Classifying Obsessive-Compulsive Disorder from Resting-State EEG using Convolutional Neural Networks: A Pilot Study

Brian A. Zaboski,[1] Sarah Kathryn Fineberg,[1] Patrick D. Skosnik,[1,2,*]

Stephen Kichuk,[1] Madison Fitzpatrick,[1] & Christopher Pittenger[1,3-7]

1. Department of Psychiatry, Yale University School of Medicine, New Haven, CT
2. Bouvé College of Health Sciences, Northeastern University, Boston, MA
3. Department of Psychology, Yale University, New Haven, CT
4. Department of Neuroscience, Yale University School of Medicine, New Haven, CT
5. Yale Child Study Center, Yale University School of Medicine, New Haven, CT
6. Center for Brain and Mind Health, Yale University School of Medicine, New Haven, CT
7. Wu Tsai Institute, Yale University, New Haven, CT

* Patrick Skosnik passed away in February, 2024

Author note:

Brian A. Zaboski: https://orcid.org/0000-0003-4567-8694

This work is funded in part by a grant provided by the Brain and Behavior Research Foundation (NARSAD, PI: BZ), by NIH grant R01MH116038 (CP), and by the State of Connecticut through its support of the Ribicoff Research Facilities at the Connecticut Mental

Health Center. This work reflects the views of the authors and not those of the State of Connecticut.

In the past three years, BZ has consulted with Biohaven Pharmaceuticals and received royalties from Oxford University Press; these relationships are not related to the work described here. SKF has consulted for Boehringer Ingelheim International GMBH, atai Life Sciences, and Oryzon Genomics in the last three years. In the past three years, CP has consulted for Biohaven Pharmaceuticals, Freedom Biosciences, Transcend Therapeutics, UCB BioPharma, Mind Therapeutics, Ceruvia Biosciences, F-Prime Capital Partners, and Madison Avenue Partners; has received research support from Biohaven Pharmaceuticals, Freedom Biosciences, and Transcend Therapeutics; owns equity in Alco Therapeutics, Mind Therapeutics, and Lucid/Care; receives royalties from Oxford University Press and UpToDate; and holds patents on pathogenic antibodies in pediatric OCD and on novel mechanisms of psychedelic drugs. None of these relationships are related to the current paper.

Correspondence concerning this article should be addressed to Brian. A. Zaboski, Yale School of Medicine, Department of Psychiatry, Yale University, Temple Medical Center (40 Temple)| 60 Temple St | New Haven, CT 06510, U.S.A. Email: brian.zaboski@yale.edu

Abstract

Objective: Classifying obsessive-compulsive disorder (OCD) using brain data remains challenging. Resting-state electroencephalography (EEG) offers an affordable and noninvasive approach, but traditional machine learning methods have limited its predictive capability. We explored whether convolutional neural networks (CNNs) applied to minimally processed EEG time-frequency representations could offer a solution, effectively distinguishing individuals with OCD from healthy controls. **Method:** We collected resting-state EEG data from 20 unmedicated participants (10 OCD, 10 healthy controls). Clean, 4-second EEG segments were transformed into time-frequency representations using Morlet wavelets. In a two-step evaluation, we first used a 2D CNN classifier using leave-one-subject-out cross-validation and compared it to a traditional support vector machine (SVM) trained on spectral band power features. Second, using multimodal fusion, we examined whether adding clinical and demographic information improved classification. **Results:** The CNN achieved strong classification accuracy (82.0%, AUC: 0.86), significantly outperforming the chance-level SVM baseline (49.0%, AUC: 0.45). Most clinical variables did not improve performance beyond the EEG data alone (subject-level accuracy: 80.0%). However, incorporating education level boosted performance notably (accuracy: 85.0%, AUC: 0.89). **Conclusion:** CNNs applied to resting-state EEG show promise for diagnosing OCD, outperforming traditional machine learning methods. Despite sample size limitations, these findings highlight deep learning's potential in psychiatric applications. Education level emerged as a potentially complementary feature, warranting further investigation in larger, more diverse samples.

Keywords: Obsessive-Compulsive Disorder, Electroencephalography, Deep Learning, Convolutional Neural Networks, Precision Psychiatry

Classifying Obsessive-Compulsive Disorder from Resting-State EEG using Convolutional Neural Networks: A Pilot Study

Obsessive-Compulsive Disorder (OCD) is a prevalent and often debilitating neuropsychiatric condition, affecting 1-3% of the population worldwide and causing significant functional impairment and distress (Kessler et al., 2012; Ruscio et al., 2010). While effective treatments exist, including cognitive-behavioral therapy (CBT) and pharmacotherapy with selective serotonin reuptake inhibitors (SSRIs), a substantial portion of patients do not achieve remission (Abramowitz & Arch, 2014; Law & Boisseau, 2019). Diagnosis currently relies on clinical interviews and standardized symptom scales, such as the Yale-Brown Obsessive-Compulsive Scale, which capture subjective reports of obsessions and compulsions (Goodman et al., 1989). The inherent subjectivity in this process, alongside the complexities in differentiating certain symptom presentations (Mattera et al., 2024), highlights a critical need for new models to aid in diagnosis and predict treatment response (Zaboski et al., 2021).

Neuroimaging modalities like positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) have identified new treatment approaches (Rance et al., 2023) as well as functional and structural alterations in individuals with OCD, particularly within cortico-striatal-thalamo-cortical circuits (Pittenger, 2017; Stein et al., 2019). However, the high cost, limited accessibility, and relatively poor temporal resolution of these techniques hinder their widespread clinical application (Constable, 2023). Electroencephalography (EEG), by contrast, offers a non-invasive, relatively inexpensive, portable method for measuring the brain's electrical activity, with millisecond temporal resolution (Zaboski et al., 2021). Despite these advantages, the use of EEG to investigate OCD has been comparatively sparse, with studies yielding heterogeneous findings regarding event-related potentials (event-related potentials

[ERPs] like the P300 and the error-related negativity [ERN]), oscillatory power abnormalities (e.g., in theta, alpha, and beta bands), and measures of signal complexity (Metin et al., 2019; Zaboski et al., 2021).

The complexity and high dimensionality of EEG data present challenges for traditional analytic methods. Machine learning (ML) approaches have emerged as powerful tools for identifying subtle patterns within complex biological data. Previous studies in OCD have employed ML techniques, often using support vector machines (SVMs) or conventional artificial neural networks (ANNs). A common characteristic of these approaches was their application to pre-defined quantitative EEG (qEEG) features—like spectral band power or cordance—that were engineered from the EEG signals prior to model training. For example, Metin et al., (2019) found that theta band power derived from qEEG (analyzed with ANNs) could predict TMS response. Similarly, Erguzel et al. (2015) used SVMs on qEEG cordance to classify OCD versus trichotillomania. While these methods have shown utility, their dependence on manual feature engineering is both cumbersome and may not fully capture the intricate patterns embedded in EEG data, a limitation that modern deep learning seeks to overcome.

