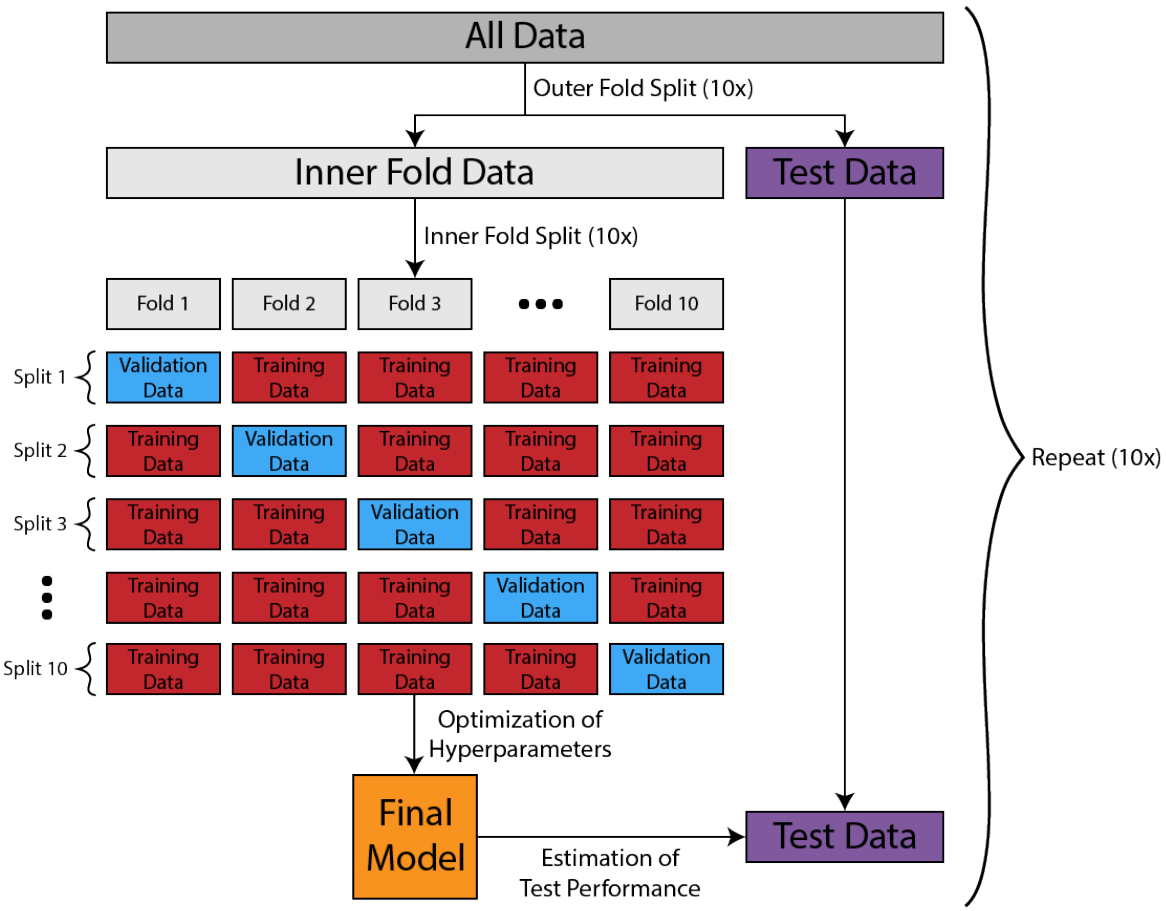

Fig. 2. **Schematic of the repeated two-layer cross-validation scheme.** In order to obtain robust and reliable estimations of how well the classifiers would perform on unseen data, we employed 10 repetitions of 10-by-10 two-layer cross-validation. The outer fold split ensured we tested on unseen data, 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.

2.8 Subtyping

We adapted the clustering methodology from Zhang et al. 2020, 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 (Witten and Tibshirani 2010; Tibshirani et al. 2001). Specifically, we employed a grid-search approach where we estimated the gap statistic for all combinations of K between 2 to 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 2 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 2 clusters (Witten and Tibshirani 2010). Pysparcl 1.4.1 (Tsumeruso 2019) 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 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 3A). 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 3B). 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 3A), 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 1/f exponent. We employed t-Distributed Stochastic Neighbor Embedding (t-SNE; van der Maaten and Hinton 2008) on the consistently selected features, and 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 3C).

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 4A). This difference was frequency specific, with the alpha band being the most prominent band (Figure 4B), 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 4C). The same connectivity patterns described for the alpha band were also observed in the theta and beta band (Supplementary Figure D1 and D2). 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 (Fig-

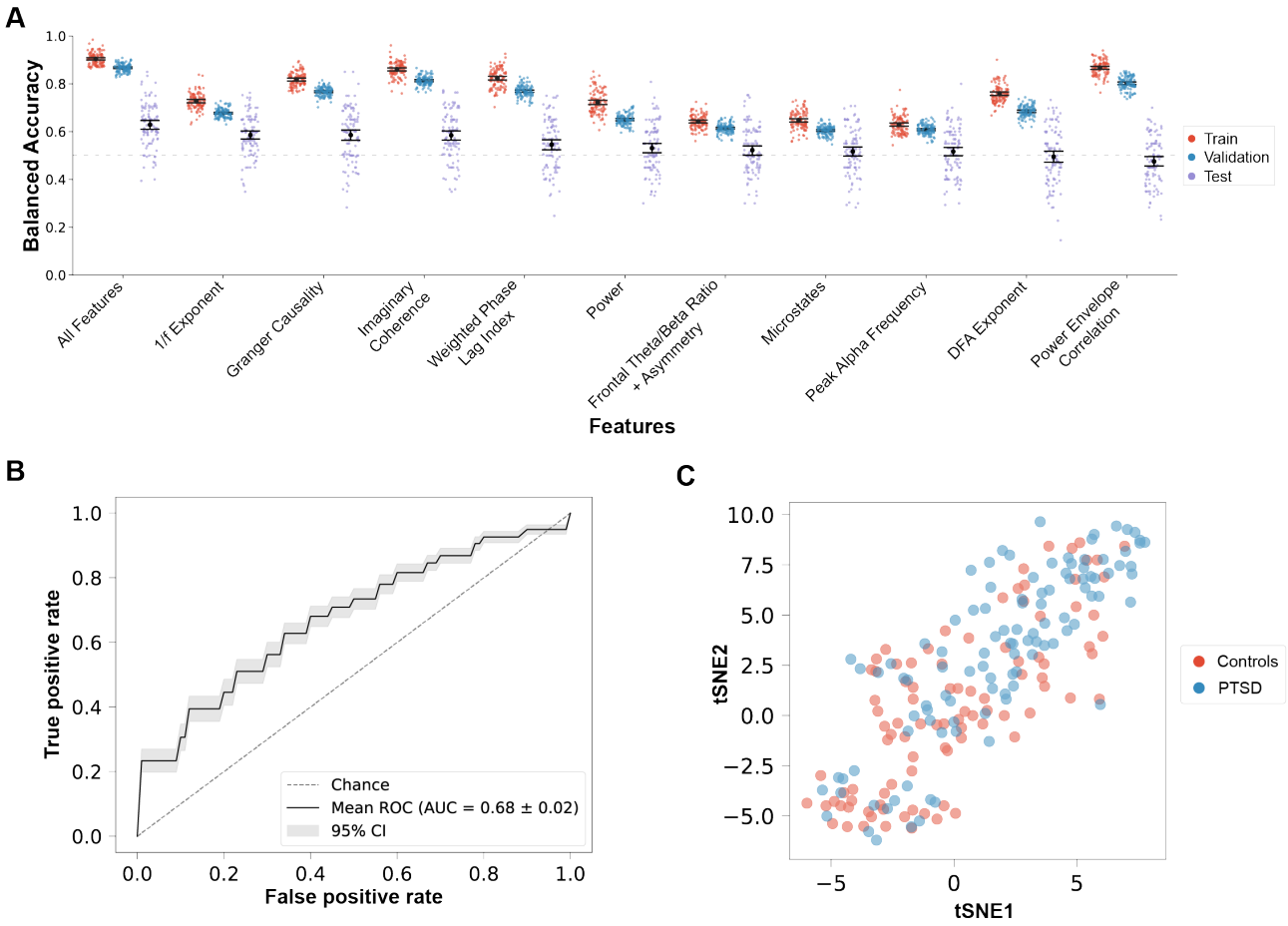

Fig. 3. **Machine learning classification of PTSD.** A) Overview of SVM classification performances. The best model, SVM with all features, obtained a balanced accuracy of 62.9%. For comparison, SVM using each feature type separately are also shown. The feature types are sorted according to average test accuracy on the unseen dataset for the 100 different outer fold splits and mean ± 95% confidence intervals are shown. The dashed line indicates chance level performance. B) Receiver operating characteristic curve for the SVM with all features C) t-SNE of the important features selected by SVM with all features. Only features that were selected in at least 20% of all CV runs were included in the t-SNE. Some separation between the two groups can be observed with the PTSD group being more prominent in the upper right corner and controls in the lower left corner, although there is still a large overlap, reflecting the moderate classification performance we obtained using SVM with all features.

ure 4D), while subtype 1 had a considerably higher difference of 0.1032 (Figure 4E). 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 4F). 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 checked the distribution of the two subtypes across the three different studies to verify the subtypes were not due to study effects. Overall, the two subtypes contained relatively similar percentages of PTSD patients from the different studies ($p = 0.195$, Fisher's exact test), albeit subtype 1 had slightly higher amount of patients from study 3 and correspondingly fewer from study 1 compared to subtype 2 (Supplementary Figure D3).

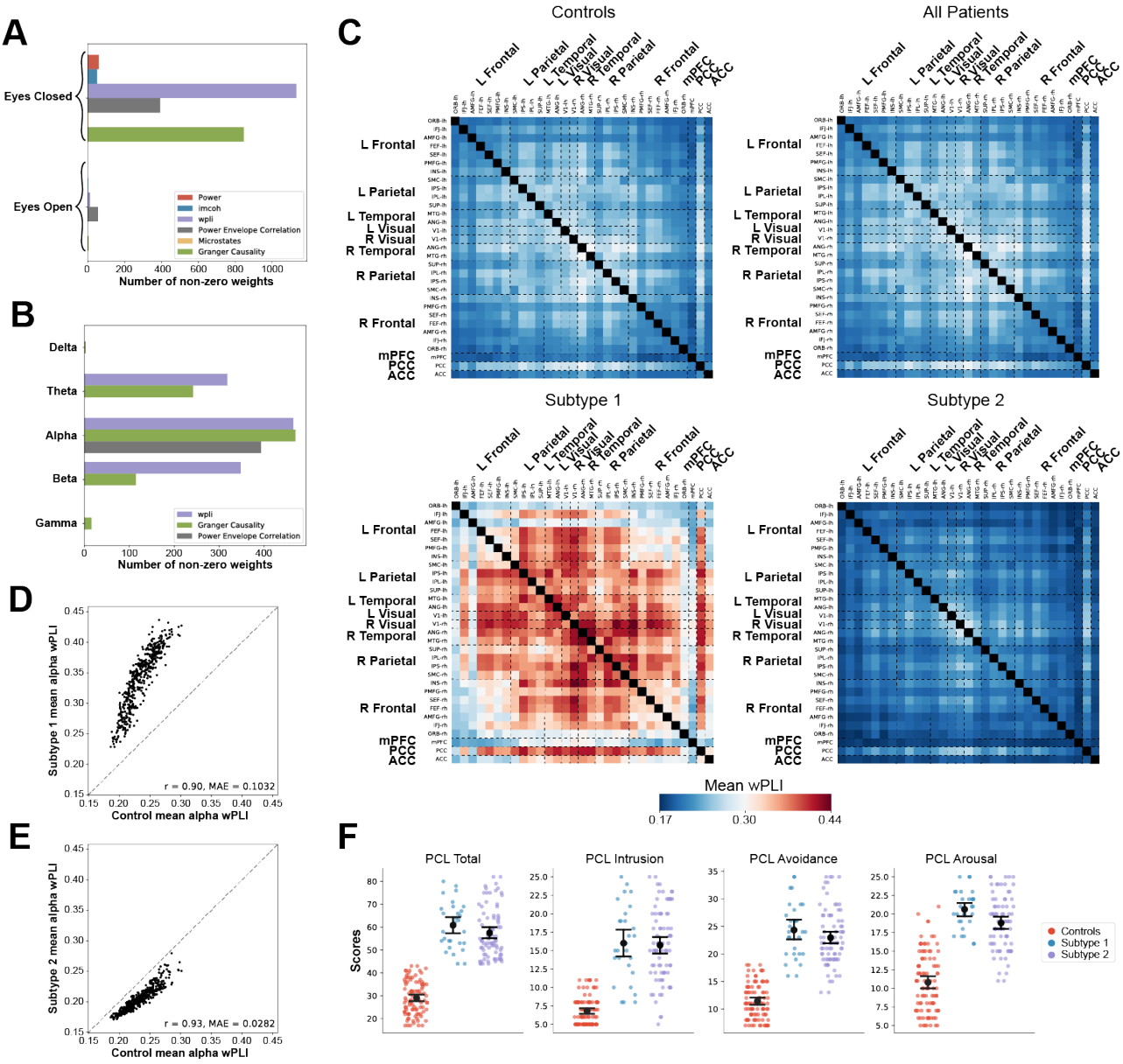

Fig. 4. **Data-driven clustering identified two neurophysiologically distinct subtypes in the PTSD group.** A) Sparse K-means was applied on all features and two clusters of the PTSD group were identified, primarily based on wPLI, GC and PEC. B) Specifically, alpha wPLI, GC and PEC connectivity were 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 $\pm$ 95% confidence intervals are shown. MAE, mean absolute error.

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; Figure 5A and 5B). 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 D4). The results were consistent across all three machine learning classifiers (Supplementary Table 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 3 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 3 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 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 5C).

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 dorsal attention network in the alpha frequency range. In total 1 theta, 8 alpha and 3 beta wPLI connections were consistently selected in more than 20% of all CV runs by the classifiers for prediction of subtype 1 (Figure 6A).

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 8 alpha and 1 beta wPLI connection showed a significant positive correlation after FDR correction. Figure 6B show examples of the 4 most significant correlations between the consistently selected wPLI features and arousal severity.

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

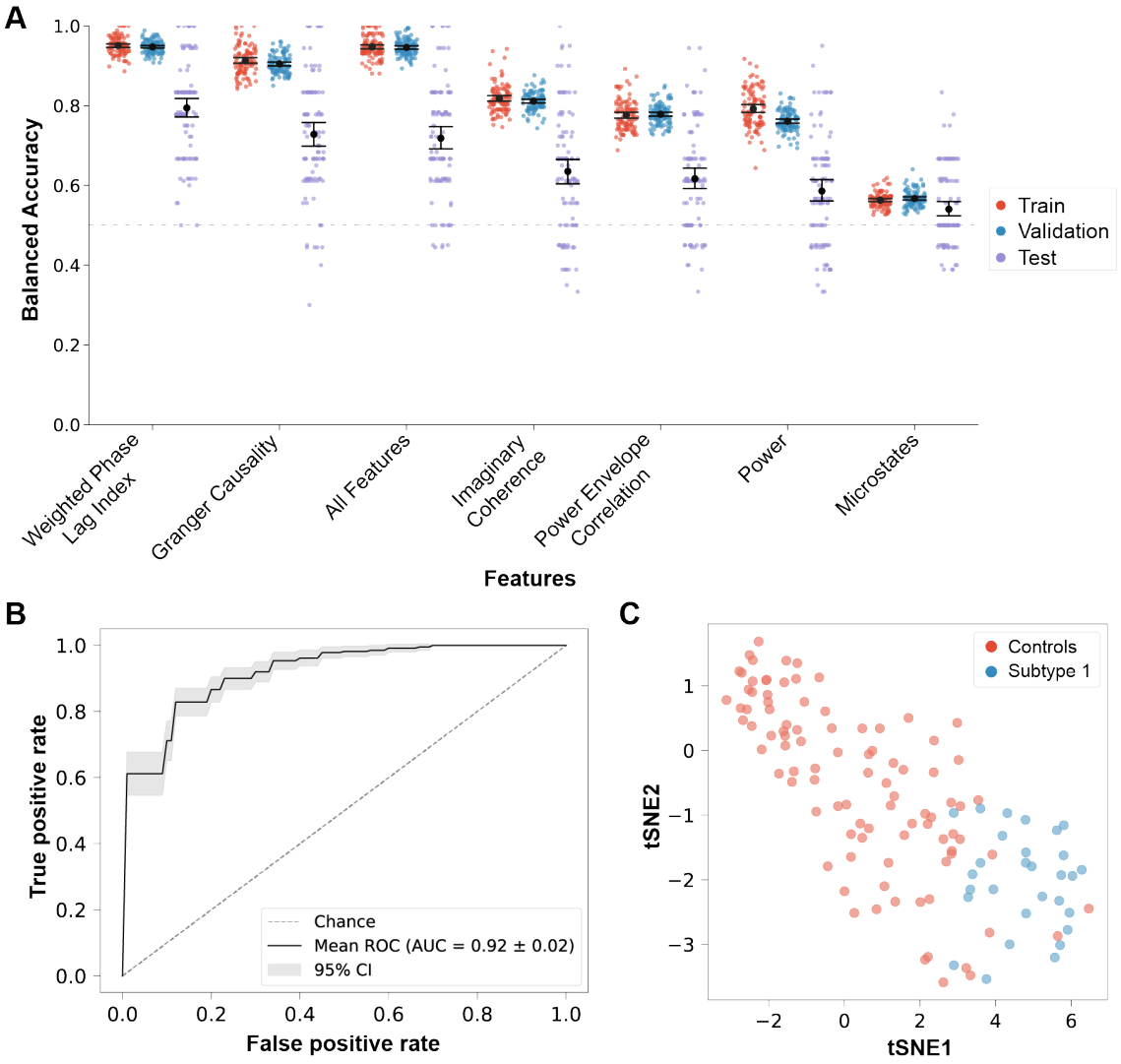

Fig. 5. **Subtyping improved machine learning prediction.** A) Performance of logistic regression using the sparse K-means identified eyes-closed EEG features. The best model using only wPLI features obtained 79.4% balanced test accuracy. The dashed line indicates chance level performance. B) Receiver 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 CV 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.

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 ventral attention network in the beta frequency band, followed by multiple alpha wPLI connections between the dorsal and ventral attention networks. Connections between the primary visual cortex and the two attention networks across all three frequency bands were also important for the classifier (Figure 6).