More recently convolutional neural networks (CNNs) have shown remarkable success in automatically extracting informative hierarchical features from raw or minimally processed data, bypassing the need for manual feature engineering (Farhad et al., 2024; Xu et al., 2018). CNNs are well suited for analyzing grid-like data structures, such as the time-frequency representations of EEG signals or even the raw multichannel time-series data itself (Farhad et al., 2024; Xu et al., 2018). Consequently, their application in EEG is growing rapidly for various neurological and psychiatric conditions (Farhad et al., 2024). However, few studies—and none in OCD—have

applied deep learning to automatically learn discriminative patterns directly from minimally processed, high-dimensional resting-state EEG data.

To begin to address this gap, we investigated the feasibility of using a CNN to capture the richness of resting state EEG's temporal signals to classify individuals diagnosed with OCD from healthy controls. By applying CNNs to resting state signals, we hope to bypass manual feature engineering (such as calculating power in specific bands) while capturing complex spatio-temporal and spectral dynamics missed by simpler models or feature sets. We tested whether a CNN can outperform a more conventional machine learning approach to classification. We then systematically added clinical and demographic predictors to the CNN model to determine if its predictive metrics would improve. Success in this endeavor could help us move towards the development of more objective, scalable, and accessible tools for OCD research and clinical practice, aligning with the broader goal of achieving precision medicine in psychiatry (Zaboski & Bednarek, 2025).

Methods

Participants

Participants were recruited from the local community for a larger 18-week clinical trial investigating brain network changes and treatment prediction in response to pharmacotherapy for OCD ("Brain network changes accompanying and predicting treatment responses to pharmacotherapy in OCD", Yale HIC#: 2000023688; PI: C. Pittenger). Recruitment methods included social media advertisements, local bus ads, community flyers, and referrals to the Yale OCD Research Clinic. The study protocol was approved by the Yale University Institutional Review Board (IRB), and all participants provided written informed consent prior to participation, including specific consent for the EEG procedures described herein. Participants

received $50 for completing the baseline clinical/self-report assessment battery and an additional $80 for the baseline EEG session.

Participants underwent an initial screening followed by a comprehensive intake evaluation (in-person or remote). The primary diagnosis of OCD was established using either the Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998) or Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders (DIAMOND; (Tolin et al., 2018) and validated by a board-certified psychologist or psychiatrist associated with the study. The same diagnostic process was used for healthy control (HC) participants. HCs were excluded if they had any current DSM-5 diagnosis, including symptoms of psychosis, mania, depression, or substance use. All participants were unmedicated.

Medical screening was performed for OCD participants as part of the parent study and included assessment of standard laboratory blood tests (including complete blood count, metabolic panel, thyroid tests), urine tests (for kidney function and drug screen), electrocardiogram (ECG), and a physical examination to ensure general health. Key exclusion criteria relevant to the larger study included pregnancy or breastfeeding, magnetic resonance imaging (MRI) contraindications (e.g., certain metallic implants), current use of specific interacting medications (e.g., monoamine oxidase inhibitors, certain serotonergic agents, warfarin unless specifically approved), inability to tolerate fluoxetine, and significant suicidal risk as determined by the investigators. Participants agreed not to start new medications during the study period.

Baseline EEG data were analyzed for the present study. Participants completed self-report scales assessing mood, OCD symptoms, personality characteristics, and other relevant factors using REDCap, a secure, web-based, HIPAA-compliant data-capture system (P. A. Harris

et al., 2009, 2019) and a baseline EEG session. 22 participants (11 OCD, 11 HC) completed the baseline EEG session. Initial preprocessing yielded unusable data for two participants (one HC, one OCD). Despite artifact correction attempts including ICA two individuals exhibited high-amplitude noise surpassing the 250 μV rejection criterion, necessitating their exclusion from the final analysis. $N = 20$ participants ($n = 10$ OCD, $n = 10$ HC) were retained for the final analysis.

Data Acquisition

Resting-state EEG data were acquired within the Psychophysiology Laboratory in the Connecticut Mental Health Center. The laboratory is specifically designed for acquiring high-quality human psychophysiology data and features an electromagnetically and acoustically shielded testing booth (8 x 6 x 9 ft) to minimize environmental noise and electrical interference during recording. Within these booths, participants were seated comfortably in a reclining chair.

The primary EEG recording system utilized was a 64-channel Biosemi ActiveTwo system. Data were collected using a standard 64-channel electrode cap, arranged according to the international 10-20 system. The recordings used an average reference; additional bipolar electrode pairs were employed to record vertical and horizontal electrooculogram (EOG) signals to monitor eye movements and blinks for subsequent artifact processing. Throughout the recording sessions, electrode impedances were maintained below 10 kΩ to ensure good signal quality. The EEG data were continuously sampled at a rate of 1024 Hz and acquired with an online hardware band-pass filter set at 0.1–100 Hz. Recordings were made under eyes open conditions for 4 minutes at a sampling rate of and saved in .bdf format.

Data Preprocessing

Offline preprocessing of the raw EEG data was conducted using the MNE-Python library (v1.7) (Gramfort, 2013; Gramfort et al., 2014) to remove artifacts and prepare the signals for

analysis. Initially, raw data files were loaded, and any non-EEG or status channels were removed. The standard 10-20 montage configuration was applied. To mitigate noise, a 60 Hz notch filter was applied to remove power line interference, followed by a band-pass filter between 0.5 Hz and 50.0 Hz (FIR filter) to eliminate slow drifts and high-frequency noise. Each participant's recording was then cropped to a uniform duration of 240 seconds to standardize the data length across the sample.

Artifacts primarily related to eye movements and blinks were addressed using independent component analysis (ICA). The FastICA algorithm (Aapo, 1999; Ablin et al., 2018; Bell & Sejnowski, 1995; Lee et al., 1999) was fitted to the filtered data to derive 15 independent components. Components exhibiting high correlation with simultaneously recorded EOG channels were automatically identified and removed. The EEG signal was then reconstructed from the remaining non-artifactual components. Reconstructed EEG data were re-referenced to the common average across all channels (Jiang et al., 2024; Tamburro et al., 2021). Finally, the cleaned, continuous data were segmented into 4-second epochs with a 2-second overlap between consecutive epochs. This yielded 190 epochs of data per subject, partially mitigating the limitations created by the small sample size in this pilot study. All generated epochs were carried forward. Metadata indicating participant ID and diagnostic group (HC/OCD) were associated with each epoch before concatenating all epochs across participants for subsequent feature extraction.

Time-Frequency Feature Generation

To capture dynamic neural activity patterns suitable for the CNN, the preprocessed EEG epochs were transformed into time-frequency representations. This was achieved using a Morlet wavelet transform implemented in MNE-Python, applied independently to each EEG channel

within each epoch. Time-frequency representations were computed for each 4-second epoch using MNE-Python's tfr_morlet function. We analyzed 40 logarithmically spaced frequencies ranging from 1 Hz to 45 Hz. The number of cycles for each wavelet was set proportionally to the frequency (n_cycles = frequency / 2.0). To manage computational load and data size, the resulting TFR output was decimated by a factor of 16. Given the original 1024 Hz sampling rate resulted in 4096 samples per 4-second epoch, this decimation yielded a final temporal dimension of 256 time points per epoch for CNN input. The resulting absolute power values were then log-transformed (base 10, scaled by 10, adding 1e-10 for stability) to approximate a normal distribution and reduce the influence of extreme values. The complete preprocessing yielded 2,277 epochs suitable for analysis. The subsequent time-frequency transformation resulted in one multi-channel power 'image' per epoch, totaling 2,277 images. Each image had the dimensions 40 frequencies, 256 time points, 68 channels. These images served as the input features for CNN classification.

Primary Classification Procedure

Convolutional Neural Network

We employed a CNN to classify individual EEG epochs as belonging to either the HC or OCD group. To rigorously control for individual subject effects and to prevent the model from gaining access to data outside the training set, we applied leave-one-subject-out (LOSO) cross-validation (Mevlevioğlu et al., 2024; Ren et al., 2024). Across the 20 folds of this procedure, the model was trained on data from 19 subjects and evaluated on all epochs from the held-out subject. Data scaling (StandardScaler) was applied within each fold, fitted only on the training data.

The classification model utilized a two-dimensional CNN implemented using the Keras API (Chollet & others, 2015) within TensorFlow (Martín Abadi et al., 2015). The architecture consisted of two sequential convolutional blocks: The first block employed 16 filters with a (3, 5) kernel size, followed by ReLU activation, batch normalization, (2, 2) max-pooling, and dropout (rate = 0.4); the second block used 32 filters with a (3, 5) kernel size and identical subsequent layers (ReLU, batch normalization, max-pooling, dropout). Following the convolutional blocks, the feature maps were flattened and passed directly to a single dense output neuron with a sigmoid activation function for binary classification (HC vs. OCD). An architecture with integrated regularization techniques (batch normalization, dropout) was chosen to mitigate the risk of overfitting, given the study's sample size. The computed log-power time-frequency representations for each epoch were reshaped to conform to the 'channels-last' input format expected by the 2D CNN. Specifically, the dimensions were permuted from (epochs, channels, frequencies, time points) to (epochs, frequencies, time points, channels). The resulting input tensor for each epoch fed into the CNN thus had dimensions: 40 (Frequencies) × 256 (Time Points) × 68 (EEG Channels). The output of the CNN's sigmoid layer was a probability (range: 0 to 1), representing the model's confidence that an epoch belonged to the OCD group. The final binary prediction (0, 1) was obtained by thresholding the probability at 0.5.

Support Vector Machine Baseline Comparison

To contextualize the performance of the CNN model, a baseline comparison was conducted using a standard support vector machine (SVM) applied to traditional EEG features (Bera et al., 2025; Rahul et al., 2024) also predicting group epochs. The SVM classifier with a radial basis function (RBF) kernel (C=1.0, gamma='scale', class_weight='balanced') was trained on features representing the mean spectral power within canonical frequency bands (delta: 1 – 4

Hz, theta: 4 – 8 Hz, alpha: 8 – 13 Hz, beta: 13 – 30 Hz, gamma: 30 – 45 Hz) calculated for each channel using Welch's method. To ensure direct comparability, an identical LOSO cross-validation procedure, including per-fold data scaling, was employed.

Input Visualization

To aid intuition for the CNN, Figure 1 contains the time-frequency power representation for channel Cz plotted for a HC participant and an OCD participant averaged across all epochs. Visual inspection revealed distinct patterns between the groups. The HC participant exhibited prominent, sustained spectral power predominantly within the high Beta frequency band (~20 - 30 Hz). In contrast, the OCD participant displayed lower overall power and notably lacked this sustained high-Beta activity band, instead showing more diffuse and attenuated power across the frequency spectrum, particularly compared to the prominent Beta activity seen in the HC individual.

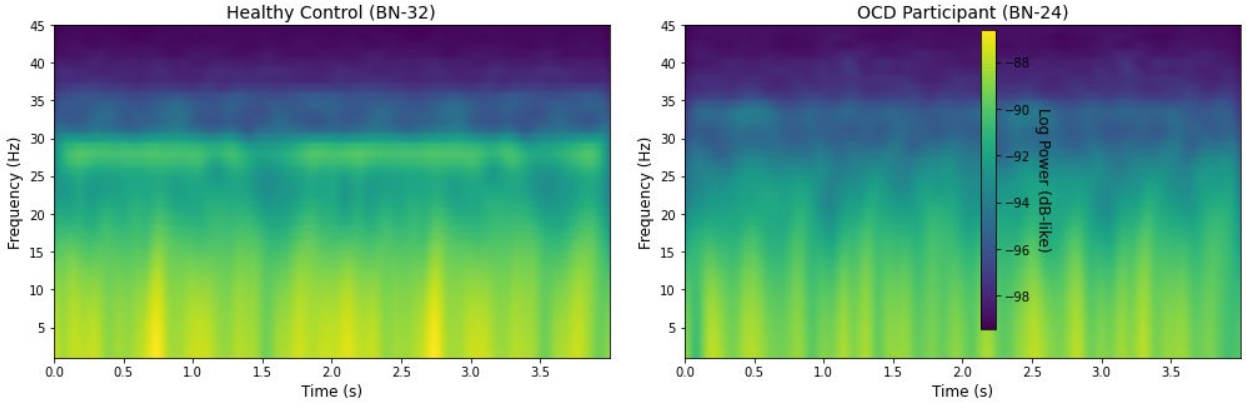

Figure 1: Example Average Time-Frequency Power (Cz) for Two Subjects

Figure 1, while illustrating differences in the averaged Beta activity at Cz for these selected individuals, highlights the need for the CNN approach for prediction. It is crucial to recognize that the CNN did not operate on these averaged, single-channel representations across all epochs (Schirrmeister et al., 2017). Instead, it processed a single, multi-channel entity. This is

in contrast to more common machine learning approaches (Chaddad et al., 2023; Hosseini et al., 2020; Lotte et al., 2007; Salehzadeh et al., 2023).

The structure of a CNN's input data is important. Consider color as an analogue: A standard color image is not three separate images (one for Red, one for Green, one for Blue). It is a single image with three channels stacked together at each pixel location. The data structure is typically [height, width, 3 color channels] (Krohn et al., 2019). In the current study, after the wavelet transform and reshaping, the data structure for a single epoch fed into the CNN was [40 Frequencies (Height), 256 Time Points (Width), 68 EEG Channels (Depth)]. By analyzing this single multi-channel entity, the network learned patterns based not only on the time-frequency content within a single channel but also on the spatial relationships between the 68 different channels across the scalp (Zhang et al., 2023).

Multimodal Classification Procedure

To explore whether readily available clinical and demographic information could provide diagnostic value beyond the EEG data, we conducted a second-stage analysis investigating the addition of clinical and demographic information to the CNN-derived EEG score using a late-fusion (stacking) approach (Breiman, 1996; Wolpert, 1992). In this stage, we predicted subject status. Output of the second-stage represented a probability, indicating confidence in the subject's classification. Similar to Stage One, the final binary prediction (0 or 1) was derived by thresholding this probability at 0.5.

This analysis began by computing subject-level EEG scores, obtained by averaging the predictions on the held-out subject from each fold of the primary CNN's LOSO evaluation. Next, subject-level EEG scores were merged with clinical scores (e.g., DOCS, BAI, and BDI-II totals) and demographic data (age, sex, education), matched by participant ID. To systematically assess

the contribution of individual variables while minimizing overfitting, we tested the addition of each clinical or demographic feature one at a time to the EEG score (Baltrušaitis et al., 2018).