Both the dorsal and ventral attention networks are key mediators for attention, and distinct attentional subprocesses have been attributed to them. The dorsal attention network is thought to mainly modulate attention towards external ob-

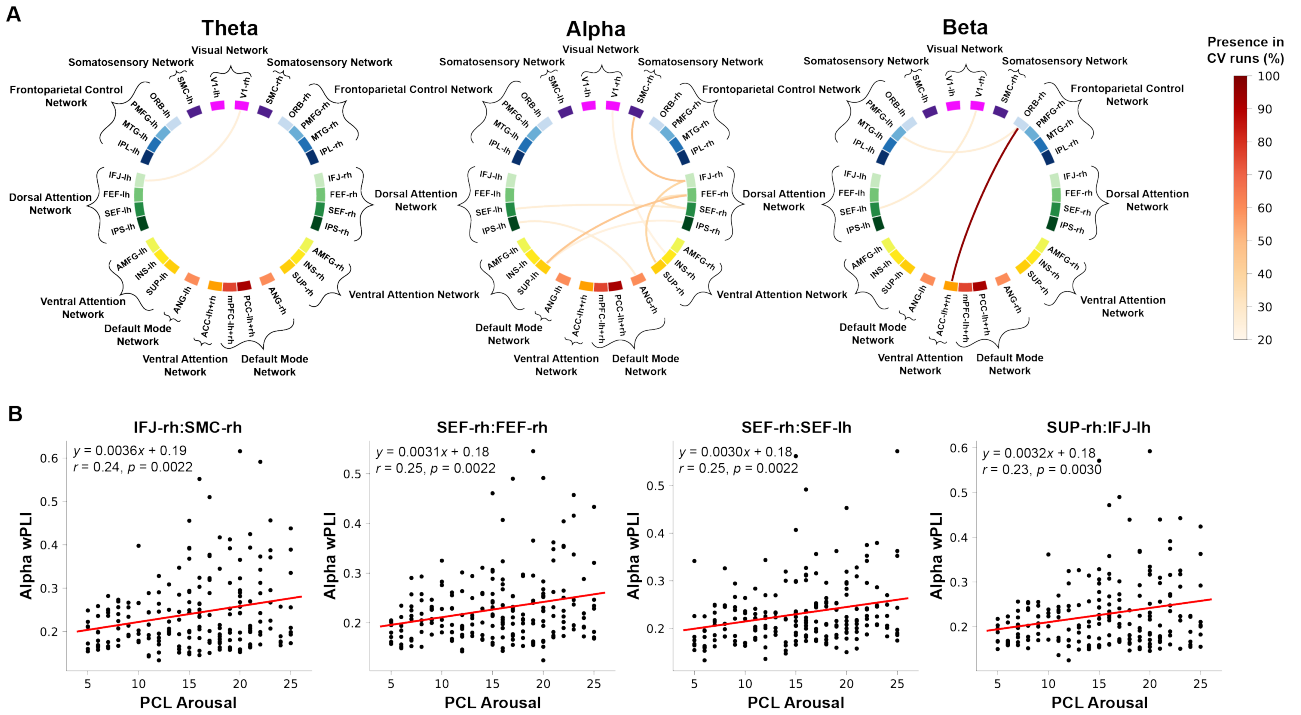

Fig. 6. **The consistently selected features are correlated with arousal symptom severity.** A) Visualization of the 12 most consistently selected wPLI features by the logistic regression trained on classifying subtype 1 from controls. Most of the features were alpha connections between the dorsal and ventral attention networks, with some connections involving the frontoparietal control network in the beta band. One connection between the visual cortex and attention networks were also observed across all three frequency bands. B) Example of the four most significantly correlated wPLI connections with arousal severity. All 8 alpha wPLI connections were significantly positively correlated with arousal severity.

jects, whereas the ventral attention network has been linked to internally directed cognition. The ventral attention network 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 dorsal and ventral attention networks 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 have 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 1-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 ventral and dorsal attention networks (Jin et al. 2021). Another study employed magnetoencephalography (MEG) and found wPLI hyperconnectivity in the high gamma band to be associated with PTSD (Dunkley et al. 2014). An fMRI study have also found an association between aberrant functional connectivity within the ventral attention network 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 dorsal and ventral attention network, were positively correlated with arousal severity (Figure 6B). Patients with PTSD often have impaired attentional control and display hypervigilance towards their sur-

Table 1. Performance and confidence intervals of selected machine learning models

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 10 repetitions of 10-by-10 cross-validation runs. BACC, balanced accuracy. CI, confidence interval.

roundings, 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 on selective attention. During processing of stimuli, alpha oscillations have been observed to be suppressed in brain areas processing the stimuli, while increased alpha activity have 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 4F). 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 metric reported is highly likely to be inflated due to overfitting (Lemm et al. 2011; Hosseini

et al. 2020a; Poldrack et al. 2020; Rashid and Calhoun 2020). The training accuracies we reported for predicting the whole PTSD group (Figure 3A) reflect the accuracies of each feature type obtained by evaluating on the same data it was trained on. Despite the moderate test accuracies for the connectivity feature types, they all 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. 2019 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 circumvent the problem of training and testing on the same dataset (Hosseini et al. 2020b). 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 3A) or the subtypes (Figure 5A and Supplementary Figure D4A). 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 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 10 repetitions of 10-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 FgMDM 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 10 repetitions of 10 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 (RVM) 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 did try subtyping using only the PEC features to compare with the results from Zhang et al. 2020, but the subtypes we found were similar to the results shown in Figure 4C with hyperconnectivity in parietal, temporal, visual areas, and posterior cingulate cortex in one subtype, while the other was more similar to controls, just with slightly lower connectivity.

While the strengths of the present study include the stringent CV scheme and the broad array of EEG features, it is important to also acknowledge the limitations. One limitation is the selected features. To mitigate the effect of feature choice, we included a broad array of different features, but there are other EEG features, which potentially could serve as biomarkers for PTSD. Besides the specific feature types, the current study is also limited to the parcellation and specific areas we investigated. We adapted the cortical parcella-

tion from Zhang et al. 2020, however the electrical activity in other brain areas or sub-cortical structures might be important for PTSD. Especially the amygdala is thought to be highly important for fear conditioning and has also been linked to all four symptom clusters in PTSD in both rodents and humans (Fenster et al. 2018). An electrical fingerprint for the amygdala activity has also been characterized and employed as neurofeedback target for PTSD treatment (Keynan et al. 2019).

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 the source localization. Nonetheless, the data analyzed in the present study were obtained at 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).

It was only subtype 1 that 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, 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 (Woo et al. (2017); Arbabshirani 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 dorsal and ventral attention networks 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 and in the future we envision an overall similar approach, using data from more samples and multiple modalities, e.g. genetics, blood samples, neuroimaging, social behavior etc, and with longitudinal data, will enable clinicians to administer tailored diagnosis and treatment for each subject.

6 Acknowledgements

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.

Data and code availability

The code used for this study is available from https://lab.compute.dtu.dk/glia/resting-state-eeg-analysis. The data is not publicly available due to privacy issues of clinical data.

References

N. Altman and M. Krzywinski. The curse(s) of dimensionality. *Nature Methods*, 15(6):399–400, 2018. ISSN 15487105. doi: 10.1038/s41592-018-0019-x.

American Psychiatric Association. *The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. American Psychiatric Association, 1994.

American Psychiatric Association. *The Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Association, 2013.

S. B. Andersen, K. I. Karstoft, M. Bertelsen, and T. Madsen. Latent trajectories of trauma symptoms and resilience: The 3-year longitudinal prospective USPER study of Danish veterans deployed in Afghanistan. *Journal of Clinical Psychiatry*, 75(9):1001–1008, 2014. ISSN 01606689. doi: 10.4088/JCP.13m08914.

A. Angelidis, M. Hagenaars, D. van Son, W. van der Does, and P. Putman. Do not look away! Spontaneous frontal EEG theta/beta ratio as a marker for cognitive control over attention to mild and high threat. *Biological Psychology*, 135:8–17, 2018. ISSN 18736246. doi: 10.1016/j.biopsycho.2018.03.002.

M. R. Arbabshirani, S. Plis, J. Sui, and V. D. Calhoun. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*, 145:137–165, 2017. ISSN 10959572. doi: 10.1016/j.neuroimage.2016.02.079.

J. Attias, A. Bleich, and S. Gilat. Classification of veterans with post-traumatic stress disorder using visual brain evoked P3s to traumatic stimuli. *British Journal of Psychiatry*, 168:110–115, 1996. ISSN 00071250. doi: 10.1192/bjp.168.1.110.

B. Babadi and E. N. Brown. A review of multitaper spectral analysis. *IEEE Transactions on Biomedical Engineering*, 61(5):1555–1564, 2014. ISSN 15582531. doi: 10.1109/TBME.2014.2311996.

A. B. Barrett, M. Murphy, M. A. Bruno, Q. Noirhomme, M. Boly, S. Laureys, and A. K. Seth. Granger causality analysis of steady-state electroencephalographic signals during propofol-induced anaesthesia. *PLoS ONE*, 7 (1):1–12, 2012. ISSN 19326203. doi: 10.1371/journal.pone.0029072.

D. Begić, L. Hotujac, and N. Jokić-Begić. Electroencephalographic comparison of veterans with combat-related posttraumatic stress disorder and healthy subjects. *International Journal of Psychophysiology*, 40(2):167–172, 3 2001. ISSN 0167-8760. doi: 10.1016/S0167-8760(00)00153-7.

A. J. Bell and T. J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, 1995. ISSN 08997667. doi: 10.1162/neco.1995.7.6.1129.

N. Bigdely-Shamlo, T. Mullen, C. Kothe, K. M. Su, and K. A. Robbins. The PREP pipeline: Standardized preprocessing for large-scale EEG analysis. *Frontiers in Neuroinformatics*, 9:1–19, 2015. ISSN 16625196. doi: 10.3389/fninf.2015.00016.

L. Breiman. Random Forests. *Machine Learning*, 45:5–32, 2001.

M. Butt, E. Espinal, R. L. Aupperle, V. Nikulina, and J. L. Stewart. The electrical aftermath: Brain signals of posttraumatic stress disorder filtered through a clinical lens. *Frontiers in Psychiatry*, 10:368, 2019. ISSN 16640640. doi: 10.3389/fpsyt.2019.00368.

K. Clancy, M. Ding, E. Bernat, N. B. Schmidt, and W. Li. Restless 'rest': Intrinsic sensory hyperactivity and disinhibition in post-traumatic stress disorder. *Brain*, 140(7):2041–2050, 7 2017. ISSN 14602156. doi: 10.1093/brain/awx116.

K. R. Dean, R. Hammamieh, S. H. Mellon, D. Abu-Amara, J. D. Flory, G. Guffanti, K. Wang, et al. Multi-omic

biomarker identification and validation for diagnosing warzone-related post-traumatic stress disorder. *Molecular Psychiatry*, 2019. ISSN 14765578. doi: 10.1038/s41380-019-0496-z.

J. DePierro, L. Lepow, A. Feder, and R. Yehuda. Translating Molecular and Neuroendocrine Findings in PTSD and Resilience to Novel Therapies. *Biological Psychiatry*, 86(6):454–463, 9 2019. ISSN 00063223. doi: 10.1016/j.biopsych.2019.07.009.

C. Destrieux, B. Fischl, A. Dale, and E. Halgren. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, 15(53):1–15, 2010. ISSN 0022-5061. doi: 10.1002/jhbs.20162.

M. Ding, Y. Chen, and S. L. Bressler. Granger Causality: Basic Theory and Application to Neuroscience. *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*, pages 437–460, 2006. doi: 10.1002/9783527609970.ch17.

T. Donoghue, M. Haller, E. J. Peterson, P. Varma, P. Sebastian, R. Gao, T. Noto, et al. Parameterizing neural power spectra into periodic and aperiodic components. *Nature Neuroscience*, 23(12):1655–1665, 2020. ISSN 15461726. doi: 10.1038/s41593-020-00744-x.

T. Donoghue, N. Schaworonkow, and B. Voytek. Methodological considerations for studying neural oscillations. *European Journal of Neuroscience*, 2021. ISSN 14609568. doi: 10.1111/ejn.15361.

B. T. Dunkley, S. M. Doesburg, P. A. Sedge, R. J. Grodecki, P. N. Shek, E. W. Pang, and M. J. Taylor. Resting-state hippocampal connectivity correlates with symptom severity in post-traumatic stress disorder. *NeuroImage: Clinical*, 5:377–384, 2014. ISSN 22131582. doi: 10.1016/j.nicl.2014.07.017.

C. L. Ehlers, S. Hurst, E. Phillips, D. A. Gilder, M. Dixon, A. Gross, P. Lau, and R. Yehuda. Electrophysiological responses to affective stimuli in American Indians experiencing trauma with and without PTSD. *Annals of the New York Academy of Sciences*, 1071(1):125–136, 7 2006. ISSN 17496632. doi: 10.1196/annals.1364.011.

A. Etkin, A. Maron-Katz, W. Wu, G. A. Fonzo, J. Huemer, P. E. Vértes, B. Patenaude, et al. Using fMRI connectivity to define a treatment-resistant form of post-traumatic stress disorder. *Science translational medicine*, 11(486): eaal3236, 4 2019. ISSN 1946-6242. doi: 10.1126/scitranslmed.aal3236.

R. J. Fenster, L. A. Lebois, K. J. Ressler, and J. Suh. Brain circuit dysfunction in post-traumatic stress disorder: from mouse to man. *Nature Reviews Neuroscience*, 19(9):535–551, 9 2018. ISSN 14710048. doi: 10.1038/s41583-018-0039-7.

B. Fischl. FreeSurfer. *NeuroImage*, 62(2):774–781, 2012. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.01.021.

J. J. Foxe and A. C. Snyder. The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, 2:154, 2011. ISSN 16641078. doi: 10.3389/fpsyg.2011.00154.

P. Fries. Rhythms for Cognition: Communication through Coherence. *Neuron*, 88(1):220–235, 10 2015. ISSN 10974199. doi: 10.1016/j.neuron.2015.09.034.

P. Fries, T. Womelsdorf, R. Oostenveld, and R. Desimone. The effects of visual stimulation and selective visual attention on rhythmic neuronal synchronization in macaque area V4. *Journal of Neuroscience*, 28(18):4823–4835, 2008. ISSN 02706474. doi: 10.1523/JNEUROSCI.4499-07.2008.

I. R. Galatzer-Levy and R. A. Bryant. 636,120 Ways to Have Posttraumatic Stress Disorder. *Perspectives on Psychological Science*, 8(6):651–662, 11 2013. ISSN 17456916. doi: 10.1177/1745691613504115.

Q. Gong, L. Li, S. Tognin, Q. Wu, W. Pettersson-Yeo, S. Lui, X. Huang, et al. Using structural neuroanatomy to identify trauma survivors with and without post-traumatic stress disorder at the individual level. *Psychological Medicine*, 44(1):195–203, 2014. ISSN 00332917. doi: 10.1017/S0033291713000561.

A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, et al. MEG and EEG data analysis with MNE-Python. *Frontiers in*

Neuroscience, 7:1–13, 2013. ISSN 1662453X. doi: 10.3389/fnins.2013.00267.

I. Guyon, J. Weston, and S. Barnhill. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, 46:389–422, 2002.

M. Hardmeier, F. Hatz, H. Bousleiman, C. Schindler, C. J. Stam, and P. Fuhr. Reproducibility of functional connectivity and graph measures based on the phase lag index (PLI) and weighted phase lag index (wPLI) derived from high resolution EEG. *PLoS ONE*, 9(10), 2014. ISSN 19326203. doi: 10.1371/journal.pone.0108648.

R. Hardstone, S. S. Poil, G. Schiavone, R. Jansen, V. V. Nikulin, H. D. Mansvelder, and K. Linkenkaer-Hansen. Detrended fluctuation analysis: A scale-free view on neuronal oscillations. *Frontiers in Physiology*, pages 1–13, 2012. ISSN 1664042X. doi: 10.3389/fphys.2012.00450.

T. Harmony. The functional significance of delta oscillations in cognitive processing. *Frontiers in Integrative Neuroscience*, 7:1–10, 2013. ISSN 16625145. doi: 10.3389/fnint.2013.00083.

J. F. Hipp, D. J. Hawellek, M. Corbetta, M. Siegel, and A. K. Engel. Large-scale cortical correlation structure of spontaneous oscillatory activity. *Nature Neuroscience*, 15(6): 884–890, 2012. ISSN 10976256. doi: 10.1038/nn.3101.

M. D. Hoskins, J. Bridges, R. Sinnerton, A. Nakamura, J. F. Underwood, A. Slater, M. R. Lee, et al. Pharmacological therapy for post-traumatic stress disorder: a systematic review and meta-analysis of monotherapy, augmentation and head-to-head approaches. *European Journal of Psychotraumatology*, 12(1), 2021. ISSN 20008066. doi: 10.1080/20008198.2020.1802920.

M. Hosseini, M. Powell, J. Collins, C. Callahan-Flintoft, W. Jones, H. Bowman, and B. Wyble. I tried a bunch of things: The dangers of unexpected overfitting in classification of brain data. *Neuroscience and Biobehavioral Reviews*, 119:456–467, 2020a. ISSN 18737528. doi: 10.1016/j.neubiorev.2020.09.036.

M.-P. Hosseini, A. Hosseini, and K. Ahi. A Review on Machine Learning for EEG Signal Processing in Bioengineering. *IEEE Reviews in Biomedical Engineering*, 2020b. ISSN 1937-3333. doi: 10.1109/rbme.2020.2969915.

C. Imperatori, B. Farina, M. I. Quintiliani, A. Onofri, P. Castelli Gattinara, M. Lepore, V. Gnoni, et al. Aberrant EEG functional connectivity and EEG power spectra in resting state post-traumatic stress disorder: A sLORETA study. *Biological Psychology*, 102(1):10–17, 10 2014. ISSN 18736246. doi: 10.1016/j.biopsycho.2014.07.011.

T. Insel, B. Cuthbert, M. Garvey, R. Heinssen, D. Pine, K. Quinn, C. Sanislow, and P. Wang. Research Domain Criteria (RDoC): Toward a New Classification Framework for Research on Mental Disorders. *American Journal of Psychiatry Online*, 167:748–751, 2010. ISSN 0002-953X.

M. Jas, D. A. Engemann, Y. Bekhti, F. Raimondo, and A. Gramfort. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage*, 159 (December 2016):417–429, 2017. ISSN 10959572. doi: 10.1016/j.neuroimage.2017.06.030.

S. Jin, C. Shin, C. Han, Y.-K. Kim, J. Lee, S. W. Jeon, S.-H. Lee, and Y.-H. Ko. Changes in Brain Electrical Activity According to Post-traumatic Stress Symptoms in Survivors of the Sewol Ferry Disaster: A 1-year Longitudinal Study. *Clinical Psychopharmacology and Neuroscience*, 19(3):537–544, 2021. ISSN 1738-1088. doi: 10.9758/cpn.2021.19.3.537.

N. Jokić-Begić and D. Begić. Quantitative electroencephalogram (qEEG) in combat veterans with post-traumatic stress disorder (PTSD). *Nordic Journal of Psychiatry*, 57(5):351–355, 1 2003. ISSN 08039488. doi: 10.1080/08039480310002688.

K. I. Karstoft, S. B. Andersen, M. Bertelsen, and T. Madsen. Diagnostic accuracy of the posttraumatic stress disorder checklist: Civilian version in a representative military sample. *Psychological Assessment*, 26(1):321–325, 2014. ISSN 10403590. doi: 10.1037/a0034889.