For each feature combination tested, a Scikit-learn (Pedregosa et al., 2011) pipeline was created to preprocess the two input features within the cross-validation loop: the numeric EEG score (and any other numeric variable being tested, like age) was standardized using StandardScaler. Any categorical variable being tested (e.g., sex, education) was one-hot encoded. A logistic regression classifier was selected as the second-stage model, and each combined model (feature pair + preprocessing + logistic regression) was evaluated using the identical LOSO cross-validation procedure employed for the original CNN. This ensured equitable comparison across feature combinations.

Lastly, in addition to using subject-level scores from the CNN, we also generated scores from the support vector machine. We then tested each variable as described above to test if the SVM prediction would improve. If the SVM led to significantly greater performance, this would imply that there is information in the clinical and demographic features that is already captured in the EEG signal as analyzed using the CNN.

Statistical Analysis and Software

The primary performance metrics for evaluating both the EEG-only CNN and the combined multimodal model were the area under the receiver operating characteristic curve (AUC) and overall classification accuracy (Naidu et al., 2023). Performance was assessed both per fold and globally by aggregating predictions across all LOSO folds. Overall confusion matrices and classification reports detailing precision, recall, and F1-scores were generated. Average performance across folds is reported as mean ± standard deviation. The analysis relied on Python (Van Rossum & Drake, 2009) and key scientific libraries including MNE-Python

(v1.7) (Gramfort, 2013; Gramfort et al., 2014), NumPy (C. R. Harris et al., 2020), Pandas (McKinney, 2010), Scikit-learn (Pedregosa et al., 2011), TensorFlow (Martín Abadi et al., 2015) with Keras (Chollet & others, 2015), Matplotlib (Hunter, 2007), and Seaborn (Waskom, 2021).

Clinical Measures

Dimensional Obsessive-Compulsive Scale

The Dimensional Obsessive-Compulsive Scale (DOCS) is a 20-item self-report tool created to assess the severity of Obsessive-Compulsive Disorder (OCD) symptoms across four well-established dimensions: contamination concerns paired with cleaning compulsions, fears of responsibility for harm leading to checking behaviors, unacceptable thoughts accompanied by mental rituals, and symmetry obsessions linked with ordering compulsions (Abramowitz et al., 2010). Initial psychometric evaluations indicated that the DOCS possesses strong internal consistency ($\alpha \approx 0.90$). Its capacity for detecting OCD symptoms (diagnostic sensitivity) is further supported by a strong positive correlation ($r = 0.69$) between DOCS total scores and scores on the Obsessive-Compulsive Inventory-Revised (Abramowitz et al., 2010; Abramowitz & Deacon, 2006).

Magical Ideation Scale

The Magical Ideation Scale (MIS) is a 30-item self-report instrument developed to measure unlikely causal beliefs, such as telepathy, superstitious thinking, and thought broadcasting (Eckblad & Chapman, 1983). It employs a true/false response format where respondents endorse statements reflecting personal beliefs and experiences related to these phenomena (e.g., "I have sometimes felt that strangers were reading my mind") (Eckblad & Chapman, 1983). The total score, derived by summing the items keyed towards magical ideation, serves as an indicator of this cognitive style, which is considered a feature of schizotypy or

psychosis-proneness (Eckblad & Chapman, 1983; Kwapil et al., 1997). Psychometric studies have generally shown good internal consistency, with reliability coefficients often reported between $r = .76$ and $r = .93$ across various samples, including adolescents and adults (Fonseca Pedrero et al., 2009; Kingdon et al., 2012). The MIS demonstrates construct validity through correlations with other measures of schizotypy, particularly positive symptoms like perceptual aberrations, and its ability to predict psychotic-like experiences or distinguish individuals at higher risk for psychosis-spectrum disorders in longitudinal studies (Chapman et al., 1982; Fonseca Pedrero et al., 2009; Kwapil et al., 1997; Martin et al., 2011).

Beck Anxiety Inventory

The Beck Anxiety Inventory (BAI) is a 21-item self-report questionnaire used to evaluate the severity of anxiety symptoms in adults. Total scores range from 0 to 63. Respondents rate how much they have been bothered by common anxiety symptoms over the past week using a 4-point Likert scale (0 = "Not at all" to 3 = "Severely") (Beck et al., 1988, 2012). The BAI focuses particularly on the somatic or physical symptoms of anxiety, such as sweating, trembling, and dizziness. The BAI has demonstrated strong internal consistency ($\alpha = .91$) and test-retest reliability ($r = .65$) in meta-analyses of its psychometric properties (Bardhoshi et al., 2016).

Beck Depression Inventory-II

The Beck Depression Inventory-II (BDI-II) is a self-report questionnaire consisting of 21 items designed to gauge the intensity of cognitive and physical symptoms commonly associated with depression (Beck et al., 1961). Respondents typically choose from four statements per item, each reflecting a different level of symptom severity and carrying a specific point value. The sum of these points yields a total score, falling between 0 and 63. This total score is used to classify depression severity into categories: minimal (0 – 13), mild (14 – 19),

moderate (20– 28), or severe (29 – 63). Studies support the BDI-II's psychometric properties, demonstrating high internal consistency ($\alpha \approx 0.90$), adequate test-retest reliability ($\alpha \approx 0.73 – 0.96$), and effectiveness in distinguishing between psychiatric patients diagnosed with depression and those without (Beck et al., 1996; Wang & Gorenstein, 2013).

Results

Descriptive Statistics

Descriptive statistics on our 20 participants (10 HC and 10 OCD) are presented in Table 1. The average participant age was in the mid-thirties ($M = 35.4$, $SD = 13.5$), with similar distributions between the HC and OCD groups. Regarding clinical scores, mean symptom severity varied across measures: DOCS and BDI-II scores were higher in the OCD group, while MIS scores were slightly lower. BAI scores were comparable between groups.

The sample consisted predominantly of males (65.0%). A high level of educational attainment characterized the cohort, with 75% of participants holding doctoral degrees or other/professional qualifications; notable differences in the distribution of specific education levels were observed between the HC and OCD groups. Data completeness was high, with missing values limited to age for one participant and MIS scores for two participants.

Table 1. *Descriptive Statistics for Demographics and Clinical Characteristics*

Continuous Variable	**Overall (*M, SD*)**	**OCD (*M, SD*)**	**HC (*M, SD*)**
Age, years [1]	35.42 (13.46)	34.89 (14.41)	35.90 (13.31)
DOCS Total	17.00 (16.98)	20.80 (17.99)	13.20 (15.91)

MIS Score [2]	7.00 (6.33)	5.90 (5.92)	8.38 (6.97)
BAI Total	13.55 (13.34)	13.30 (11.87)	13.80 (15.32)
BDI-II Total	11.50 (13.34)	16.00 (14.50)	7.00 (10.98)
Categorical Variables	***N* (%)**	***n* (%)**	***n* (%)**
Sex			
Male	13 (65.0%)	6 (60.0%)	7 (70.0%)
Female	7 (35.0%)	4 (40.0%)	3 (30.0%)
Education Level			
Some College	2 (10.0%)	2 (20.0%)	0 (0.0%)
Bachelor Degree	2 (10.0%)	2 (20.0%)	0 (0.0%)
Master's Degree	1 (5.0%)	0 (0.0%)	1 (10.0%)
Doctoral Degree	6 (30.0%)	4 (40.0%)	2 (20.0%)
Other/Professional	9 (45.0%)	2 (20.0%)	7 (70.0%)

Note: *N = 20* total; *n* = 10 participants in each group. HC = Healthy Control; OCD = Obsessive-Compulsive Disorder; DOCS: Dimensional Obsessive-Compulsive Scale; MIS: Magical Ideation Scale; BAI: Beck Anxiety Inventory; BDI-II: Beck Depression Inventory-II.

Primary Model

CNN-Based Classification

Aggregating the predictions across all 20 held-out subjects, the CNN's overall classification accuracy across all 2277 test epochs reached 81.95%. The primary metric for

discriminative performance, the ROC AUC, calculated from the combined predictions across all folds, was 0.86. Further analysis of the aggregated predictions revealed balanced performance, with an overall weighted average precision of 0.82, recall of 0.82, and F1-score of 0.82 (detailed metrics are presented in Table 1).

Table 1: *Overall Classification Performance of the CNN Model*

Class	Precision	Recall	F1-Score	Support
HC	0.84	0.79	0.81	1148
OCD	0.80	0.85	0.82	1129
Accuracy	-	-	0.82	2277
Macro Avg	0.82	0.82	0.82	2277
Wtd Avg	0.82	0.82	0.82	2277

Notes: F1-Score: A single metric that balances precision and recall; support: Number of epochs belonging to that true class in the aggregated test sets; accuracy: Proportion of correctly classified epochs; macro avg: The unweighted average of precision, recall, F1 across both classes; weighted avg: The average of the metric across both classes, weighted by the support (number of true instances) for each class.

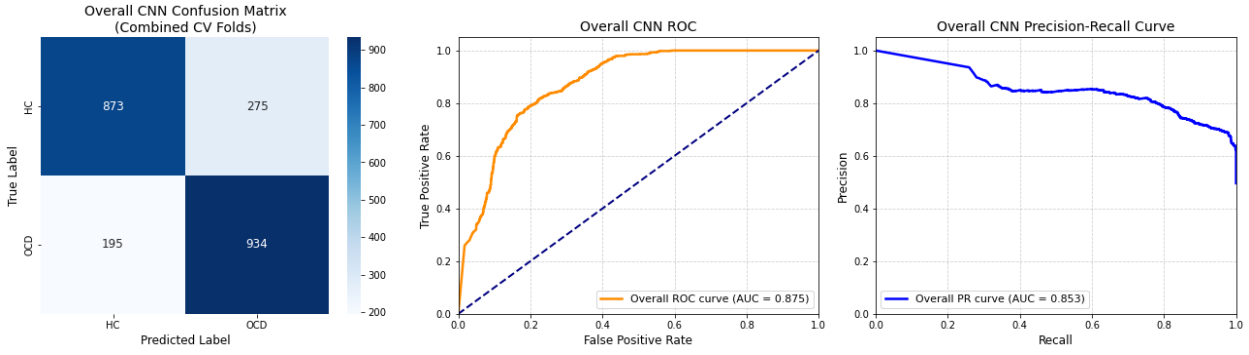

Figure 2: CNN Confusion Matrix, Receiver Operator Curve (ROC), and Precision-Recall Curve

The model's overall performance characteristics are illustrated in Figure 2. The confusion matrix details the specific classification outcomes across all test epochs: 873 True Negatives (HC correctly identified), 934 True Positives (OCD correctly identified), 195 False Negatives (OCD misclassified as HC), and 275 False Positives (HC misclassified as OCD). The overall ROC curve illustrates the model's strong ability to distinguish between classes across different thresholds, consistent with the high overall AUC. The overall Precision-Recall curve yielded an area under the curve of 0.875, further indicating robust performance, particularly in correctly identifying positive (OCD) cases while maintaining reasonable precision.

Consistent with the nature of LOSO cross-validation on datasets with inherent inter-subject variability (Arevalillo-Herráez et al., 2019; Mennes et al., 2010), the classification accuracy for individual held-out subjects varied considerably across the folds (mean per-fold accuracy = 81.29% ± 32.17%). Metrics such as per-fold AUC were undefined due to the single-class nature of the test set in each iteration. Therefore, the aggregated performance metrics and overall curve analyses (Table 1, Figure 2) provide the most comprehensive assessment of the model's generalizability in distinguishing HC from OCD participants within this cohort. These results suggest that the time-frequency patterns derived from resting-state EEG, as processed by the CNN, contain significant information relevant to OCD diagnostic status.

Support Vector Machine Classification

For comparison, we constructed a SVM trained on spectral power features (delta: 1 – 4 Hz, theta: 4 – 8 Hz, alpha: 8 – 13 Hz, beta: 13 – 30 Hz, gamma: 30 – 45 Hz). Aggregating predictions across all 20 folds, the SVM model did not achieve prediction better than chance level. The overall accuracy was 49.01%, and the overall ROC AUC was 0.449 (Figure 2), indicating an inability to reliably discriminate between HC and OCD groups using these features;

indeed, the AUC suggests performance slightly worse than random guessing. The overall weighted average F1-score was 0.49. Table 2 contains the model metrics; the corresponding confusion matrix is shown in Figure 3.

Table 2. *Overall Classification Performance of the SVM Baseline Model*

Class	Precision	Recall	F1-Score	Support
HC (0)	0.49	0.52	0.51	1148
OCD (1)	0.48	0.46	0.47	1129
Accuracy	-	-	0.49	2277
Macro Avg	0.49	0.49	0.49	2277
Wtd Avg	0.49	0.49	0.49	2277

Notes: F1-Score: A single metric that balances precision and recall; support: Number of epochs belonging to that true class in the aggregated test sets; accuracy: Proportion of correctly classified epochs; macro avg: The unweighted average of precision, recall, F1 across both classes; weighted avg: The average of the metric across both classes, weighted by the support (number of true instances) for each class.