A. H. Kemp, K. Griffiths, K. L. Felmingham, S. A. Shankman, W. Drinkenburg, M. Arns, C. R. Clark, and R. A. Bryant. Disorder specificity despite comorbidity: Resting EEG alpha asymmetry in major depressive disorder and post-traumatic stress disorder. *Biological Psy-*

chology, 85(2):350–354, 10 2010. ISSN 03010511. doi: 10.1016/j.biopsycho.2010.08.001.

J. N. Keynan, A. Cohen, G. Jackont, N. Green, N. Goldway, A. Davidov, Y. Meir-Hasson, et al. Electrical fingerprint of the amygdala guides neurofeedback training for stress resilience. *Nature Human Behaviour*, 3(1):63–73, 2019. ISSN 23973374. doi: 10.1038/s41562-018-0484-3.

Y.-w. Kim, S. Kim, M. Shim, M. Jin, and H. Jeon. Riemannian classifier enhances the accuracy of machine-learning-based diagnosis of PTSD using resting EEG. *Progress in Neuropsychopharmacology & Biological Psychiatry*, 102:109960, 2020. ISSN 0278-5846. doi: 10.1016/j.pnpbp.2020.109960.

K. Kiran. mRMR feature selection, 2017. url: *https://github.com/stochasticresearch/featureselect*.

W. Klimesch, P. Sauseng, and S. Hanslmayr. EEG alpha oscillations: The inhibition-timing hypothesis, 1 2007. ISSN 01650173.

J. P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela. Measuring phase synchrony in brain signals. *Human Brain Mapping*, 8(4):194–208, 1999. ISSN 10659471. doi: 10.1002/(SICI)1097-0193(1999)8:4¡194::AID-HBM4¿3.0.CO;2-C.

D. Lehmann, H. Ozaki, and I. Pal. EEG alpha map series: brain micro-states by space-oriented adaptive segmentation. *Electroencephalography and Clinical Neurophysiology*, 67(3):271–288, 1987. ISSN 00134694. doi: 10.1016/0013-4694(87)90025-3.

S. Lemm, B. Blankertz, T. Dickhaus, and K. R. Müller. Introduction to machine learning for brain imaging. *NeuroImage*, 56(2):387–399, 2011. ISSN 10538119. doi: 10.1016/j.neuroimage.2010.11.004.

K. Linkenkaer-Hansen, V. V. Nikouline, J. M. Palva, and R. J. Ilmoniemi. Long-range temporal correlations and scaling behavior in human brain oscillations. *Journal of Neuroscience*, 21(4):1370–1377, 2001. ISSN 02706474. doi: 10.1523/jneurosci.21-04-01370.2001.

F. Liu, B. Xie, Y. Wang, W. Guo, J. P. Fouche, Z. Long, W. Wang, et al. Characterization of Post-traumatic Stress Disorder Using Resting-State fMRI with a Multi-level Parametric Classification Approach. *Brain Topography*, 28 (2):221–237, 2014. ISSN 15736792. doi: 10.1007/s10548-014-0386-2.

S. Makeig, A. J. Bell, T.-P. Jung, and T. J. Sejnowski. Independent Component Analysis of Electroencephalographic Data. *Advances in neural information processing systems*, 8(8):145–151, 1996. ISSN 10495258.

V. Menon. Large-scale brain networks and psychopathology: A unifying triple network model. *Trends in Cognitive Sciences*, 15(10):483–506, 2011. ISSN 13646613. doi: 10.1016/j.tics.2011.08.003.

C. M. Michel and D. Brunet. EEG source imaging: A practical review of the analysis steps. *Frontiers in Neurology*, 10, 2019. ISSN 16642295. doi: 10.3389/fneur.2019.00325.

S. Y. Moon, Y. B. Choi, H. K. Jung, Y. I. Lee, and S. H. Choi. Increased frontal gamma and posterior delta powers as potential neurophysiological correlates differentiating posttraumatic stress disorder from anxiety disorders. *Psychiatry Investigation*, 15(11):1087–1093, 11 2018. ISSN 19763026. doi: 10.30773/pi.2018.09.30.

Y. Neria. Functional neuroimaging in PTSD: From discovery of underlying mechanisms to addressing diagnostic heterogeneity. *American Journal of Psychiatry*, 178(2):128–135, 2021. ISSN 15357228. doi: 10.1176/appi.ajp.2020.20121727.

NIPY. Nitime - Time-series analysis of neuroscience data, 2019. url: *http://nipy.org/nitime/index.html*.

G. Nolte, O. Bai, L. Wheaton, Z. Mari, S. Vorbach, and M. Hallett. Identifying true brain interaction from EEG data using the imaginary part of coherency. *Clinical Neurophysiology*, 115(10):2292–2307, 10 2004. ISSN 13882457. doi: 10.1016/j.clinph.2004.04.029.

P. L. Nunez, R. Srinivasan, A. F. Westdorp, R. S. Wijesinghe, D. M. Tucker, R. B. Silberstein, and P. J. Cadusch. EEG coherency I: Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. *Electroencephalography and Clinical Neurophysiology*, 103(5):499–515, 1997. ISSN 00134694. doi: 10.1016/S0013-4694(97)00066-7.

S. Palva and J. M. Palva. Discovering oscillatory interaction networks with M/EEG: Challenges and breakthroughs. *Trends in Cognitive Sciences*, 16(4):219–230, 2012. ISSN 1879307X. doi: 10.1016/j.tics.2012.02.004.

R. D. Pascual-Marqui, C. M. Michel, and D. Lehmann. Segmentation of Brain Electrical Activity into Microstates; Model Estimation and Validation. *IEEE Transactions on Biomedical Engineering*, 42(7):658–665, 1995. ISSN 15582531. doi: 10.1109/10.391164.

F. Pedrogosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, et al. Scikit-learn: Machine Learning in Python Fabian. *Journal of Machine Learning Research*, 12:2825–2830, 2011. ISSN 15529924. doi: 10.1289/EHP4713.

H. Peng, F. Long, and C. Ding. Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8):1226–1238, 2005. ISSN 01628828. doi: 10.1109/TPAMI.2005.159.

F. Perrin, J. Pernier, and O. Bertrand. Spherical splines for scalp potential and current density mapping. *Electroencephalograp and clinical Neurophysiology*, pages 184–187, 1989.

R. A. Poldrack, G. Huckins, and G. Varoquaux. Establishment of Best Practices for Evidence for Prediction: A Review, 2020. ISSN 2168622X.

S. Rabe, A. Beauducel, T. Zöllner, A. Maercker, and A. Karl. Regional brain electrical activity in posttraumatic stress disorder after motor vehicle accident. *Journal of Abnormal Psychology*, 115(4):687–698, 2006. ISSN 0021843X. doi: 10.1037/0021-843X.115.4.687.

S. Raschka. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*, 3(24):638, 2018. ISSN 2475-9066. doi: 10.21105/joss.00638.

B. Rashid and V. Calhoun. Towards a brain-based predictome of mental illness. *Human Brain Mapping*, 41(12):3468–3535, 2020. ISSN 10970193. doi: 10.1002/hbm.25013.

T. A. Rihs, C. M. Michel, and G. Thut. Mechanisms of selective inhibition in visual spatial attention are indexed by α-band EEG synchronization. *European Journal of Neuroscience*, 25(2):603–610, 1 2007. ISSN 1460-9568. doi: 10.1111/J.1460-9568.2007.05278.X.

T. Ros, P. Frewen, J. Théberge, A. Michela, R. Kluetsch, A. Mueller, G. Candrian, et al. Neurofeedback Tunes Scale-Free Dynamics in Spontaneous Brain Activity. *Cerebral Cortex*, 27(10):4911–4922, 9 2017. ISSN 14602199. doi: 10.1093/cercor/bhw285.

C. Schölzel. Nonlinear measures for dynamical systems (Nolds), 2019. url: *https://doi.org/10.5281/zenodo.3814723*.

S. A. Shankman, S. M. Silverstein, L. M. Williams, P. J. Hopkinson, A. H. Kemp, K. L. Felmingham, R. A. Bryant, et al. Resting electroencephalogram asymmetry and posttraumatic stress disorder. *Journal of Traumatic Stress*, 21(2): 190–198, 4 2008. ISSN 08949867. doi: 10.1002/jts.20319.

E. Shvil, H. L. Rusch, G. M. Sullivan, and Y. Neria. Neural, psychophysiological, and behavioral markers of fear processing in PTSD: A review of the literature. *Current Psychiatry Reports*, 15(5):358, 5 2013. ISSN 15233812. doi: 10.1007/s11920-013-0358-3.

C. J. Stam, G. Nolte, and A. Daffertshofer. Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources. *Human Brain Mapping*, 28(11):1178–1193, 2007. ISSN 10659471. doi: 10.1002/hbm.20346.

M. M. Steenkamp, B. T. Litz, C. W. Hoge, and C. R. Marmar. Psychotherapy for military-related PTSD: A review of randomized clinical trials. *JAMA - Journal of the American Medical Association*, 314(5):489–500, 8 2015. ISSN 15383598. doi: 10.1001/jama.2015.8370.

S. K. Sutton and R. J. Davidson. PREFRONTAL BRAIN ASYMMETRY : Inhibition Systems. *Psychological science*, 8(3):204–210, 1997. ISSN 0956-7976.

B. A. Terpou, S. B. Shaw, J. Théberge, F. Victor, C. M. Michel, M. C. McKinnon, R. A. Lanius, and T. Ros. EEG microstates in PTSD: Using machine

learning to identify neuromarkers. *SSRN*, 2022. doi: http://dx.doi.org/10.2139/ssrn.4061516.

R. Tibshirani, G. Walther, and T. Hastie. Estimating the number of clusters in a data set via the gap statistic. *J. R. Statist. Soc. B*, 63:411–423, 2001. ISSN 01676423. doi: 10.1111/1467-9868.00293.

D. Todder, J. Levine, A. Abujumah, M. Mater, H. Cohen, and Z. Kaplan. The quantitative electroencephalogram and the low-resolution electrical tomographic analysis in posttraumatic stress disorder. *Clinical EEG and Neuroscience*, 43(1):48–53, 1 2012. ISSN 15500594. doi: 10.1177/1550059411428716.

R. T. Toll, W. Wu, S. Naparstek, Y. Zhang, M. Narayan, B. Patenaude, C. De Los Angeles, et al. An Electroencephalography Connectomic Profile of Posttraumatic Stress Disorder. *The American journal of psychiatry*, 177(3):233–243, 2020. ISSN 15357228. doi: 10.1176/appi.ajp.2019.18080911.

Tsumeruso. Pysparcl, 2019. url: *https://github.com/tsurumeso/pysparcl*.

L. Q. Uddin, B. T. Yeo, and R. N. Spreng. Towards a Universal Taxonomy of Macro-scale Functional Human Brain Networks. *Brain Topography*, 32(6):926–942, 2019. ISSN 15736792. doi: 10.1007/s10548-019-00744-6.

L. van der Maaten and G. Hinton. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008. ISSN 15729338. doi: 10.1007/s10479-011-0841-3.

M. D. Veltmeyer, A. C. Mcfarlane, and R. A. Bryant. Integrative assessment of brain function in PTSD: Brain stability and working memory. *Journal of Integrative Neuroscience*, 5(1):123–138, 2006.

M. Vinck, R. Oostenveld, M. Van Wingerden, F. Battaglia, and C. M. Pennartz. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage*, 55(4):1548–1565, 2011. ISSN 10538119. doi: 10.1016/j.neuroimage.2011.01.055.

F. von Wegner and H. Laufs. Information-theoretical analysis of EEG microstate sequences in python. *Frontiers in Neuroinformatics*, 12(June):1–10, 2018. ISSN 16625196. doi: 10.3389/fninf.2018.00030.

S. Vossel, J. J. Geng, and G. R. Fink. Dorsal and ventral attention systems: Distinct neural circuits but collaborative roles. *Neuroscientist*, 20(2):150–159, 7 2014. ISSN 10894098. doi: 10.1177/1073858413494269.

H. Wahbeh and B. S. Oken. Peak high-frequency HRV and peak alpha frequency higher in PTSD. *Applied Psychophysiology Biofeedback*, 38(1):57–69, 3 2013. ISSN 10900586. doi: 10.1007/s10484-012-9208-z.

F. W. Weathers, B. T. Litz, D. S. Herman, J. a. Huska, and T. M. Keane. The PTSD Checklist (PCL): Reliability, Validity, and Diagnostic Utility. *Paper presented at the Annual Convention of the International Society for Traumatic Stress Studies, San Antonio, TX*, 2:90–92, 1993.

D. M. Witten and R. Tibshirani. A framework for feature selection. *American Statistician*, 105(490):713–726, 2010. ISSN 0162-1459. doi: 10.1198/jasa.2010.tm09415.A.

C. W. Woo, L. J. Chang, M. A. Lindquist, and T. D. Wager. Building better biomarkers: Brain models in translational neuroimaging. *Nature Neuroscience*, 20(3):365–377, 2017. ISSN 15461726. doi: 10.1038/nn.4478.

H. Yuan, R. Phillips, C. K. Wong, V. Zotev, M. Misaki, B. Wurfel, F. Krueger, et al. Tracking resting state connectivity dynamics in veterans with PTSD. *NeuroImage: Clinical*, 19:260–270, 1 2018. ISSN 22131582. doi: 10.1016/j.nicl.2018.04.014.

A. Zandvakili, N. S. Philip, S. R. Jones, A. R. Tyrka, B. D. Greenberg, and L. L. Carpenter. Use of machine learning in predicting clinical response to transcranial magnetic stimulation in comorbid posttraumatic stress disorder and major depression: A resting state electroencephalography study. *Journal of Affective Disorders*, 252:47–54, 6 2019. ISSN 15732517. doi: 10.1016/j.jad.2019.03.077.

Y. Zhang, W. Wu, R. T. Toll, S. Naparstek, A. Maron-Katz, M. Watts, J. Gordon, et al. Identification of psychiatric disorder subtypes from functional connectivity

patterns in resting-state electroencephalography. *Nature Biomedical Engineering*, 2020. ISSN 2157846X. doi: 10.1038/s41551-020-00614-8.

Appendices

Appendix A: Overview of PTSD resting-state EEG power spectral density findings

	Delta δ (1 - 4Hz)	Theta θ (4 - 8Hz)	Alpha α (8 - 13Hz)	Beta β (13 - 30Hz)	Gamma γ (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	3, 6

1. Begić et al 2001
2. Ehlers et al 2006
3. Imperatori et al 2014
4. Kemp et al 2010
5. Todder et al 2012
6. Wahbeh & Oken 2013
7. Moon et al 2018
8. Clancy et al 2017
9. Veltmeyer et al 2006
10. Rabe et al 2006
11. Shankman et al 2008
12. Jokić-Begić & Begić 2003

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.

Appendix B: Supplementary methods

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

Spectral analysis

EEG power was computed using multitaper spectral estimation with 7 discrete prolate spheroidal sequences (DPSS) 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):

$$\text{Asymmetry} = \ln(Power_{rh}) - \ln(Power_{lh}) \tag{B1}$$

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 exponent 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 6Hz as the start frequency, until the R^2 was above the thresholds. The end range of the fit was set to 40Hz. 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.

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 data (Hardstone et al. 2012). Nolds 0.5.2 (Schölzel 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.

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.

Source functional connectivity

Interpolation of channels might lead to spurious connectivity, especially with neighboring 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: 1) it makes it possible to compare results across the studies, 2) by giving estimates on cortical brain areas instead of the electrode locations, 3) while also alleviating volume conduction (Palva and Palva 2012, Donoghue et al. 2021). We adapted the source localization methodology from Zhang et al. 2020, 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 20484 vertices, followed by extraction of time series from 31 cortical regions of interest (ROI). 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. 2020. The areas were: Visual Area 1 (V1); Somatosensory Cortex (SMC); Orbital Gyrus (ORB); Posterior Middle Frontal Gyrus (PMFG); Middle Temporal Gyrus (MTG); Inferior Parietal Lobule (IPL); Inferior Frontal Junction (IFJ); Frontal Eye Fields (FEF); Supplementary Eye Fields (SEF); Intraparietal Sulcus (IPS); Anterior Middle Frontal Gyrus (AMFG); Insula (INS); Supramarginal Gyrus (SUP); Angular Gyrus (ANG); Posterior Cingulate Cortex (PCC); Medial Prefrontal Cortex (mPFC); Anterior Cingulate Cortex (ACC). The 31 ROIs were originally derived from ICA 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.

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 with 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).

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

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.

Weighted Phase Lag Index

The cross-spectral density contains information about both the amplitudes and the phase difference between the two signals of interest. When looking at 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 lead/lag 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).

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

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 weighted phase lag index 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).

$$\begin{aligned} \text{wPLI} &= \frac{\frac{1}{T}\left|\sum_{t=1}^{T} |\sin(\Delta\phi_t)| \cdot sign[\sin(\Delta\phi_t)]\right|}{\frac{1}{T}\sum_{t=1}^{T} |\sin(\Delta\phi_t)|} \\ \Leftrightarrow \text{wPLI} &= \frac{\left|\sum_{t=1}^{T} |\sin(\Delta\phi_t)| \cdot sign[\sin(\Delta\phi_t)]\right|}{\sum_{t=1}^{T} |\sin(\Delta\phi_t)|} \end{aligned} \tag{B4}$$

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

Power Envelope Correlations

PEC were estimated following Toll et al. 2020. Briefly, the time series for each signal pair were bandpass 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.

Granger Causality

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 (Eq. B5) with a restricted model (Eq. B6):

$$\begin{bmatrix} X(t) \\ Y(t) \end{bmatrix} = \sum_{k=1}^{P} A_k \cdot \begin{bmatrix} X(t - kt_0) \\ Y(t - kt_0) \end{bmatrix} + \begin{bmatrix} E_x(t) \\ E_y(t) \end{bmatrix} \tag{B5}$$

$$Y(t) = \sum_{k=1}^{P} B_k \cdot Y(t - kt_0) + \hat{E}_y(t) \tag{B6}$$

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 $\hat{E}_y$ 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):

$$GC_{X \to Y} = \ln\left(\frac{var(\hat{E}_y(t))}{var(E_y(t))}\right) \tag{B7}$$

The provided equations are for estimation of GC in the time domain, however by performing Fourier transformation it is possible to estimate GC for each of the canonical frequency bands (Ding et al. 2006). GC estimation using autoregressive models assume 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 nitime 0.8.1 (NIPY 2019) python library was used for implementation of GC.

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]

(Continued on next page)

Table C1 – continued from previous page

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

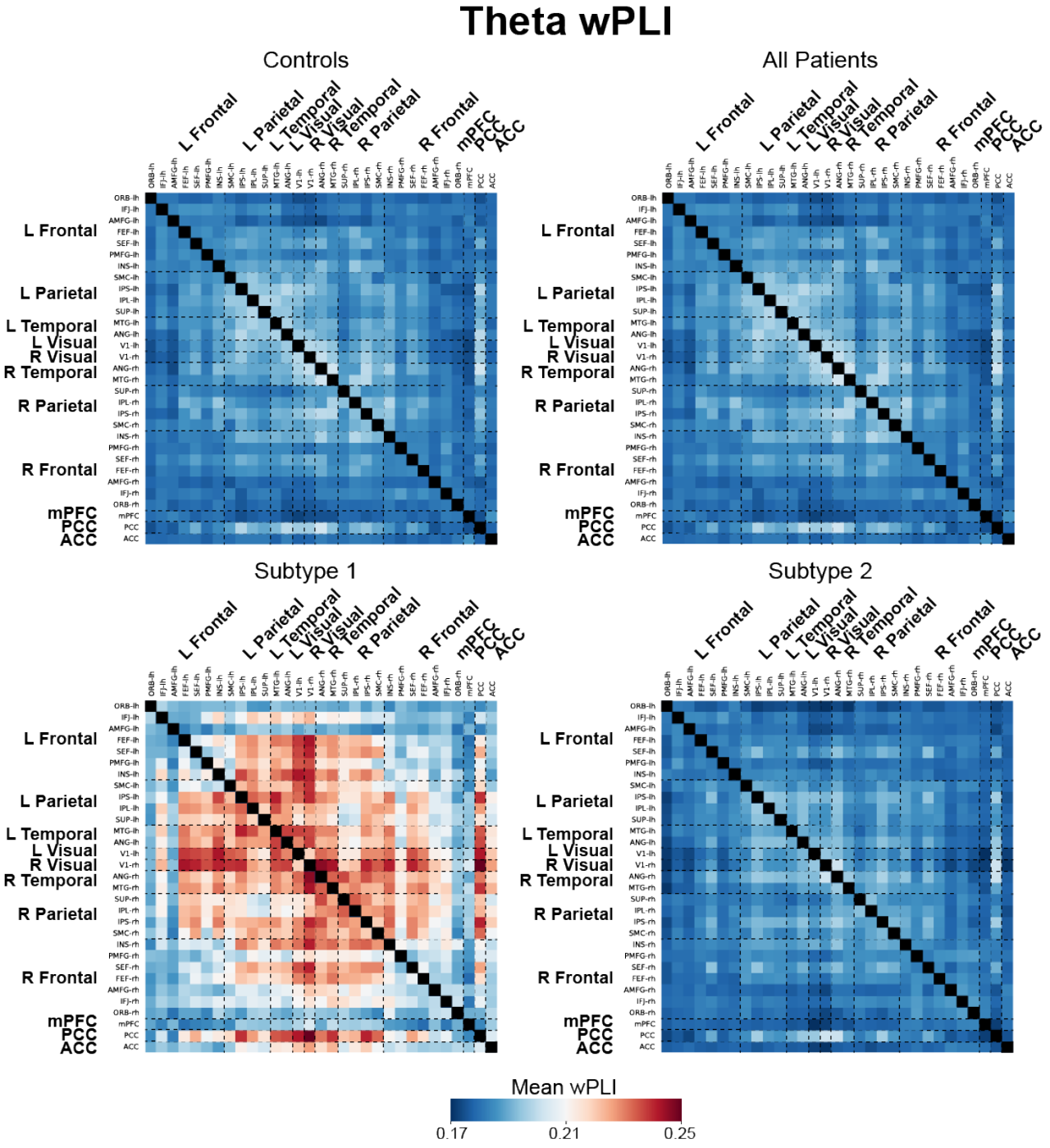

Fig. D1. **Mean theta wPLI connectivity.** No differences were observed between the PTSD and control group, but after clustering we observed higher wPLI in subtype 1, while subtype 2 was relatively similar to controls.

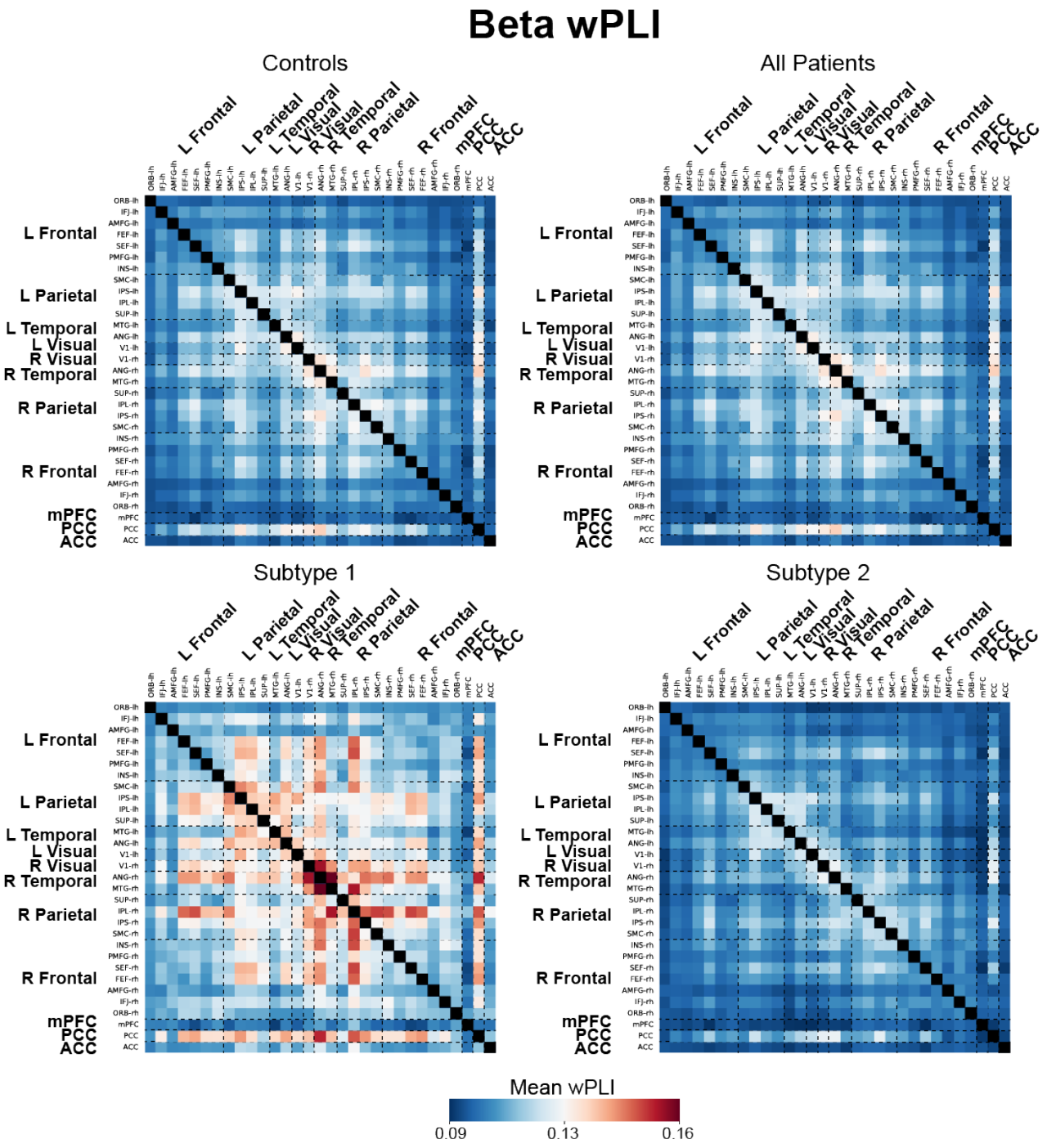

Fig. D2. **Mean beta wPLI connectivity.** No differences were observed between the PTSD and control group, but after clustering we observed higher wPLI in subtype 1, while subtype 2 was relatively similar to controls.

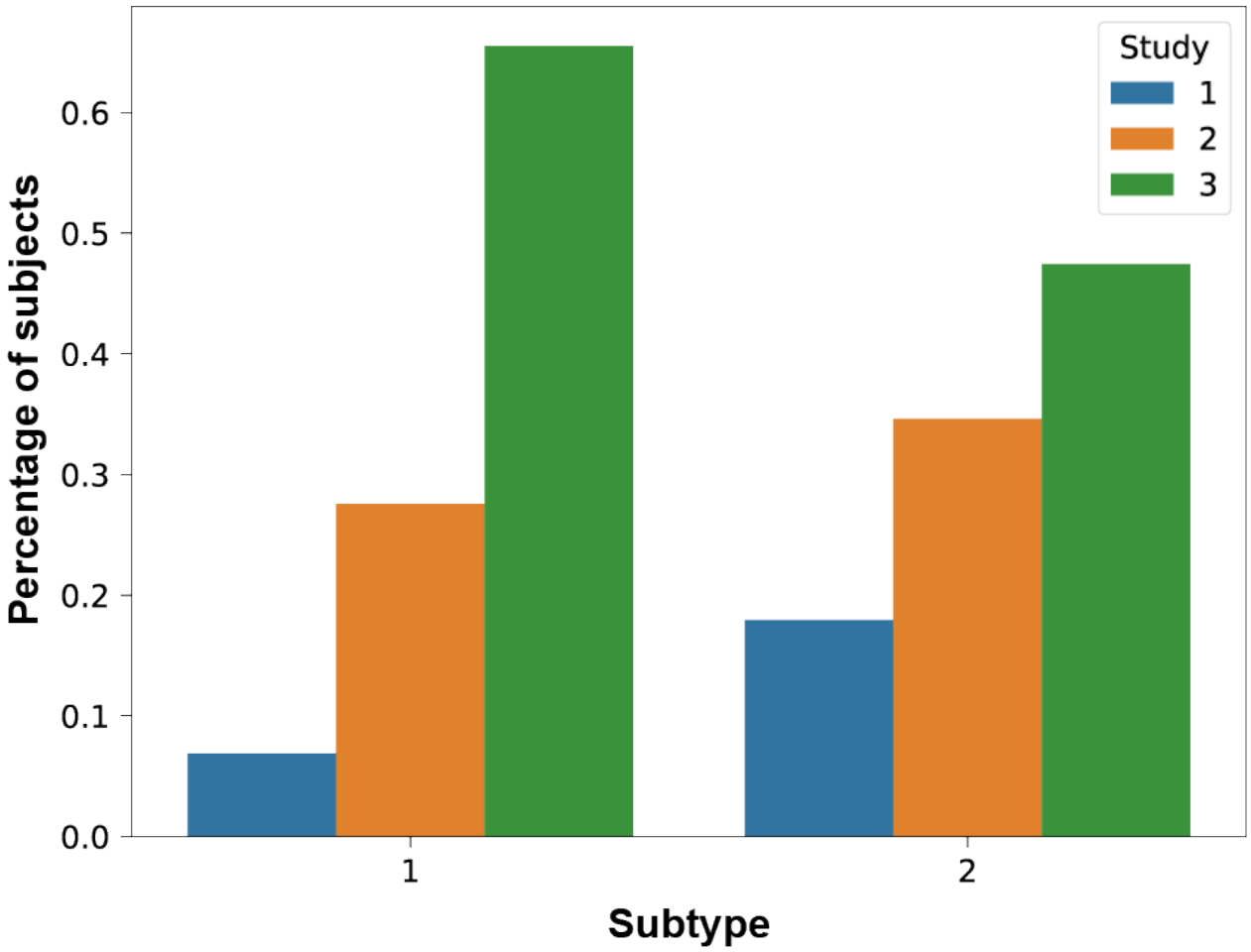

Fig. D3. **Distribution of the subjects from the two PTSD subtypes across the three studies.** Overall, the two subtypes contained relatively similar percentages of patients from the different studies, albeit subtype 1 had slightly higher amount of patients from study 3 and correspondingly fewer from study 1 compared to subtype 2.

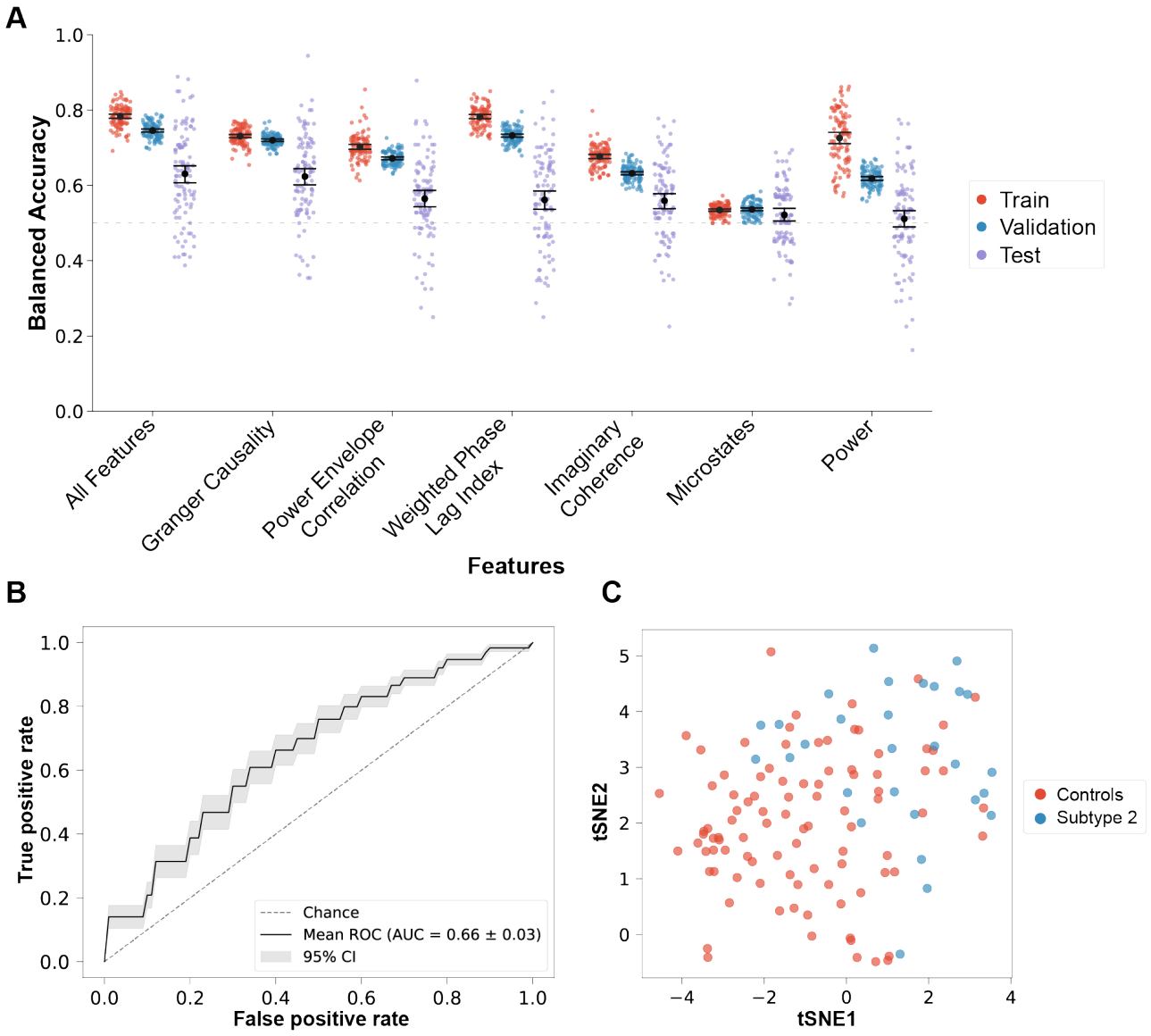

Fig. D4. **Prediction of subtype 2 was not better than prediction of the whole PTSD group.** A) Performance of SVM using the sparse K-means identified eyes-closed EEG features. The best model using all the sparse K-means identified features obtained 63.1% balanced test accuracy. The dashed line indicates 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 CV 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.

APPENDIX B

ASD Paper

Title: Adults with autism spectrum disorder show typical resting-state EEG activity

Authors: Qianliang Li*, Ricarda F. Weiland*, Ivana Konvalinka, Huibert D. Mansvelder, Tobias S. Andersen, Dirk J. A. Smit and Sander Begeer and Klaus Linkenkaer-Hansen

*Shared first-author

Journal: Scientific Reports

Status: Under review

ADULTS WITH AUTISM SPECTRUM DISORDER SHOW TYPICAL RESTING-STATE EEG ACTIVITY

Qianliang Li[1*†], Ricarda F. Weiland[2*], Ivana Konvalinka[1], Huibert D. Mansvelder[3], Tobias S. Andersen[1], Dirk J. A. Smit[4], Sander Begeer[2], and Klaus Linkenkaer-Hansen[3]

[1]Section for Cognitive Systems, DTU Compute, Technical University of Denmark, Denmark,

[2]Faculty of Behavioural and Movement Sciences, Vrije Universiteit Amsterdam, The Netherlands,

[3]Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, The Netherlands,

[4]Department of Psychiatry, Amsterdam UMC, University of Amsterdam, The Netherlands,

[*]These authors contributed equally to this work

[†]Corresponding author. Email: glia@dtu.dk

ABSTRACT

There is broad interest in discovering quantifiable physiological biomarkers for psychiatric disorders to aid diagnostic assessment. However, finding biomarkers for autism spectrum disorder (ASD) has proven particularly difficult, partly due to high heterogeneity. Here, we recorded five minutes eyes-closed rest EEG from 186 adults (51% with ASD and 49% without ASD) and investigated the potential of EEG biomarkers to classify ASD using four conventional machine learning models with two-layer cross-validation. Comprehensive characterization of spectral, temporal and spatial dimensions of source-modelled EEG resulted in 3443 biomarkers per recording. We found no significant group-mean and group-variance differences for any of the EEG features. Interestingly, we obtained validation accuracies above 80%; however, the best machine learning model merely distinguished ASD from the non-autistic comparison group with a mean balanced test accuracy of 56% on the entirely unseen test set. The large drop in model performance between validation and testing, stress the importance of proper model evaluation, and further highlights the high heterogeneity in ASD. Overall, the lack of significant differences and weak classification indicates that, at the group level, adults with ASD show resting-state EEG that is remarkably similar to the non-autistic comparison group.

***Keywords*:** ASD, Machine Learning, Resting-state EEG.

1 Introduction

Autism spectrum disorder (ASD) is defined by persistent differences in social interactions, atypical sensory reactivity, and restricted and repetitive behavior (1). High heterogeneity exists among individuals receiving the diagnosis, since the criteria allow a broad spectrum of symptoms, and the neural mechanisms underlying ASD remain unclear. To elucidate the neurobiological mechanisms behind ASD, many studies have used neuroimaging (2–4). Discovery of biomarkers for ASD has the potential to support diagnosis and might disentangle the heterogeneity (5). Recently, there has been a growing interest in discovering resting-state electroencephalographic (EEG) biomarkers for various neuropsychiatric conditions, as EEG has good clinical practicality, due to being non-invasive, portable, widely available, and low cost.

Many different resting state EEG features have been investigated in regards to ASD, with spectral power being the most commonly used feature. Decreased theta power (6), alpha power (7–10) and gamma power have been observed in ASD (11, 12). However, increased alpha power (13) and gamma power have also been reported in ASD (14). Other spectral features, e.g. peak alpha frequency (15), theta/beta ratio (16) and asymmetry (17) have also been associated with

ASD. Besides spectral features, abnormal functional connectivity (18), microstates (19), and measurements of criticality (20, 21) have also been found in ASD. The role of each individual feature and their implications on ASD are outside the scope of this paper (for reviews, see 2, 3, 22).