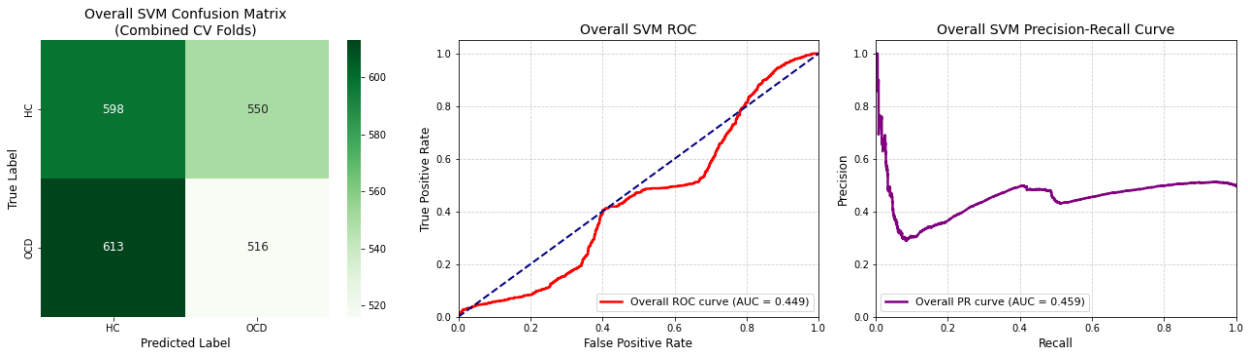

Figure 3: CNN Confusion Matrix, Receiver Operator Curve (ROC), and Precision-Recall Curve

Multimodal Clinical and Demographic Classification

CNN Augmentation

To determine if classification accuracy could be enhanced, we implemented a second-stage analysis using a late-fusion approach incorporating individual clinical and demographic variables. Performance using only the EEG score served as the subject-level baseline for comparison, yielding an AUC of 0.86 and an overall accuracy of 80.0%. The performance metrics for models incorporating clinical scores (DOCS, MIS, BAI, BDI-II) and demographic variables (age, sex, education) are detailed in Table 3.

Table 3: *Predictive Performance for CNN + Clinical and Demographic Features*

Feature	*N*	AUC	Accuracy	Precision	Recall	F1	Change
Baseline EEG	20	0.86	0.80	0.80	0.80	0.80	-
DOCS Total	20	0.86	0.80	0.80	0.80	0.80	None
MIS Score	18	0.81	0.78	0.75	0.75	0.75	Worse (All)
BAI Total	20	0.86	0.80	0.80	0.80	0.80	None
BDI-II Total	20	0.84	0.80	0.80	0.80	0.80	Worse (AUC)
Age	19	0.76	0.79	0.80	0.80	0.80	Worse (AUC)
Sex	20	0.81	0.80	0.80	0.80	0.80	Worse (AUC)
Education	20	0.89	0.85	0.82	0.90	0.86	Improved

Note: Baseline metrics include only the convolutional neural network's subject-level EEG scores in the second stage. *N* = number of subjects remaining after handling missing values. DOCS: Dimensional Obsessive-Compulsive Scale; MIS: Magical Ideation Scale; BAI: Beck Anxiety Inventory; BDI-II: Beck Depression Inventory-II.

The addition of most variables—symptom scores (DOCS, MIS), anxiety/depression scores (BAI, BDI-II), participant age, and sex—did not improve classification performance beyond the baseline EEG CNN model. Several of these combinations resulted in slightly lower AUC values (range 0.76 - 0.86) while maintaining similar overall accuracy (~80%).

By contrast, incorporating the participant's education level alongside the EEG score resulted in improved classification performance. This optimal multimodal model achieved an AUC of 0.89 and an overall accuracy of 85.0% across the LOSO cross-validation folds. Compared to baseline, this represents an improvement of 0.03 in AUC and a 5 percentage point increase in accuracy. Examining the class-specific performance for this model, it achieved a recall (sensitivity) of 0.90 and a precision of 0.82 (F1-score = 0.86) for correctly identifying participants with OCD, and a recall of 0.80 and precision of 0.89 for identifying HCs. These findings suggest that while the resting-state EEG features captured by the CNN contain substantial discriminative information, integrating subject education level provides unique complementary value. The confusion matrix is presented in Figure 4.

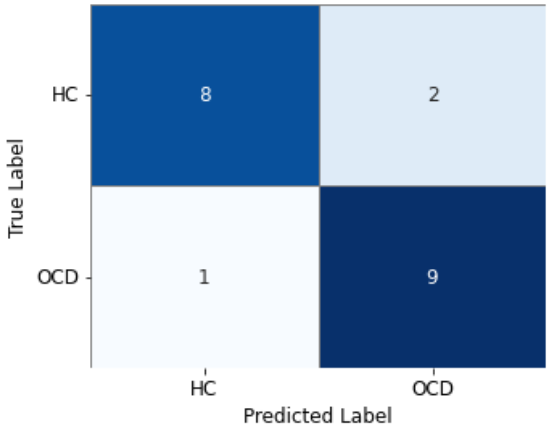

Figure 4: Confusion matrix for the multimodal model with EEG and education

SVM Augmentation

Separately, we conducted an analogous multimodal analysis starting from the subject-level score derived from the baseline SVM classifier. This SVM-only baseline, evaluated using the same LOSO framework, performed poorly (AUC = 0.39, Accuracy = 55.0%), indicating limited discriminative information captured from the traditional band power features at the subject level. The results of systematically adding individual clinical and demographic features to the SVM's baseline EEG score are presented in Table 4.

Table 4: *Predictive Performance for Clinical and Demographic Features*

Feature	*N*	AUC	Accuracy	Precision	Recall	F1	Change
Baseline EEG	20	0.39	0.55	0.55	0.60	0.57	*Baseline*
DOCS Total	20	0.97	0.85	0.89	0.80	0.84	Improved
MIS Score	18	0.62	0.67	0.71	0.56	0.63	Improved
BAI Total	20	0.94	0.90	0.90	0.90	0.90	Improved
BDI-II Total	20	0.90	0.90	1.00	0.80	0.89	Improved
Age	19	0.28	0.37	0.36	0.44	0.40	Worse
Sex	20	0.18	0.35	0.39	0.50	0.44	Worse
Education	20	0.38	0.45	0.43	0.30	0.353	Worse

Note: Baseline metrics include only the support vector machine's subject-level EEG scores in the second stage. *N* = number of subjects remaining after handling missing values. DOCS: Dimensional Obsessive-Compulsive Scale; MIS: Magical Ideation Scale; BAI: Beck Anxiety Inventory; BDI-II: Beck Depression Inventory-II.

Results showed that certain clinical symptom scores led to substantial increases in performance, particularly the total DOCS (AUC = 0.97), BAI (AUC = 0.94), and BDI-II (AUC = 0.90). The MIS score showed only a modest improvement (AUC = 0.62). Conversely, adding demographic variables (age, sex, education) resulted in performance metrics below the already poor baseline (AUC range 0.18 – 0.38). The dramatic performance increase observed when incorporating scores like the DOCS strongly suggests that these clinical variables, rather than the EEG score itself, were the primary drivers of classification success in these specific combinations.

Discussion

This pilot study investigated the feasibility and efficacy of using CNN-based classification on resting-state EEG data to differentiate individuals with OCD from HC. Our primary finding is that the CNN operating on minimally processed time-frequency representations of EEG signals achieved robust classification performance, significantly outperforming a traditional machine learning baseline model (SVM) trained on standard spectral power features. These results are generally aligned with previous research indicating better predictive power with neural networks (Salomoni et al., 2009) and the nonlinearity of OCD (Zaboski et al., 2024). Furthermore, a second-stage analysis integrating the CNN-derived EEG score with clinical and demographic variables revealed that incorporating participant education level enhanced classification accuracy, suggesting unique complementary information provided by this demographic factor.