In addition to identifying group mean differences of EEG features in ASD, many recent studies also investigated the potential of classifying ASD with predictive machine learning models. The advantages of predictive modelling is that it is optimized for predicting new subjects or future outcomes, which aligns with the goal of diagnostic/prognostic tools (23). Using linear support vector machines (SVM), Zhao et al. (15) was able to classify children with ASD with an accuracy of 81.7% using alpha power, 75.2% using peak alpha frequency, and using both alpha power and peak alpha frequency the overall accuracy was 92.7%. Kang et al. (24) also used a linear SVM, and estimated power features from all five canonical frequency bands, but only obtained a 68% accuracy, however when they the EEG data with eye tracking data they obtained a 85.4% accuracy. Another study used graph theoretical metrics estimated from EEG and obtained a 73.7% accuracy, and after combining EEG features with eye tracking features the accuracy improved to 95% (25). Alpha amplitude variability (standard deviation) was also found to be able to classify autistic children with an accuracy of 90.0% (16). Therefore, predictive machine learning has been fairly successful thus far in classifying ASD from non autistic participants based on EEG spectral power, obtaining even higher accuracies when EEG features were combined with eye tracking.

All of these studies, when viewed in isolation, paint a promising picture about the use of EEG for ASD prediction and reflect the interest in discovering biomarkers for ASD. However, the results are also limited due to either low sample sizes, imbalanced datasets, or lack of reproducibility because of differences in the choice of analysis. Additionally, many of the clinical machine learning models have also not been evaluated on an unseen dataset or employed cross-validation. Thus, currently there is no clear consensus about the physiological correlates of the disorder, which may relate to the complexity and heterogeneity of ASD (5).

To disentangle the potential role of EEG biomarkers for characterization of ASD, the present study estimated many of the commonly used resting-state EEG features that have shown promising results. Specifically, we source-modelled the EEG time series and computed spectral features (power, asymmetry, theta/beta ratio, peak alpha frequency and 1/f exponent), measures of criticality (long-range temporal correlations [DFA exponent] and functional excitation inhibition ratio [fEI]) and functional connectivity (coherence [Coh], imaginary part of coherence [Imcoh], phase locking value [PLV], weighted phase lag index [wPLI] and power envelope correlations [PEC]). A relatively large sample size of 186 participants were recruited and feature selection methods and machine learning models were applied to evaluate the biomarker potential of the selected EEG features in distinguishing between individuals with and without ASD. To not have our results be dependent on one particular machine learning algorithm, we trained variations of three commonly used classifiers, namely support vector machine (SVM), logistic regression and random forest. Additionally, to obtain a robust result, we employed repeated two-layer cross-validation, in order to train the models on a training set, estimate hyperparameters on a validation set and finally evaluate the generalization performance on an unseen separate test set (see Figure 1 for overall analysis framework). To our knowledge, no studies aiming to predict ASD using resting-state EEG with machine learning have employed such a comprehensive biomarker set, and stringent repeated cross-validation scheme for evaluating their models.

2 Results

2.1 Multiple EEG features exhibited correlation with age

For each EEG feature type, we computed how many features were significantly correlated with age after false discovery rate (FDR) correction (Pearson's correlations, $p < 0.05$) and observed that spectral power, theta/beta ratio, 1/f exponents and PEC had greater than chance-level correlations with age. Table 1 shows the percentage of all features within each feature type that were significantly correlated with age.

Further investigation revealed that the correlations between power, PEC and age were frequency-band dependent. Specifically, absolute delta, theta and alpha power, relative beta and gamma power, and delta PEC exhibited high correlations with age. To take into account the difference in age between the ASD and non-autistic comparison groups, we

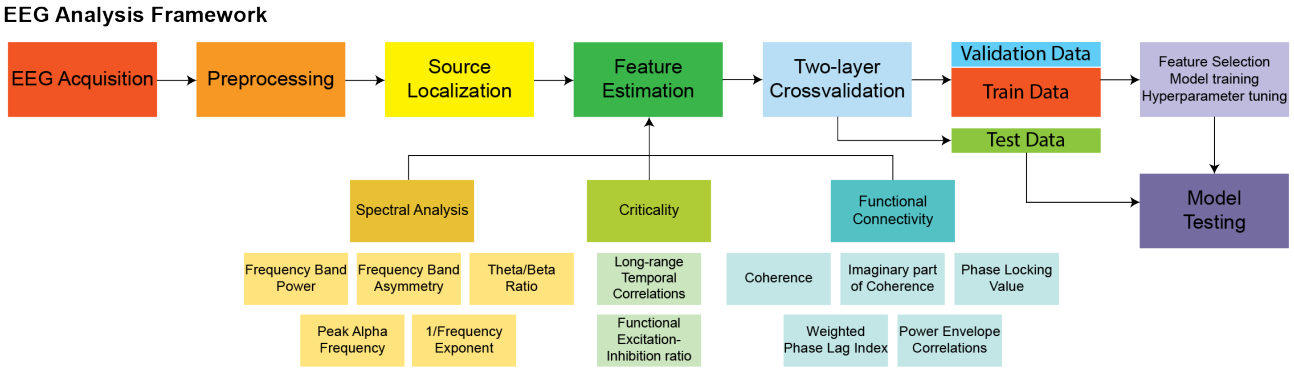

Fig. 1. Overview of the EEG analysis framework. The EEG data was preprocessed and source modelled, followed by estimation of commonly used EEG biomarkers. Repeated 10-by-10 two-layer cross-validation was performed to split the data. Feature selection, model training and hyperparamter tuning were performed on the inner fold training and validation set, while the generalization performance was evaluated on the entirely unseen test set. Logistic regression, random forest, and support vector machine were employed to classify ASD.

Table 1. Percentage of significant correlations with age

Power	Theta/beta	Asymmetry	Peak Alpha Freq
20.1	**100**	0.0	1.4
1/f	Coh	Imcoh	PLV
73.5	0.0	0.0	0.0
wPLI	PEC	DFA	fEI
0.0	**11.1**	1.2	0.3

Numbers in **bold** indicate more than 5% of the features were significantly correlated with age after FDR correction.

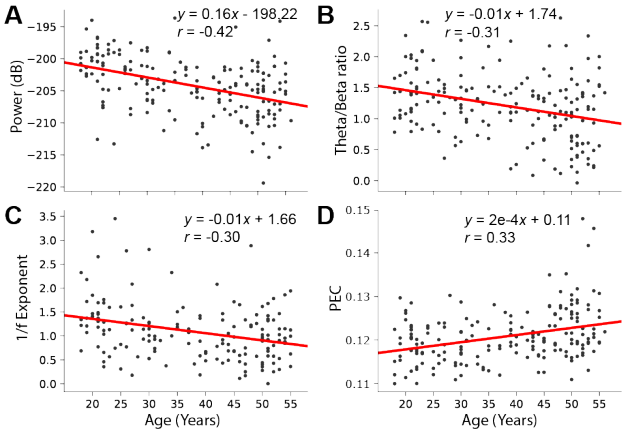

Fig. 2. Several EEG features exhibit age dependence. Scatter plots illustrating correlations between age and A) absolute power, B) theta/beta ratio, C) 1/f exponent, and D) PEC of source-modelled signals.

removed the linear trend for 1/f exponents, theta/beta ratio, and power and PEC in the above-mentioned frequency bands. Figure 2 show examples of the strongest correlations and age for each of the highlighted feature types prior to age-effect correction.

The importance of correcting for age-related effects can be observed when evaluating the mean differences in 1/f exponents between the ASD and non-autistic comparison group before and after age correction. Without age-effect correction, the 1/f exponents were significantly lower in the ASD group (Figure 3A). However, this effect was confounded by the negative correlation between age and 1/f exponents (Figure 2C), which becomes apparent after age-effect correction, where no significant effects were observed in 1/f exponents between the ASD and non-autistic comparison group (Figure 3B).

2.2 The ASD and non-autistic comparison group did not differ in group-mean or group- variability

No EEG features differed significantly between the ASD and non-autistic comparison group after age-effect correction (Figure 4 and 5). We performed permutation tests with FDR correction for each feature type separately, but did not observe any significant difference between the two groups ($p >$ 0.05 for all features).

Considering the large variation in symptomatology in

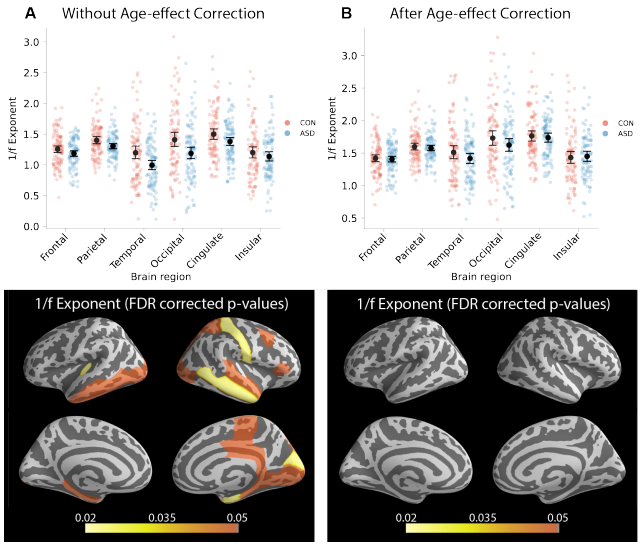

Fig. 3. **The effect of age on 1/f exponents.** A) Prior to age-effect correction of 1/f exponents, we observed a significant difference in group-means. B) However, this effect was confounded by age, as evident by the lack of significant effects after we corrected for the age-effect. Means with 95% confidence intervals are shown.

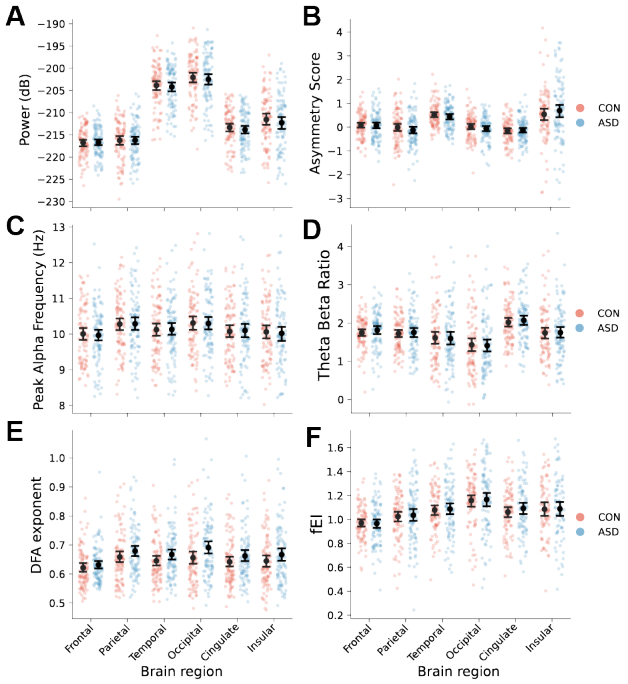

Fig. 4. **The ASD and non-autistic comparison groups have remarkably similar quantitative EEG features.** Representative examples of computed EEG feature values averaged across the brain regions for: A) Absolute Alpha Power, B) Alpha Asymmetry, C) Peak Alpha Frequency, D) Theta/Beta Ratio, E) Alpha DFA exponents, and F) Alpha fEI. Mean with 95% confidence intervals are shown.

ASD, it is plausible that the ASD and non-autistic comparison groups would show differences in variability in spite of the lack of mean differences. However, using Levene's test with FDR correction for each feature type separately, we did not observe any significant difference in variability between the two groups ($p > 0.05$).

2.3 Machine learning models based on EEG features predicted ASD around chance-level

The statistical tests indicated that each individual EEG feature was not significantly different between the two groups. To investigate whether combinations of features could potentially serve as biomarkers for prediction of ASD, we employed multivariate machine learning models. To estimate how well the models would potentially predict on new unseen subjects, we trained and tested using two-layer cross-validation. The data was divided into a training set, which the model was trained on, a validation set, which were used to tune the hyperparameters, and a test set, which was the unseen data that was used to estimate the generalization performance of the models. Although we obtained decent training and validation accuracies, the performance dropped to around chance-level for most of the models when tested on completely unseen data. The best full classification model, which had access to all the features, was logistic regression with L1 regularization with a balanced test accuracy of 50.0%. None of the four full classification models performed better than chance.

We also tested each feature type separately, i.e., the same four classification models were applied, but the feature selection was limited to each individual feature type. Here, the best classification performance was obtained by logistic regression with L1 regularization using peak alpha frequency with a balanced test accuracy of 55.8% (Figure 6), followed by PLV with a balanced test accuracy of 55.1% and fEI with a balanced test accuracy of 53.7%. All three feature types performed significantly better than chance (one-sided Wilcoxon signed-rank test $p < 0.05$), albeit the effects were small. Additionally, the performance of both peak alpha frequency and PLV was significantly better than chance-level in three out of the four classifiers, while fEI performed better than chance-level in two out of the four classifiers (See Supplementary

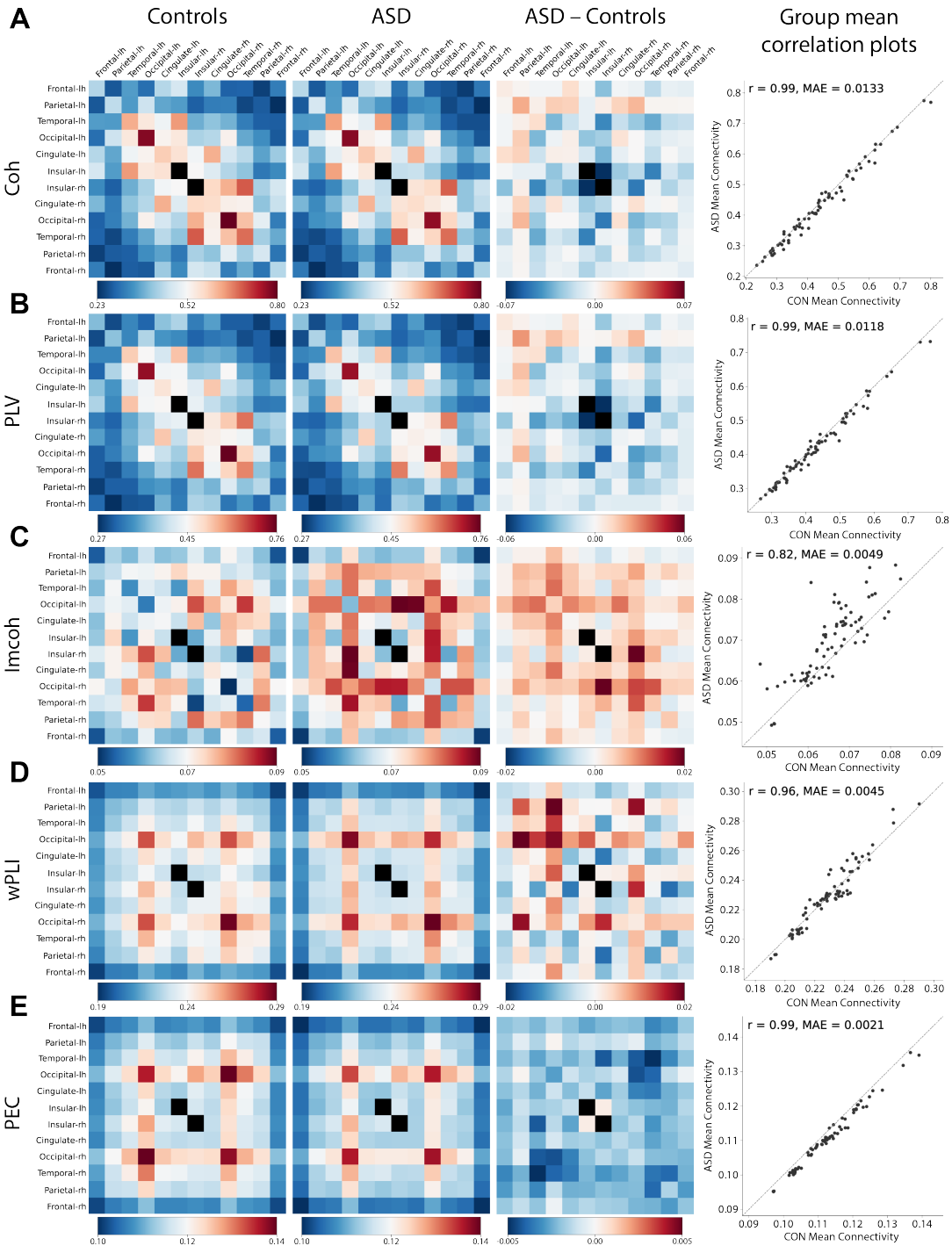

Fig. 5. **The ASD and non-autistic comparison groups have remarkably similar connectivity patterns.** Examples of the mean A) Coherence, B) Phase Locking Value, C) Imaginary Coherence, D) Weighted Phase Lag Index, and E) Power Envelope Correlations for the ASD and non-autistic comparison group, their difference, and correlation in the alpha band. The closer the points are to the diagonal line in the correlation plots, the better the correlation between the group-mean connectivities. Self-connectivity was excluded, hence the insular regions, which only contain one patch in each hemisphere, are colored black to reflect the missing value. Pearson's correlation was used to compute r. MAE, mean absolute error.

Table S1 for the performance of all combinations of features and classification models).

2.4 Mean-level predictions of AQ and SPQ

Besides classifying ASD, we also investigated if combinations of EEG features would be able to predict the autism-spectrum quotient short questionnaire (AQ) score and the visual and auditory scores of the sensory perception quotient short questionnaire (SPQvis and SPQaud) using linear regression. The best full model was linear regression with L1 regularization, which obtained a normalized mean absolute error (nMAE) of 1.051 for AQ. None of the models with access to all features obtained a nMAE below 1, hence they did not perform better than predicting the mean on any of the questionnaire scales. We also tested each feature type separately and found the best performance for linear regression with L1 norm using fEI, which obtained a test nMAE of 0.960 for prediction of AQ. This result was significantly better than mean prediction (one-sided Wilcoxon signed-rank test $p < 0.05$), suggesting that fEI variability may be related to autism severity. However, no significant correlations between fEI and AQ were found after FDR correction, and the prediction was merely 4% better than the mean prediction MAE. The only other feature type with better than mean-level predictions was the theta/beta ratio, which obtained a test nMAE of 0.993 using linear regression with L1 regularization, and a test nMAE of 0.979 using linear regression with L2 regularization for AQ prediction. All other features did not perform better than predicting the mean on any of the questionnaire scales (See Supplementary Table S2 for the performance of all combinations of features, regression models, and target variables).

3 Discussion

In spite of many attempts to identify EEG correlates of ASD, there is a lack of consistency of measures used and findings reported (for reviews, see 2, 3, 22). With the aim of finding something more robust, we recruited, to our knowledge, one of the largest single-center adult ASD cohorts, and tested a comprehensive set of EEG features and classification algorithms using two-layer cross-validation. This approach reinforced the picture that adults with ASD have remarkably similar resting-state EEG to adults without ASD. No group-mean or group-variance differences were observed across the 3443 features investigated, and our machine learning models only performed slightly better than chance-level.

One caveat with an exploratory analysis involving many different features is the need for multiple testing correction, which might have resulted in a relatively lower statistical power compared to studies conducted in a hypothesis-driven paradigm. However, to not be too conservative and be more relatable to the other resting-state EEG studies, we only corrected for the number of features within each feature type separately, and even with this consideration, we still did not achieve any significant group-mean or group-variance differences. Visual inspection of distributions of EEG feature values (e.g. Figure 4) also did not invoke the impression that the lack of significance is due to harsh multiple comparison correction.

Furthermore, the primary aim of our study was not to identify group-mean differences within single features, but to combine the many different features, and utilize machine learning to infer which combinations of features might have potential value as biomarkers for ASD. Even if each individual feature has small effect sizes, they might still provide relevant predictive information if they are combined in a multivariate model (26). However, our machine learning models were unable to find an EEG pattern characteristic for the ASD group.

This is inconsistent with some of the studies using machine learning on resting-state EEG in ASD, which reported up to 85% (24), 95% (25), and 96.4% accuracy (10). However, these three studies did not perform any cross-validation, which means they trained and tested on the same data and this has the caveat of the models overfitting to the data, which would result in the models being unable to generalize to new data (27–30). Thus the reported performance of their models would correspond to the high training accuracies we observed (e.g. the red points in Figure 6).

Other machine learning studies that employed cross-validation still obtained up to 93% (15, 31), 94% (32), and 99.7% (33) accuracy. However, the machine learning models they employed have hyperparameters, which have to be tuned. If a single layer of cross-validation is used for training and tuning, then there is no test set to estimate how well the model would generalize to new unseen subjects (28). It would correspond to the validation accuracies we observed, where we obtained more than 80% accuracy in multiple folds using

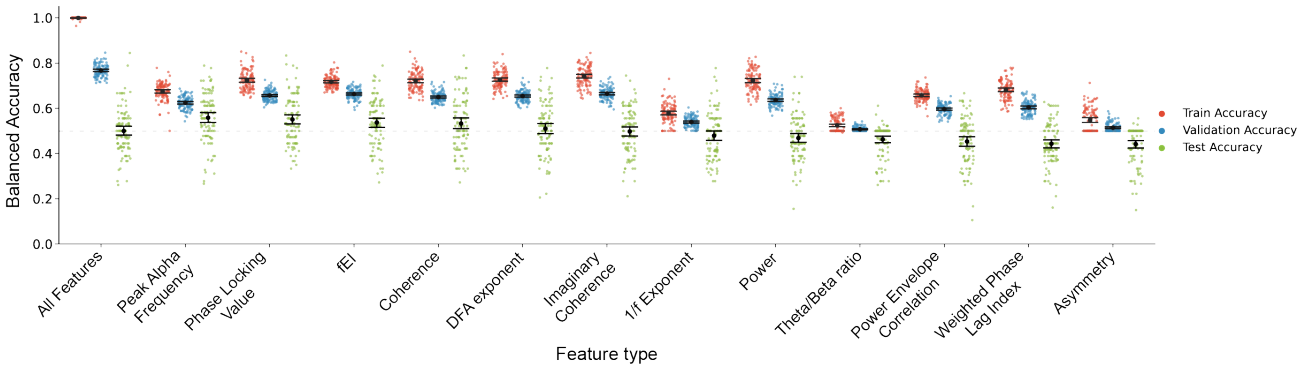

Fig. 6. Performance of each feature type with logistic regression with L1 regularization. The logistic regression with all features are shown on the left, followed by the performances of each individual feature type, sorted according to average balanced test accuracy (green points). The dashed line indicates chance-level performance. Most of the feature types displayed significant training (red points) and validation (blue points) accuracies, which did not survive when tested on completely unseen data (cf. Figure 2). Mean with 95% confidence intervals are shown.

the model with all features (e.g. the blue points in Figure 6).

To obtain a robust estimation of the predictive performance of our machine learning models, we employed two-layer cross-validation, where we tuned the hyperparameters and number of features in the inner layer, and tested the generalization performance in the outer layer. Surprisingly, although we obtained decent validation accuracies, there was a large drop in performance when we evaluated on entirely unseen data, to the point where our best models merely performed slightly better than chance-level. This finding is consistent with a recently published EEG study on ASD that obtained up to 57% balanced accuracy for ASD, when testing on unseen data using a nested cross-validation scheme 34. Interestingly, we observed a high variance in test accuracies, reflecting how big an impact the exact split of the data could have on our machine learning models.

One of the most common arguments for inconsistent results across different studies is that methodological differences confound the results. However, the high variation we observed in test accuracies suggest that even if the facilities, equipment, preprocessing steps, algorithms, and researchers are the same, the results can still vary substantially based on the specific participants in the sample. Specifically, we obtained performance values ranging from not better than chance-level up to around 80% balanced test accuracy on classifying ASD. This clearly highlights the highly heterogeneous nature of ASD, and combined with differing recruitment methods, facilities, analysis methods and relatively small sample sizes in clinical ASD EEG studies, it is not surprising that inconsistent findings occur, and that robust neural biomarkers have not yet been found for ASD. The high variance in test accuracies also stress the importance of proper model evaluation schemes, e.g. two-layer cross-validation for models with hyperparameters (28), for identification of robust biomarkers.

The problem with small sample sizes not being representative of the clinical population is also clearly illustrated by the paradox that machine learning models seem to perform worse on more neuroimaging data (23, 35). This negative correlation with sample sizes does not mean that the models trained on more data are worse, but instead that with few samples the models are more likely to overfit to bias within the sample, that is not reflective for the whole clinical population. Thus, subtle within group differences might be misinterpreted as between group differences. Additionally, this issue might be further inflated by publication bias (36).

Apart from methodological differences, there are theoretical considerations that should be taken into account as well. The definition of ASD has become a topic of interest recently for three reasons: Firstly, ASD prevalence has more than tripled in recent years, from 0.67% in the year 2000 to 2.3% in 2018 (37). Secondly, the criteria for ASD according to the DSM has changed considerably in the past decades, most prominently between its fourth and fifth edi-

tion where multiple previously distinct conditions have been merged into the current condition "Autism Spectrum Disorder". Thirdly, it has been shown that psychological and neurological effect sizes in ASD research have been continually decreasing in the past decades while no such effect was shown for schizophrenia research (38). While the increase in ASD prevalence could be attributed to an increased public awareness or other external factors such as pollution (39), it can also partly be attributed to ASD being used for a broader spectrum of symptoms and disorders. The increased heterogeneity could partly explain the decreased effect sizes in ASD research, since higher heterogeneity would result in increased within group variance (40). Our results also support the notion of high heterogeneity in ASD.

We acknowledge that the present study has limitations. The machine learning classifiers were trained on commonly used EEG features and are thus dependent on these specific feature types. We developed a comprehensive framework to encompass a broad number of feature types to investigate whether each feature or combinations of features could characterize ASD, but other combinations of EEG features remain unexplored, e.g. graph theory based network metrics, entropy, and microstates have also been associated with ASD (41–43).

Another limitation to the present study is the experimental paradigm. We investigated the EEG activity during 5 min eyes-closed rest, which might be different from eyes-open rest or task-based paradigms (44). One previous study found group differences in alpha power and coherence during eyes-open EEG in adults with ASD, but not in the eyes-closed condition (13).

The study sample should also be highlighted. The ASD group is comprised of adults with average or above average intelligence, often employed, and well educated, while many previous ASD EEG studies focused on children. One might argue that because many individuals in the ASD group were not diagnosed at an early age, they might have less pronounced autistic traits or been better at compensatory strategies, however we did not see an association between AQ score and age of diagnosis (Supplementary Figure S1).

Overall, our results indicate that adults with ASD have eyes-closed resting-state EEG activity within the typical range. This does not necessarily mean that resting-state EEG traces contain no meaningful information about ASD, however, the electrophysiological effects were likely too subtle to be picked up by our models relative to the high heterogeneity in ASD, even with a sample size of close to 200 participants. Future studies should try to increase effect sizes or mitigate the effect of heterogeneity by increasing sample sizes to better represent the whole ASD population and/or look for prototypes (40, 45) or distinct subtypes within ASD. Identification of subtypes might improve diagnosis and enable better tailored treatment on an individual/subgroup level (23, 35, 46). Differences might also be more prominent during specific tasks targeted at the autistic traits, e.g. crucial differences could be picked up by employing interactive paradigms, and studying interpersonal mechanisms in ASD (47–52). Encompassing longitudinal data or data from multiple modalities might also increase effect sizes (26), e.g. combining genetics (heritability was estimated to be around 83% for ASD; 53), neuroimaging, psychological, and social information. The analysis framework we developed can be readily expanded to integrate with all of these modalities and support the development of a biopsychosocial model for ASD (54, 55).

4 Methods

4.1 Participants

The participants were part of a larger project investigating mechanisms underlying ASD in adults and recruited through the Netherlands Autism Register (56). Inclusion criteria for the ASD cohort were a clinical diagnosis of ASD (according to DSM-5; 1), Asperger's syndrome, pervasive developmental disorder-not otherwise specified, autism (according to DSM-IV; 57), and age between 18 and 55 years. Exclusion criteria for the non-autistic comparison group was a diagnosis of ASD or a diagnosis of ASD in a direct family member. Table 2 presents the characteristics of the included participants. The protocol of this study was approved by the ethics committee of the VU University Medical Center (approval number 2013/45). The study was conducted in accordance with the guidelines and regulations approved by the respective ethical committee and in compliance with the provisions of the declaration of Helsinki. All participants provided informed consent and were financially reimbursed. Data and scripts from this study are available upon request.

Table 2. Characteristics of study participants

	CON	ASD
Sample Size	91	95
Mean Age (SD), years	32.34 (12.25)	43.65 (8.99)
Sex (Female), %	57.1	53.7
Mean AQ (SD)	50.97 (10.05)	85.73 (9.88)
Mean Auditory SPQ (SD)	9.13 (2.51)	6.66 (2.51)
Mean Visual SPQ (SD)	10.54 (2.98)	7.07 (3.15)

AQ, Autism-spectrum Quotient. SPQ, Sensory Perception Quotient. SD, standard deviation.

4.2 Clinical Measures

Autism-Spectrum Quotient Short Questionnaire

The autism-spectrum quotient short questionnaire (hereafter AQ; 58) is an abridged version of the autism-spectrum quotient (59). It consists of 28 self-report items, which can be clustered into two main factors: Social Behavior, and Numbers and Patterns. The items under Social Behavior can further be divided into the subfactors: Social Skills, Routine, Attention Switching, and Imagination. All items are scored on a four-point Likert scale and range from "Definitely agree" to "Definitely disagree". The items are summed to obtain total AQ scores. Higher AQ score suggest more autistic traits.

Sensory Perception Quotient Short Questionnaire

The sensory perception quotient short questionnaire (hereafter SPQ; 60, 61) is a 35-item self-report assessing sensory perception of touch, smell, vision, hearing, and taste. We only used the items pertaining to the factors vision (6 items) and hearing (5 items). All items are scored on a four-point Likert scale and range from "Strongly agree" to "Strongly disagree". The items are summed to obtain factor scores (SPQvis and SPQaud). Lower scores on the SPQ suggest a lower sensory threshold, thus higher sensory sensitivity.

4.3 EEG acquisition

Resting-state EEG was recorded during 5 min of eyes-closed rest with a 64-channel BioSemi system sampled at 2048 Hz. The participants received the instructions: "Please keep your eyes closed, relax, and try not to fall asleep". Impedance across all electrodes was kept below 5 kΩ. Additionally, four electrodes were placed at the left and right outer canthi to capture horizontal eye movements, and an electrode underneath each eye for vertical eye movements and blinking.

4.4 EEG pre-processing

The EEG data were processed using MNE-Python 0.24.3 (62). First, the data were band-pass filtered at 1 to 100 Hz, notch filtered at 50 Hz, downsampled to 500 Hz, and divided into 4 second epochs without overlap. Bad epochs and channels with gross non-ocular artefacts were rejected by visual inspection. The data were re-referenced to the common average and ocular and ECG artefacts were removed using Piccard independent component analysis (63) with the number of components set to 32. Autoreject 0.2.2 (64) was employed to catch any remaining artefacts, which guided a final visual inspection. Subjects with less than 2 min of clean signal were excluded from further analysis (n = 3), resulting in a total sample size of 95 ASD and 91 non-autistic participants. The EEG features were estimated in the five canonical frequency bands: delta (1.25–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), beta (13–30 Hz), and gamma (30–48 Hz), unless otherwise stated.

4.5 EEG source localization

We used L2 minimum norm estimation, as implemented by MNE-Python, to obtain cortical current estimates from our sensor level data. The FreeSurfer average brain template from FreeSurfer 6 (65) was used to construct the boundary element head model and forward operator for the source modelling. The regularization parameter was set to $\lambda^2 = 1/9$. A diagonal matrix with 0.2 µV values was used for the covariance matrix, which was the default values for EEG provided by MNE-Python. Unconstrained orientations were allowed, and principal component analysis was employed on the whole source time series at each vertex to reduce the three-dimensional signals to one-dimensional time series of the dominant principal component. The time series of the 20484 source vertices were further collapsed into 68 cortical patches based on the Desikan Killiany atlas, by first aligning the dipole orientations by shifting vertices with opposite polarity to the majority of vertices by π, followed by averaging the amplitudes of all vertices within a patch. The phase shifting prevents the vertices with opposite polarities from canceling each other out during the averaging operation. Some of the estimated EEG features were computed across cortical brain regions, i.e. patches within frontal, parietal, temporal, occipital, cingulate, and insular regions were averaged. The individual patches were

mapped to the brain regions according to the appendix in (66).

4.6 EEG feature estimation

A brief description of each EEG feature type is provided (see Figure 1 for overview of all features), as all the estimated features are already well-established. More detailed information and equations can be found in the supplementary information.

4.6.1 Spectral analysis

Multitaper spectral estimation (67) was employed to estimate absolute and relative power in the five canonical frequency bands. EEG power asymmetry was calculated by subtracting each left hemispheric patch from the corresponding patch in the right hemisphere (68), followed by averaging across the brain regions. Theta/beta ratio was computed by dividing theta power by beta power and averaged in brain regions. Peak alpha frequency and 1/f exponent was estimated using the FOOOF algorithm (69).

4.6.2 Criticality

Long-range temporal correlations (70) were estimated following the detrended fluctuation analysis (DFA) procedure described by (71) to obtain DFA exponents. Functional excitation/inhibition ratio (fEI) was estimated following 20.

4.6.3 Functional connectivity

All functional connectivity measurements were computed for all pairwise combinations of the 68 brain patches. This yields 2346 connections, which results in 56950 features when taking into account all connections are calculated for every frequency bands and five different connectivity measurements. This vast number is many times higher than the number of subjects we have, which can lead to overfitting for the machine learning models (also known as curse of dimensionality; 72). To reduce the number of dimensions, we averaged the connectivity measurements in brain regions and for each hemisphere, thus reducing the number of connections to 76, which across frequency bands and five connectivity feature types resulted in 1900 features.

Coherence (Coh) was 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 (73). The imaginary part of coherence (Imcoh) was also used as a standalone feature, due to its lower sensitivity towards volume conduction (74). We also computed the phase locking value (PLV), which measures connectivity as a function of phase difference variability (75), and weighted phase lag index (wPLI), which also measures phase synchronization but is less sensitive towards volume conduction (76). Lastly, power envelope correlations (PEC) were estimated following (77).

4.7 Prediction

All the EEG features were combined and turned into a data matrix of number of subjects by 3443 features. Given that there were more features than samples, and in order to decrease the dimensionality and reduce overfitting (72), we applied multiple dimensionality-reduction methods. First, minimal-redundancy-maximum-relevance (mRMR) (78) was applied to each EEG feature type to attenuate the effect of imbalance in the number of features between each of the EEG feature types. To not have our results be dependent on one particular machine learning algorithm, we trained variations of four commonly used classifiers to predict ASD: 1) support vector machine (SVM) with recursive feature elimination, 2) logistic regression with Ridge regularization (L2) and sequential forward selection, 3) logistic regression with Lasso regularization (L1) and 4) random forest. For prediction of questionnaire scores, we employed linear regression with either L1 or L2 norm. Model training, hyperparameter tuning and evaluation of the models were conducted in a stratified 10-by-10 fold two-layer cross-validation scheme repeated 10 times to take into account the effect of random splits, resulting in 100 final models (10 repetitions of 10-fold outer cross-validation; Figure 7). The stratification ensured the proportion of people with ASD were balanced across folds. Balanced accuracy, the mean of sensitivity and specificity, was estimated for the classifier performances. A balanced accuracy around 50% indicates chance-level prediction for a two-class classification. Normalized mean absolute error (nMAE), the ratio of MAE of the model to the MAE obtained when predicting the mean value of the training set, was computed for evaluation of the regression models. A nMAE = 1 indicates prediction around the level of just predicting the mean, while a lower nMAE indicates superior performance compared to predicting the mean value. We used mlxtend 0.19.0 (79) for sequential forward selection, scikit-learn 1.0.1 (80) for most the machine learning algorithms, and a python implementation of mRMR (81).

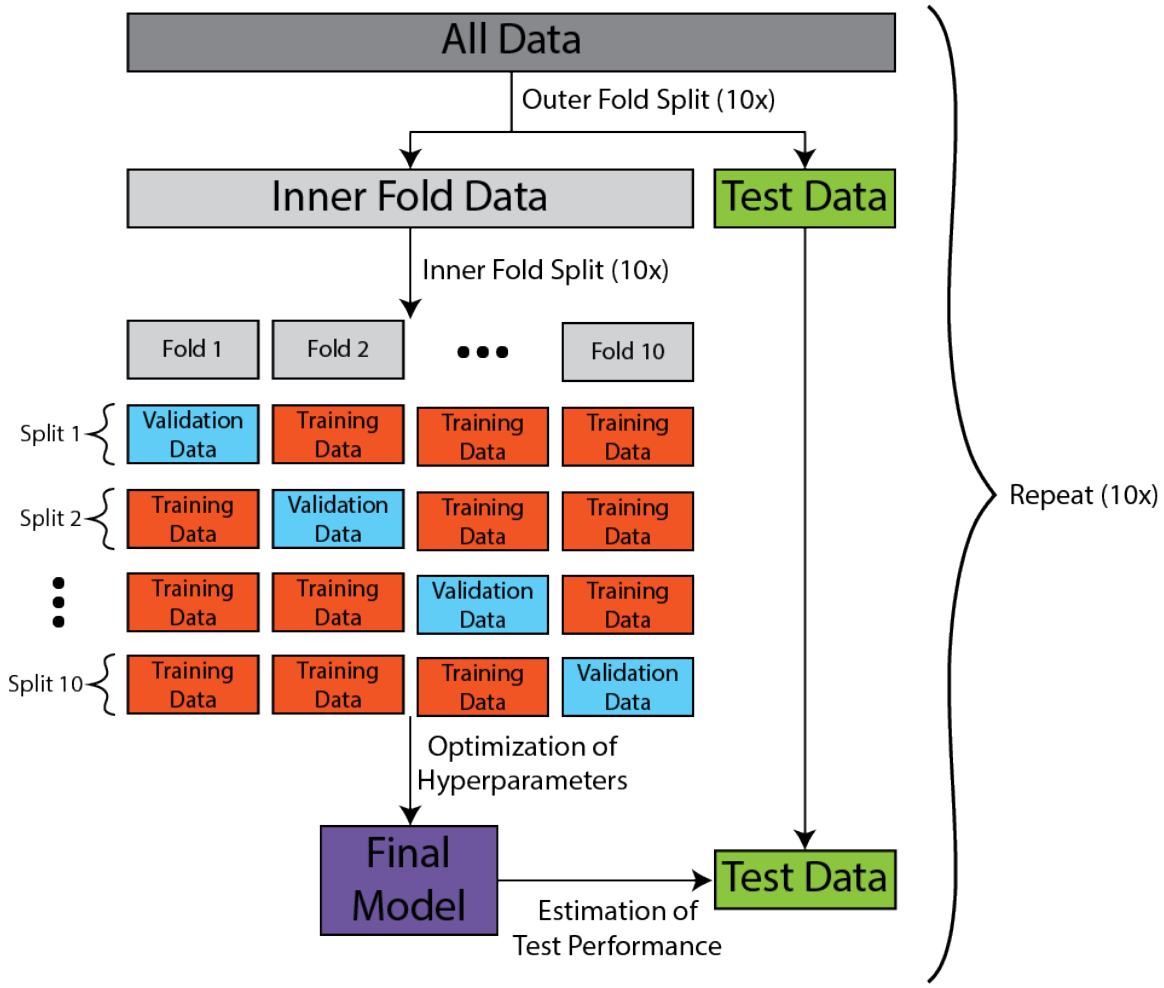

Fig. 7. Schematic of the repeated two-layer cross-validation scheme. In order to obtain robust and reliable estimations of how well the machine learning models would perform on unseen data, we employed 10 repetitions of 10-by-10 two-layer cross-validation. The outer fold split ensured we tested on unseen data, while the inner fold split alleviates overfitting when we optimized model hyperparameters, i.e., regularization strength for logistic regression and SVM, number of trees and depth for random forest, and the number of features used by the models after feature selection.