The successful application of the CNN, achieving an aggregated AUC of 0.86 and accuracy of 81.95%, underscores the potential of deep learning to extract diagnostically relevant patterns from complex, high-dimensional resting-state EEG data in OCD populations. This

finding aligns with the growing body of literature demonstrating the power of CNNs in EEG analysis across various neurological and psychiatric conditions (Farhad et al., 2024; Xu et al., 2018), often circumventing the need for manual feature selection. The ability to distinguish OCD patients from HC participants using resting-state EEG is particularly relevant given the need for rich data sources that can be used in psychiatric research (Zaboski et al., 2021). While previous EEG studies in OCD have yielded heterogeneous findings (Metin et al., 2019), our results suggest that advanced computational techniques applied to time-frequency data may capture consistent group differences.

The marked contrast between the CNN's performance and the chance level performance of the SVM (AUC 0.45, Accuracy 49.01%) is noteworthy. The SVM was trained on canonical frequency band power features, a common approach in quantitative EEG analysis. Its performance suggests that these specific, averaged spectral features may not sufficiently capture the complex temporal and spectral dynamics differentiating OCD patients from HCs in this cohort, or that the patterns are too subtle or variable across subjects to be effectively learned by the SVM with these features. The CNN's ability to learn hierarchical features directly from the 2D time-frequency images likely allowed it to identify more intricate patterns related to oscillatory activity, cross-frequency, and spatial interactions that were missed by the traditional band power approach (Hartmann et al., 2018; Schirrmeister et al., 2017).

Our exploration of multimodal classification using a late-fusion (stacking) approach yielded intriguing preliminary results. The baseline model using only the subject-level EEG score derived from the CNN LOSO predictions already performed well (AUC 0.86, Accuracy 80.0%). Contrary to expectations, adding specific symptom scores (DOCS, MIS) or general anxiety/depression scores (BAI, BDI-II), participant age, or sex did not improve classification

performance beyond this EEG baseline. This could suggest that the variance captured by these clinical scales was either largely represented within the complex EEG features learned by the CNN, or that these scales did not provide sufficient independent discriminative information at the subject level. Notably, descriptive statistics showed high variability among scores (e.g., DOCS), as well as scores that were similar between groups (BAI). This may have limited their utility for multimodal classification in this cohort due to added noise. While heterogeneity in clinical samples is expected, it underscores the need for replicating our results in larger datasets.

Nevertheless, the inclusion of education alongside the EEG score in this sample resulted in a notable improvement, boosting the AUC to 0.89 and accuracy to 85.0%. This warrants further investigation. Education may be a proxy for a range of other relevant characteristics, such as cognitive reserve, socioeconomic status, coping mechanisms, or even subtle differences in brain structure or function related to long-term learning. These might interact with the neural signatures of OCD captured by EEG (Pérez-Vigil et al., 2018).

The findings with education warrant careful interpretation in light of the fact that our sample was highly educated but not matched across groups. While education could theoretically relate to neural signatures captured by EEG, its predictive power here may stem, at least partially, from the differences in educational attainment between the HC and OCD groups within our specific cohort (Table 1). 70% of the OCD group fell into the 'Other/Professional' category compared to only 20% of the HC group, while 40% of the HC group held doctoral degrees compared to 20% of the OCD group. The multimodal model may have leveraged these group-specific distributional differences, meaning education might be acting partly as a proxy for group status in this sample, rather than reflecting a useful neurophysiological interaction. Therefore, while a potentially interesting variable, education level requires cautious interpretation and

necessitates replication in larger, more educationally diverse samples to determine its true potential as a complementary factor in EEG-based OCD classification.

The baseline multimodal performance comparison (CNN vs. SVG) demonstrated the superior discriminative capability of the CNN-derived EEG features over the traditional SVM/band power approach. Furthermore, the multimodal analyses revealed distinct patterns for each approach. The CNN approach showed that adding education provided complementary information, modestly enhancing an already strong EEG-based signal (AUC improving from 0.86 to 0.89). In sharp contrast, the SVM required the addition of potent clinical symptom scores (DOCS, BAI, BDI-II) to achieve high classification metrics. This improvement largely compensated for the underlying weakness of the SVM's baseline EEG score, indicating the clinical scores themselves, not the SVM's limited EEG features, were responsible for the higher performance. Therefore, while the SVM + DOCS combination yielded the highest numerical AUC (0.97) in our analyses, this result primarily reflects the discriminative power of the DOCS score in this sample, not the resting state EEG signals. As such, the model achieving the most robust classification performance demonstrably driven by the EEG signal itself remains the CNN, particularly the multimodal CNN + education model.

Strengths and Limitations

Strengths of this pilot investigation include the application of a modern deep learning architecture (CNN) to minimally processed EEG data, the rigorous LOSO cross-validation approach controlling for individual subject effects, and a direct comparison with a traditional SVM baseline using identical cross-validation. Moreover, we implemented systematic investigation of multimodal integration. The preprocessing pipeline also followed established best practices using MNE-Python.

Despite these strengths, limitations must also be acknowledged. The most significant limitation is the small sample (n = 10 per group). While LOSO cross-validation maximizes the use of available data for training in each fold (Mennes et al., 2010; Ren et al., 2024), the overall model generalizability is limited, and the results require replication in larger datasets. The high variability observed in per-fold accuracy (81.29% ± 32.17%) also underscores the influence of individual subject heterogeneity due to sample size. Secondly, the high level of education in our sample may limit the generalizability of our findings to the broader population of individuals with OCD. The descriptive statistics also warrant caution in interpretation and call for replication. Third, this study focused solely on resting-state EEG; task-based paradigms might elicit different or more pronounced group differences. Fourth, the analyses were cross-sectional, using only baseline data, precluding insights into treatment effects or longitudinal changes. Finally, although the CNN performed well, the neural features driving the classification remain relatively opaque, a common challenge with deep learning models ("black box" problem).

Future Directions

Replication in larger datasets with more diverse demographic and clinical characteristics is paramount to confirm the robustness and generalizability of the CNN's performance and the contribution of education level. Investigating the underlying reasons for education's predictive value, perhaps by examining its relationship with cognitive performance or specific EEG features, would be valuable. Applying similar deep learning models to task-based EEG data (e.g., error monitoring, symptom provocation tasks) in OCD could reveal complementary diagnostic information. Utilizing advanced interpretability techniques (e.g., saliency mapping, layer-wise relevance propagation) could help elucidate the specific time-frequency patterns the CNN identified as discriminative (Cui et al., 2023; Lopes et al., 2023). Exploring different deep

learning architectures, potentially incorporating temporal dependencies more explicitly, leveraging graph neural networks to model electrode relationships, or using transformer models could yield further improvements (Cisotto et al., 2020; Klepl et al., 2024; Kuruvila et al., 2021; Yao et al., 2024). Finally, longitudinal studies are needed to assess whether these EEG-based models can predict treatment response or track clinical changes over time.

Conclusion

This pilot study demonstrates the feasibility of using CNNs applied to resting-state EEG time-frequency data to differentiate individuals with OCD from healthy controls with high accuracy, outperforming a traditional SVM approach based on spectral power. While EEG data alone contained substantial discriminative information, the integration of participant education level further improved classification performance. Despite the limitations imposed by the small sample size, these findings highlight the potential of deep learning techniques for developing powerful predictive models on complex data types. Further research and validation in larger, more diverse samples are crucial next steps towards translating these advanced computational approaches into tools for research and clinical practice.