4.8 Age-effect correction

It is well-established that some EEG features have an age-dependent effect. Both peak alpha frequency (82), theta/beta ratio (83) and 1/f exponents have been found to be lower in old compared to young adults (69, 84–86). DFA exponents have also been found to increase from childhood into early adulthood before it stabilizes (87). Although the age range of both groups fell between 19 and 55 years, the ASD group was, on average, 11 years older than the non-autistic comparison group (Permutation test, $p = 0.0001$). We investigated the linear correlations between each EEG feature type and age. For all the features that showed more significant Pearson's correlations than chance-level after false-discovery rate (FDR) correction, we corrected for the age-effect by removing the linear trend.

4.9 Statistical analysis

Results are shown as mean with 95% confidence intervals. A non-parametric permutation test with FDR correction was used to test for group mean differences. Levene's test with FDR correction was used to assess equality of variances.

Pearson's correlation coefficient and Student's t-distribution was used to test for significant linear correlations. Non-parametric Wilcoxon signed-rank test was used to test differences between classifiers. One-sided tests were used for baseline comparisons (chance-level for the classifiers or mean prediction for the regressors). A p-value < 0.05 was considered significant for rejection of the null hypothesis.

References

[1] American Psychiatric Association. *The Diagnostic and Statistical Manual of Mental Disorders (DSM-5)*. American Psychiatric Association, 2013, pp. 1–947.

[2] Jun Wang et al. "Resting state EEG abnormalities in autism spectrum disorders". In: *Journal of Neurodevelopmental Disorders* 5.1 (2013), p. 1.

[3] Christian O'Reilly, John D. Lewis, and Mayada Elsabbagh. "Is functional brain connectivity atypical in autism? A systematic review of EEG and MEG studies". In: *PLoS ONE* 12.5 (2017), e0175870.

[4] Stefan Holiga et al. "Patients with autism spectrum disorders display reproducible functional connectivity alterations". In: *Science Translational Medicine* 11.481 (2019).

[5] Corey Horien et al. "Functional connectome-based predictive modelling in autism". In: *Biological Psychiatry* 0.0 (2022).

[6] Taylor Hornung et al. "Dopaminergic hypo-activity and reduced theta-band power in autism spectrum disorder: A resting-state EEG study". In: *International Journal of Psychophysiology* 146 (2019), pp. 101–106.

[7] Sarah Pierce et al. "Associations between sensory processing and electrophysiological and neurochemical measures in children with ASD: an EEG-MRS study". In: *Journal of Neurodevelopmental Disorders* 13.1 (2021).

[8] Lisa E Mash et al. "Atypical relationships between spontaneous EEG and fMRI activity in Autism". In: *Brain Connectivity* 10.1 (2020), pp. 18–28.

[9] Brandon Keehn et al. "Autism, Attention, and Alpha Oscillations: An Electrophysiological Study of Attentional Capture". In: *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* 2.6 (2017), pp. 528–536.

[10] Ali Sheikhani et al. "Detection of abnormalities for diagnosing of children with autism disorders using of quantitative electroencephalography analysis". In: *Journal of Medical Systems* 36.2 (2012), pp. 957–963.

[11] Rachel R. Romeo et al. "Parental Language Input Predicts Neuroscillatory Patterns Associated with Language Development in Toddlers at Risk of Autism". In: *Journal of Autism and Developmental Disorders* (2021), pp. 1–15.

[12] Christina R. Maxwell et al. "Atypical Laterality of Resting Gamma Oscillations in Autism Spectrum Disorders". In: *Journal of Autism and Developmental Disorders* 45.2 (2015), pp. 292–297.

[13] Karen J. Mathewson et al. "Regional EEG alpha power, coherence, and behavioral symptomatology in autism spectrum disorder". In: *Clinical Neurophysiology* 123.9 (2012), pp. 1798–1809.

[14] Eric van Diessen et al. "Increased power of resting-state gamma oscillations in autism spectrum disorder detected by routine electroencephalography". In: *European Archives of Psychiatry and Clinical Neuroscience* 265.6 (2015), pp. 537–540.

[15] Jie Zhao et al. "A study on EEG feature extraction and classification in autistic children based on singular spectrum analysis method". In: *Brain and Behavior* 10.12 (2020), e01721.

[16] Agnes S. Chan and Winnie W.M. Leung. "Differentiating autistic children with quantitative encephalography: A 3-month longitudinal study". In: *Journal of Child Neurology* 21.5 (2006), pp. 391–399.

[17] Courtney P. Burnette et al. "Anterior EEG asymmetry and the modifier model of autism". In: *Journal of Autism and Developmental Disorders* 41.8 (2011), pp. 1113–1124.

[18] Tianyi Zhou et al. "Early childhood developmental functional connectivity of autistic brains with non-negative matrix factorization". In: *NeuroImage: Clinical* 26 (2020), p. 102251.

[19] D'Croz-Baron DF et al. "EEG Microstates Analysis in Young Adults With Autism Spectrum Disorder During Resting-State". In: *Frontiers in human neuroscience* 13 (2019).

[20] Hilgo Bruining et al. "Measurement of excitation-inhibition ratio in autism spectrum disorder using critical brain dynamics". In: *Scientific Reports* 10.1 (2020), pp. 1–15.

[21] Huibin Jia and Dongchuan Yu. "Attenuated long-range temporal correlations of electrocortical oscillations in patients with autism spectrum disorder". In: *Developmental Cognitive Neuroscience* 39 (2019), p. 100687.

[22] Jennifer J. Newson and Tara C. Thiagarajan. "EEG Frequency Bands in Psychiatric Disorders: A Review of Resting State Studies". In: *Frontiers in Human Neuroscience* 12 (2019), p. 521.

[23] Mohammad R. Arbabshirani et al. "Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls". In: *NeuroImage* 145 (2017), pp. 137–165.

[24] Jiannan Kang et al. "The identification of children with autism spectrum disorder by SVM approach on EEG and eye-tracking data". In: *Computers in Biology and Medicine* 120 (2020).

[25] Shasha Zhang et al. "Children ASD Evaluation Through Joint Analysis of EEG and Eye-Tracking Recordings With Graph Convolution Network". In: *Frontiers in Human Neuroscience* 15 (2021), p. 651349.

[26] Eva Loth et al. "The meaning of significant mean group differences for biomarker discovery". In: *PLoS computational biology* 17.11 (2021), e1009477.

[27] Steven Lemm et al. "Introduction to machine learning for brain imaging". In: *NeuroImage* 56.2 (2011), pp. 387–399.

[28] Mahan Hosseini et al. "I tried a bunch of things: The dangers of unexpected overfitting in classification of brain data". In: *Neuroscience and Biobehavioral Reviews* 119 (2020), pp. 456–467.

[29] Russell A Poldrack, Grace Huckins, and Gael Varoquaux. *Establishment of Best Practices for Evidence for Prediction: A Review*. 2020.

[30] Barnaly Rashid and Vince Calhoun. "Towards a brain-based predictome of mental illness". In: *Human Brain Mapping* 41.12 (2020), pp. 3468–3535.

[31] Enzo Grossi, Chiara Olivieri, and Massimo Buscema. "Diagnosis of autism through EEG processed by advanced computational algorithms: A pilot study". In: *Computer Methods and Programs in Biomedicine* 142 (2017), pp. 73–79.

[32] Shixin Peng et al. "Early Screening of Children With Autism Spectrum Disorder Based on Electroencephalogram Signal Feature Selection With L1-Norm Regularization". In: *Frontiers in Human Neuroscience* 15 (2021), p. 656578.

[33] Ridha Djemal et al. "EEG-Based computer aided diagnosis of autism spectrum disorder using wavelet, entropy, and ANN". In: *BioMed Research International* 2017 (2017).

[34] Pilar Garcés et al. "Resting state EEG power spectrum and functional connectivity in autism: a cross-sectional analysis". In: *Molecular Autism* 13.1 (2022), pp. 1–16.

[35] Choong Wan Woo et al. "Building better biomarkers: Brain models in translational neuroimaging". In: *Nature Neuroscience* 20.3 (2017), pp. 365–377.

[36] P. J. Easterbrook et al. "Publication bias in clinical research". In: *The Lancet* 337.8746 (1991), pp. 867–872.

[37] CDC. *Data & Statistics on Autism Spectrum Disorder*. 2022.

[38] Eya Mist Rødgaard et al. "Temporal Changes in Effect Sizes of Studies Comparing Individuals with and Without Autism: A Meta-analysis". In: *JAMA Psychiatry* 76.11 (2019), pp. 1124–1132.

[39] Giovanni Imbriani et al. "Early-life exposure to environmental air pollution and autism spectrum disorder: A review of available evidence". In: *International Journal of Environmental Research and Public Health* 18.3 (2021), pp. 1–24.

[40] Laurent Mottron. "A radical change in our autism research strategy is needed: Back to prototypes". In: *Autism Research* 14.10 (2021), pp. 2213–2220.

[41] Pablo Barttfeld et al. “Organization of brain networks governed by long-range connections index autistic traits in the general population”. In: *Journal of Neurodevelopmental Disorders* 5.1 (2013), pp. 1–9.

[42] William Bosl et al. “EEG complexity as a biomarker for autism spectrum disorder risk”. In: *BMC Medicine* 9.1 (2011), pp. 1–16.

[43] Huibin Jia and Dongchuan Yu. “Aberrant Intrinsic Brain Activity in Patients with Autism Spectrum Disorder: Insights from EEG Microstates”. In: *Brain Topography* 32.2 (2019), pp. 295–303.

[44] Meera E. Modi and Mustafa Sahin. “Translational use of event-related potentials to assess circuit integrity in ASD”. In: *Nature Reviews Neurology* 13.3 (2017), pp. 160–170.

[45] Michael V. Lombardo. “Prototyping as subtyping strategy for studying heterogeneity in autism”. In: *Autism Research* 14.10 (2021), pp. 2224–2227.

[46] Seok Jun Hong et al. “Toward Neurosubtypes in Autism”. In: *Biological Psychiatry* 88.1 (2020), pp. 111–128.

[47] Hiroki C. Tanabe et al. “Hard to ”tune in”: Neural mechanisms of live face-to-face interaction with high-functioning autistic spectrum disorder”. In: *Frontiers in Human Neuroscience* 6 (2012), p. 268.

[48] Leonhard Schilbach et al. “Toward a second-person neuroscience”. In: *Behavioral and Brain Sciences* 36.4 (2013), pp. 393–414.

[49] Victoria Leong and Leonhard Schilbach. “The promise of two-person neuroscience for developmental psychiatry: using interaction-based sociometrics to identify disorders of social interaction”. In: *The British Journal of Psychiatry* 215.5 (2019), pp. 636–638.

[50] Yafeng Pan and Xiaojun Cheng. “Two-Person Approaches to Studying Social Interaction in Psychiatry: Uses and Clinical Relevance”. In: *Frontiers in Psychiatry* 11 (2020), p. 301.

[51] Jana A. Kruppa et al. “Brain and motor synchrony in children and adolescents with ASD - A fNIRS hyperscanning study”. In: *Social Cognitive and Affective Neuroscience* 16.1-2 (2021), pp. 103–116.

[52] Leonhard Schilbach. “Autism and other disorders of social interaction: where we are and where to go from here”. In: *European Archives of Psychiatry and Clinical Neuroscience* 272.2 (2022), pp. 173–175.

[53] Sven Sandin et al. “The heritability of autism spectrum disorder”. In: *JAMA - Journal of the American Medical Association* 318.12 (2017), pp. 1182–1184.

[54] Nancy Doyle. “Neurodiversity at work: a biopsychosocial model and the impact on working adults”. In: *British Medical Bulletin* 135.1 (2020), p. 108.

[55] Cristina Panisi and Marina Marini. “Dynamic and Systemic Perspective in Autism Spectrum Disorders: A Change of Gaze in Research Opens to A New Landscape of Needs and Solutions”. In: *Brain Sciences* 12.2 (2022).

[56] Netherlands Autism Register. *Netherlands Autism Register*. 2022.

[57] American Psychiatric Association. *The Diagnostic and Statistical Manual of Mental Disorders (DSM-IV)*. American Psychiatric Association, 1994, pp. 1–358.

[58] Rosa A. Hoekstra et al. “The construction and validation of an abridged version of the autism-spectrum quotient (AQ-short)”. In: *Journal of Autism and Developmental Disorders* 41.5 (2011), pp. 589–596.

[59] Simon Baron-Cohen et al. “The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians”. In: *Journal of Autism and Developmental Disorders* 31.1 (2001), pp. 5–17.

[60] Teresa Tavassoli, Rosa A. Hoekstra, and Simon Baron-Cohen. “The Sensory Perception Quotient (SPQ): Development and validation of a new sensory questionnaire for adults with and without autism”. In: *Molecular Autism* 5.1 (2014), pp. 1–10.

[61] Ricarda F. Weiland et al. “The Dutch Sensory Perception Quotient-Short in adults with and without autism”. In: *Autism* 24.8 (2020), pp. 2071–2080.

[62] Alexandre Gramfort et al. “MEG and EEG data analysis with MNE-Python”. In: *Frontiers in Neuroscience* 7 (2013), pp. 1–13.

[63] Pierre Ablin, Jean Francois Cardoso, and Alexandre Gramfort. “Faster independent component analysis by preconditioning with hessian approximations”. In: *IEEE Transactions on Signal Processing* 66.15 (2018), pp. 4040–4049.

[64] Mainak Jas et al. “Autoreject: Automated artifact rejection for MEG and EEG data”. In: *NeuroImage* 159.December 2016 (2017), pp. 417–429.

[65] Bruce Fischl. “FreeSurfer”. In: *NeuroImage* 62.2 (2012), pp. 774–781.

[66] Arno Klein and Jason Tourville. “101 labeled brain images and a consistent human cortical labeling protocol”. In: *Frontiers in Neuroscience* 0.DEC (2012), p. 171.

[67] Behtash Babadi and Emery N. Brown. “A review of multitaper spectral analysis”. In: *IEEE Transactions on Biomedical Engineering* 61.5 (2014), pp. 1555–1564.

[68] Steven K Sutton and Richard J Davidson. “PREFRONTAL BRAIN ASYMMETRY : Inhibition Systems”. In: *Psychological science* 8.3 (1997), pp. 204–210.

[69] Thomas Donoghue et al. “Parameterizing neural power spectra into periodic and aperiodic components”. In: *Nature Neuroscience* 23.12 (2020), pp. 1655–1665.

[70] Klaus Linkenkaer-Hansen et al. “Long-range temporal correlations and scaling behavior in human brain oscillations”. In: *Journal of Neuroscience* 21.4 (2001), pp. 1370–1377.

[71] Richard Hardstone et al. “Detrended fluctuation analysis: A scale-free view on neuronal oscillations”. In: *Frontiers in Physiology* (2012), pp. 1–13.

[72] Naomi Altman and Martin Krzywinski. “The curse(s) of dimensionality”. In: *Nature Methods* 15.6 (2018), pp. 399–400.

[73] Paul L. Nunez et al. “EEG coherency I: Statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales”. In: *Electroencephalography and Clinical Neurophysiology* 103.5 (1997), pp. 499–515.

[74] Guido Nolte et al. “Identifying true brain interaction from EEG data using the imaginary part of coherency”. In: *Clinical Neurophysiology* 115.10 (2004), pp. 2292–2307.

[75] Jean Philippe Lachaux et al. “Measuring phase synchrony in brain signals”. In: *Human Brain Mapping* 8.4 (1999), pp. 194–208.

[76] Cornelis J. Stam, Guido Nolte, and Andreas Daffertshofer. “Phase lag index: Assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources”. In: *Human Brain Mapping* 28.11 (2007), pp. 1178–1193.

[77] Russell T. Toll et al. “An Electroencephalography Connectomic Profile of Posttraumatic Stress Disorder”. In: *The American journal of psychiatry* 177.3 (2020), pp. 233–243.

[78] Hanchuan Peng, Fuhui Long, and Chris Ding. “Feature selection based on mutual information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27.8 (2005), pp. 1226–1238.

[79] Sebastian Raschka. “MLxtend: Providing machine learning and data science utilities and extensions to Python’s scientific computing stack”. In: *Journal of Open Source Software* 3.24 (2018), p. 638.

[80] Fabian Pedrogosa et al. “Scikit-learn: Machine Learning in Python Fabian”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[81] Samuele Mazzanti. *mRMR Python Implementation*. 2021.

[82] Thomas H Grandy et al. “Peak individual alpha frequency qualifies as a stable neurophysiological trait marker in healthy younger and older adults”. In: *Psychophysiology* 50.6 (2013), pp. 570–582.

[83] Anna J Finley et al. “Periodic and aperiodic contributions to theta-beta ratios across adulthood”. In: *Psychophysiology* (2022), pp. 1–20.

[84] Bradley Voytek et al. “Age-related changes in 1/f neural electrophysiological noise”. In: *Journal of Neuroscience* 35.38 (2015), pp. 13257–13265.

[85] Marius Tröndle et al. “Decomposing age effects in EEG alpha power”. In: *bioRxiv* (2021).

[86] Ashley Merkin et al. “Age differences in aperiodic neural activity measured with resting EEG”. In: *bioRxiv* (2021), pp. 1–30.

[87] Dirk J.A. Smit et al. “Scale-free modulation of resting-state neuronal oscillations reflects prolonged brain maturation in humans”. In: *Journal of Neuroscience* 31.37 (2011), pp. 13128–13136.

[88] Christopher Schölzel. *Nonlinear measures for dynamical systems (Nolds)*. 2019.

[89] Martin Vinck et al. “An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias”. In: *NeuroImage* 55.4 (2011), pp. 1548–1565.

[90] Martin Hardmeier et al. “Reproducibility of functional connectivity and graph measures based on the phase lag index (PLI) and weighted phase lag index (wPLI) derived from high resolution EEG”. In: *PLoS ONE* 9.10 (2014).

[91] Joerg F. Hipp et al. “Large-scale cortical correlation structure of spontaneous oscillatory activity”. In: *Nature Neuroscience* 15.6 (2012), pp. 884–890.

List of abbreviations

ASD: Autism spectrum disorder; EEG: Electroencephalography; SVM: Support vector machine; DSM: The Diagnostic and Statistical Manual of Mental Disorders; AQ: Autism quotient short questionnaire; SPQ: Sensory perception quotient short questionnaire; DFA: Detrended fluctuation analysis; fEI: Functional excitation/inhibition ratio; Coh: Coherence; Imcoh: Imaginary coherence; PLV: Phase locking value; wPLI: Weighted phase lag index; PEC: Power envelope correlations; nMAE: Normalized mean absolute error; FDR: False discovery rate.

Acknowledgements

This research was supported and funded by NWO ZonMW Top grant (2017/02015/ZONMW and 2017/02186/ZONMW).

Availability of data and materials

Data are available from SB and the code is available from QL on request.

Authors’ contributions

RFW, DJAS, KLH and SB conceived and conducted the experiments. QL, IK, TSA, and KLH designed the analysis. QL implemented the code and generated the first draft of the manuscript. RFW, HDM, DJAS, SB and KLH contributed to interpretation of the results. All authors reviewed the manuscript.

Competing interests

No conflicts of interest are declared by the authors.

Supplementary materials

Supplementary methods

Detailed information about how each individual EEG feature was computed can be found here.

Power

EEG power was computed using multitaper spectral estimation with 7 discrete prolate spheroidal sequences (DPSS) 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 (1.25–48 Hz).

Asymmetry

EEG power asymmetry was computed as the log transformed raw power (i.e., prior to absolute/relative transformation) in each left hemispheric patch subtracted from the corresponding patch in the right hemisphere (68):

$$\text{Asymmetry} = \ln(Power_{rh}) - \ln(Power_{lh}) \tag{1}$$

The asymmetry values were averaged across the brain lobes to yield frontal, parietal, temporal, occipital, cingulate, and insular asymmetry for each of the five frequency bands.

Theta/beta ratio

Theta/beta ratio was estimated by taking the averaged raw theta and beta power in the six brain lobes followed by computing their ratio and applying the natural logarithm.

Peak alpha frequency and 1/f exponent

Peak alpha frequency and 1/f exponent was estimated using the FOOOF algorithm (69). The FOOOF algorithm assume the power spectral density is composed of an the 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-squared of the fit of the full model to evaluate how well the algorithm worked. Based on visual inspection of the fits and the R-squared values, we decided that for our data the peak alpha frequency estimations of models with an R-squared > 0.90 was reliable. For the 1/f exponents, we set the thresholds to be R-squared > 0.95. If the R-squared 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 the power spectra, the R-squared 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 lower frequency and if the R-squared was above the thresholds then we used the estimated PAF and 1/f exponents. The end range of the fit was set to the end of the gamma-frequency range (48 Hz). Peak alpha frequency and 1/f exponents were calculated for each patch and a global peak alpha frequency was also computed as the average over all patches.

Long-range temporal correlations

Long-range temporal correlations were estimated following the detrended fluctuation analysis (DFA) procedure described by 71. Window sizes were equally spaced on a logarithmic scale and specified to be between 5 and 30 s. Specifically, the lowest window size was 5.54 s and the highest window size was 28.94 s. The DFA exponent indicates that the signal exhibits long-range anti-correlations if DFA < 0.5, DFA exponent ~ 0.5 indicates it is indistinguishable from a random process (i.e., uncorrelated),

while a DFA exponent > 0.5 indicates the signal exhibits long-range positive correlations. DFA exponents for each of the five canonical frequency bands were computed. Nolds 0.5.2 (88) was used to implement DFA.

Functional excitation/inhibition ratio

Functional excitation/inhibition ratio (fEI) was estimated following 20, and as recommended we only estimated fEI if DFA > 0.6. Windows of 5 seconds with 50% overlap was used for the computation. Sub-critical networks have fEI < 1, critical networks have fEI ~ 1, while super-critical networks have fEI > 1. fEI for each of the five canonical frequency bands were computed.

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 was 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 (73).

$$\text{Coh} = \frac{|G_{xy}|}{\sqrt{G_{xx}G_{yy}}} \tag{2}$$

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. A coherence value close to 1 indicates strong synchronization, while a value close to 0 reflects no synchronization between the two signals. Coherence values were calculated for each epoch and then averaged.

Imaginary Coherence

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 (74).

$$\text{Imcoh} = \frac{Im(G_{xy})}{\sqrt{G_{xx}G_{yy}}} \tag{3}$$

The drawback of using imaginary coherence 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.

Phase Locking Value

The cross-spectral density contains information about both the amplitudes and the phase difference between the two signals of interest. When looking at 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 (75). The phase-locking value was originally defined as the inter-trial variability of the phase difference at time point t, which is appropriate when looking at event-related potentials (ERPs). However, for resting-state data, the variability of the phase difference cannot be calculated across trials, so here we calculate it over time.

$$\text{PLV} = \frac{1}{T}\left|\sum_{t=1}^{T} e^{i\Delta\phi_t}\right| \tag{4}$$

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. If the phase differences are consistent over time, PLV is equal to 1, while if the phase differences are randomly distributed over time, PLV will be close to 0. The continuous wavelet transform

was used to decompose into frequencies. After estimation of PLV in each epoch the values were averaged to obtain a more robust estimation.

Weighted Phase Lag Index

A common source due to volume conduction might lead to consistent zero-phase differences and would thus inflate PLV. Thus, the phase lag index was developed to circumvent this problem, by calculating how consistent one signal lead/lag 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 (76).

$$\text{PLI} = \frac{1}{T}\left|\sum_{t=1}^{T} sign[\sin(\Delta\phi_t)]\right| \tag{5}$$

Similar to the idea behind imaginary coherence, the weighted phase lag index (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 (89; 90).

$$\begin{aligned} \text{wPLI} &= \frac{\frac{1}{T}\left|\sum_{t=1}^{T} |\sin(\Delta\phi_t)| \cdot sign[\sin(\Delta\phi_t)]\right|}{\frac{1}{T}\sum_{t=1}^{T} |\sin(\Delta\phi_t)|} \\ \Leftrightarrow \text{wPLI} &= \frac{\left|\sum_{t=1}^{T} |\sin(\Delta\phi_t)| \cdot sign[\sin(\Delta\phi_t)]\right|}{\sum_{t=1}^{T} |\sin(\Delta\phi_t)|} \end{aligned} \tag{6}$$

The denominator normalizes the wPLI to the interval $0 \leq \text{wPLI} \leq 1$, where higher values indicates more synchronization.

Power Envelope Correlations

Power envelope correlations (PEC) were estimated following 77. Briefly, the time series for each signal pair were bandpass 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 (91). Following orthogonalization, the power envelope of the orthogonalized analytical signals are estimated and log-transformed. Finally, Pearson's correlations were calculated and Fischer's r-to-z transform applied.

Supplementary figures

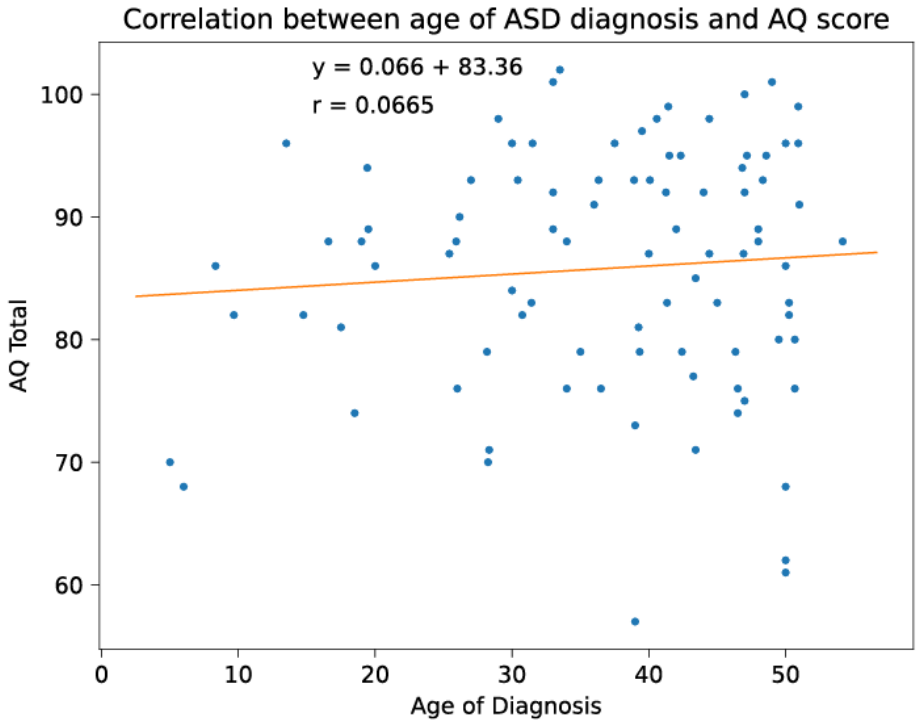

Fig. S1. **Lack of correlation between age of diagnosis and AQ scores.** The AQ scores were not correlated with the age of ASD diagnosis.

Supplementary tables

Supplementary Table S1 contains the classifier performances for all combinations of machine learning models and feature types. Please see the corresponding Excel table.

Supplementary Table S2 contains the regression performances for all combinations of machine learning models and feature types. Please see the corresponding Excel table.

Supplementary Table S1: Classification performance for all combinations of models and features

Feature	Model	Training Accuracy	Validation Accuracy	Test Accuracy
All Features	Logistic Regression with L1 norm	1.000	0.767	0.500
	Logistic Regression with L2 norm	0.862	0.844	0.483
	Linear SVM	0.943	0.801	0.473
	Random Forest	0.945	0.694	0.494
Power	Logistic Regression with L1 norm	0.723	0.636	0.468
	Logistic Regression with L2 norm	0.731	0.718	0.451
	Linear SVM	0.729	0.670	0.476
	Random Forest	0.772	0.632	0.482
Theta/Beta Ratio	Logistic Regression with L1 norm	0.525	0.507	0.462
	Logistic Regression with L2 norm	0.551	0.546	0.459
	Linear SVM	0.528	0.518	0.454
	Random Forest	0.696	0.548	0.497
Asymmetry	Logistic Regression with L1 norm	0.548	0.514	0.442
	Logistic Regression with L2 norm	0.612	0.600	0.432
	Linear SVM	0.597	0.558	0.444
	Random Forest	0.723	0.562	0.498
Peak Alpha Frequency	Logistic Regression with L1 norm	0.675	0.625	**0.558**
	Logistic Regression with L2 norm	0.708	0.692	0.502
	Linear SVM	0.683	0.646	**0.541**
	Random Forest	0.761	0.613	**0.532**
1/f Exponent	Logistic Regression with L1 norm	0.579	0.539	0.479
	Logistic Regression with L2 norm	0.639	0.636	0.453
	Linear SVM	0.602	0.583	0.481
	Random Forest	0.669	0.565	0.482
DFA Exponent	Logistic Regression with L1 norm	0.727	0.654	0.510
	Logistic Regression with L2 norm	0.750	0.739	0.529
	Linear SVM	0.741	0.687	0.510
	Random Forest	0.766	0.648	**0.542**
fEI	Logistic Regression with L1 norm	0.717	0.663	**0.537**
	Logistic Regression with L2 norm	0.760	0.756	0.518
	Linear SVM	0.735	0.701	**0.538**
	Random Forest	0.794	0.659	0.525
Coherence	Logistic Regression with L1 norm	0.721	0.650	**0.533**
	Logistic Regression with L2 norm	0.738	0.724	0.509
	Linear SVM	0.730	0.680	**0.537**
	Random Forest	0.765	0.640	0.503
Imaginary Coherence	Logistic Regression with L1 norm	0.742	0.665	0.497
	Logistic Regression with L2 norm	0.765	0.748	0.501
	Linear SVM	0.754	0.699	0.500
	Random Forest	0.786	0.650	**0.526**
Phase Locking Value	Logistic Regression with L1 norm	0.724	0.656	**0.551**
	Logistic Regression with L2 norm	0.743	0.730	**0.543**
	Linear SVM	0.739	0.681	**0.539**
	Random Forest	0.765	0.627	0.487
Weighted Phase Lag Index	Logistic Regression with L1 norm	0.681	0.604	0.443
	Logistic Regression with L2 norm	0.707	0.695	0.435
	Linear SVM	0.697	0.635	0.441
	Random Forest	0.753	0.617	0.500
Power Envelope Correlation	Logistic Regression with L1 norm	0.657	0.596	0.453
	Logistic Regression with L2 norm	0.690	0.687	0.452
	Linear SVM	0.659	0.624	0.453
	Random Forest	0.735	0.620	0.483

Accuracy are mean balanced accuracy over 10 repetitions of 10-by-10 two-layer crossvalidation runs (100 outer fold runs). Numbers in **bold** indicate the performance was significantly better than chance-level (One-sided Wilcox test with FDR correction for the 12 feature types for a given model type)

Supplementary Table S2: Regression performance for all combinations of models and features

Target Variable	Feature	Model	Training Error	Test Error
Autism Quotient	All Features	Linear Regression with L1 norm	0.444	1.051
		Linear Regression with L2 norm	0.315	1.243
	Power	Linear Regression with L1 norm	0.833	1.097
		Linear Regression with L2 norm	0.768	1.182
	Theta/Beta Ratio	Linear Regression with L1 norm	0.966	**0.993**
		Linear Regression with L2 norm	0.957	**0.979**
	Asymmetry	Linear Regression with L1 norm	0.998	1.003
		Linear Regression with L2 norm	0.948	1.079
	Peak Alpha Frequency	Linear Regression with L1 norm	0.918	1.003
		Linear Regression with L2 norm	0.860	1.023
	1/f Exponent	Linear Regression with L1 norm	0.865	1.048
		Linear Regression with L2 norm	0.822	1.064
	DFA Exponent	Linear Regression with L1 norm	0.826	1.009
		Linear Regression with L2 norm	0.786	1.050
	fEI	Linear Regression with L1 norm	0.731	**0.960**
		Linear Regression with L2 norm	0.683	0.988
	Coherence	Linear Regression with L1 norm	0.800	1.027
		Linear Regression with L2 norm	0.755	1.042
	Imaginary Coherence	Linear Regression with L1 norm	0.740	1.055
		Linear Regression with L2 norm	0.694	1.042
	Phase Locking Value	Linear Regression with L1 norm	0.791	1.025
		Linear Regression with L2 norm	0.741	1.018
	Weighted Phase Lag Index	Linear Regression with L1 norm	0.878	1.045
		Linear Regression with L2 norm	0.821	1.072
	Power Envelope Correlation	Linear Regression with L1 norm	0.876	1.082
		Linear Regression with L2 norm	0.815	1.123
Sensory Perception Quotient (Vision)	All Features	Linear Regression with L1 norm	0.545	1.125
		Linear Regression with L2 norm	0.396	1.282
	Power	Linear Regression with L1 norm	0.890	1.116
		Linear Regression with L2 norm	0.799	1.104
	Theta/Beta Ratio	Linear Regression with L1 norm	0.999	1.005
		Linear Regression with L2 norm	0.990	1.020
	Asymmetry	Linear Regression with L1 norm	1.000	1.000
		Linear Regression with L2 norm	0.983	1.036
	Peak Alpha Frequency	Linear Regression with L1 norm	0.951	1.028
		Linear Regression with L2 norm	0.869	1.050
	1/f Exponent	Linear Regression with L1 norm	0.970	1.022
		Linear Regression with L2 norm	0.908	1.067
	DFA Exponent	Linear Regression with L1 norm	0.890	1.116
		Linear Regression with L2 norm	0.837	1.167
	fEI	Linear Regression with L1 norm	0.832	1.158
		Linear Regression with L2 norm	0.798	1.202
	Coherence	Linear Regression with L1 norm	0.811	0.995
		Linear Regression with L2 norm	0.778	0.984
	Imaginary Coherence	Linear Regression with L1 norm	0.859	1.028
		Linear Regression with L2 norm	0.842	1.054

(Continued on next page)

	Phase Locking Value	Linear Regression with L1 norm	0.819	1.011
		Linear Regression with L2 norm	0.785	1.004
	Weighted Phase Lag Index	Linear Regression with L1 norm	0.901	1.009
		Linear Regression with L2 norm	0.891	1.005
	Power Envelope Correlation	Linear Regression with L1 norm	0.915	1.053
		Linear Regression with L2 norm	0.888	1.063
Sensory Perception Quotient (Auditory)	All Features	Linear Regression with L1 norm	0.614	1.171
		Linear Regression with L2 norm	0.453	1.332
	Power	Linear Regression with L1 norm	0.906	1.105
		Linear Regression with L2 norm	0.857	1.154
	Theta/Beta Ratio	Linear Regression with L1 norm	1.000	1.000
		Linear Regression with L2 norm	0.999	1.012
	Asymmetry	Linear Regression with L1 norm	0.995	1.005
		Linear Regression with L2 norm	0.967	1.037
	Peak Alpha Frequency	Linear Regression with L1 norm	0.950	1.004
		Linear Regression with L2 norm	0.894	0.994
	1/f Exponent	Linear Regression with L1 norm	0.998	1.008
		Linear Regression with L2 norm	0.932	1.076
	DFA Exponent	Linear Regression with L1 norm	0.857	1.119
		Linear Regression with L2 norm	0.818	1.150
	fEI	Linear Regression with L1 norm	0.839	1.121
		Linear Regression with L2 norm	0.802	1.124
	Coherence	Linear Regression with L1 norm	0.874	1.021
		Linear Regression with L2 norm	0.832	1.030
	Imaginary Coherence	Linear Regression with L1 norm	0.903	1.115
		Linear Regression with L2 norm	0.868	1.157
	Phase Locking Value	Linear Regression with L1 norm	0.867	1.016
		Linear Regression with L2 norm	0.835	1.030
	Weighted Phase Lag Index	Linear Regression with L1 norm	0.905	1.048
		Linear Regression with L2 norm	0.885	1.053
	Power Envelope Correlation	Linear Regression with L1 norm	0.920	1.059
		Linear Regression with L2 norm	0.886	1.090

Errors are mean normalized absolute errors. Numbers in **bold** indicate significantly better performance than predicting the mean (One-sided Wilcox test with FDR correction for the 12 feature types for a given model