References

Aapo, H. (1999). Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks, 10*(3), 626–634.

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., . . . Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems* [White paper]. Google Research. https://download.tensorflow.org/paper/whitepaper2015.pdf

Ablin, P., Cardoso, J.-F., & Gramfort, A. (2018). Faster independent component analysis by preconditioning with Hessian approximations. *IEEE Transactions on Signal Processing, 66*(15), 4040–4049. https://doi.org/10.1109/TSP.2018.2844203

Abramowitz, J. S., & Arch, J. J. (2014). Strategies for improving long-term outcomes in cognitive behavioral therapy for obsessive-compulsive disorder: Insights from learning theory. *Cognitive and Behavioral Practice, 21*(1), 20–31.

Abramowitz, J. S., & Deacon, B. J. (2006). Psychometric properties and construct validity of the Obsessive–Compulsive Inventory—Revised: Replication and extension with a clinical sample. *Journal of Anxiety Disorders, 20*(8), 1016–1035. https://doi.org/10.1016/j.janxdis.2006.03.001

Abramowitz, J. S., Deacon, B. J., Olatunji, B. O., Wheaton, M. G., Berman, N. C., Losardo, D., Timpano, K. R., McGrath, P. B., Riemann, B. C., Adams, T., Björgvinsson, T., Storch, E. A., & Hale, L. R. (2010). Assessment of obsessive-compulsive symptom dimensions: Development and evaluation of the Dimensional Obsessive-Compulsive Scale. *Psychological Assessment, 22*(1), 180–198. https://doi.org/10.1037/a0018260

Arevalillo-Herráez, M., Cobos, M., Roger, S., & García-Pineda, M. (2019). Combining inter-subject modeling with a subject-based data transformation to improve affect recognition from EEG signals. *Sensors (Basel, Switzerland), 19*(13), Article 2999. https://doi.org/10.3390/s19132999

Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 41*(2), 423–443. https://doi.org/10.1109/TPAMI.2018.2798607

Bardhoshi, G., Duncan, K., & Erford, B. T. (2016). Psychometric meta-analysis of the English version of the Beck Anxiety Inventory. *Journal of Counseling & Development, 94*(3), 356–373. https://doi.org/10.1002/jcad.12090

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety: Psychometric properties. *Journal of Consulting and Clinical Psychology, 56*(6), 893–900. https://doi.org/10.1037/0022-006X.56.6.893

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (2012). *Beck Anxiety Inventory* [Dataset]. American Psychological Association. https://doi.org/10.1037/t02025-000

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Beck Depression Inventory–II* [Database record]. American Psychological Association. https://doi.org/10.1037/t00742-000

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*(6), 561–571. https://doi.org/10.1001/archpsyc.1961.01710120031004

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

Bera, S., Geem, Z. W., Cho, Y.-I., & Singh, P. K. (2025). A comparative study of machine learning and deep learning models for automatic Parkinson's disease detection from electroencephalogram signals. *Diagnostics, 15*(6), Article 773. https://doi.org/10.3390/diagnostics15060773

Breiman, L. (1996). Stacked regressions. *Machine Learning, 24*(1), 49–64. https://doi.org/10.1007/BF00117832

Chaddad, A., Wu, Y., Kateb, R., & Bouridane, A. (2023). Electroencephalography signal processing: A comprehensive review and analysis of methods and techniques. *Sensors, 23*(14), Article 6434. https://doi.org/10.3390/s23146434

Chapman, L. J., Chapman, J. P., & Miller, E. N. (1982). Reliabilities and intercorrelations of eight measures of proneness to psychosis. *Journal of Consulting and Clinical Psychology, 50*(2), 187–195. https://doi.org/10.1037/0022-006X.50.2.187

Chollet, F. (2015). *Keras* [Computer software]. https://keras.io

Cisotto, G., Zanga, A., Chlebus, J., Zoppis, I., Manzoni, S., & Markowska-Kaczmar, U. (2020). *Comparison of attention-based deep learning models for EEG classification* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2012.01074

Constable, R. T. (2023). Challenges in fMRI and its limitations. In S. H. Faro & F. B. Mohamed (Eds.), *Functional neuroradiology* (pp. 471–486). Springer. https://doi.org/10.1007/978-3-031-10909-6_22

Cui, J., Yuan, L., Wang, Z., Li, R., & Jiang, T. (2023). Towards best practice of interpreting deep learning models for EEG-based brain computer interfaces. *Frontiers in Computational Neuroscience, 17*, Article 1232925. https://doi.org/10.3389/fncom.2023.1232925

Eckblad, M., & Chapman, L. J. (1983). Magical ideation as an indicator of schizotypy. *Journal*

of Consulting and Clinical Psychology, 51(2), 215–225. https://doi.org/10.1037/0022-006X.51.2.215

Erguzel, T. T., Ozekes, S., Sayar, G. H., Tan, O., & Tarhan, N. (2015). A hybrid artificial intelligence method to classify trichotillomania and obsessive compulsive disorder. *Neurocomputing, 161*, 220–228. https://doi.org/10.1016/j.neucom.2015.02.039

Farhad, S., Metin, S. Z., Uyulan, Ç., Makouei, S. T. Z., Metin, B., Ergüzel, T. T., & Tarhan, N. (2024). Application of hybrid deep learning architectures for identification of individuals with obsessive compulsive disorder based on EEG data. *Clinical EEG and Neuroscience, 55*(5), 543–552. https://doi.org/10.1177/15500594231222980

Fonseca Pedrero, E., Paino Piñeiro, M. de las M., Lemos Giráldez, S., García Cueto, E., Villazón García, U., & Muñiz Fernández, J. (2009). Psychometric properties of the Perceptual Aberration Scale and the Magical Ideation Scale in Spanish college students. *International Journal of Clinical and Health Psychology, 9*(2), 319–331. https://digibuo.uniovi.es/dspace/bitstream/handle/10651/9944/Psychometric%20properties%20of%20the%20Perceptual%20Aberration%20Scale%20and%20the%20Magical%20Ideation%20Scale%20in%20Spanish%20college%20students.pdf?sequence=1

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., Heninger, G. R., & Charney, D. S. (1989). The Yale-Brown obsessive compulsive scale: I. Development, use, and reliability. *Archives of General Psychiatry, 46*(11), 1006–1011. https://doi.org/10.1001/archpsyc.1989.01810110048007

Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience, 7*, Article 267. https://doi.org/10.3389/fnins.2013.00267

Gramfort, A., Luessi, M., Larson, E., Engemann, D. A., Strohmeier, D., Brodbeck, C.,

Parkkonen, L., & Hämäläinen, M. S. (2014). MNE software for processing MEG and EEG data. *NeuroImage, 86*, 446–460. https://doi.org/10.1016/j.neuroimage.2013.10.027

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

Harris, P. A., Taylor, R., Minor, B. L., Elliott, V., Fernandez, M., O'Neal, L., McLeod, L., Delacqua, G., Delacqua, F., Kirby, J., & Duda, S. N. (2019). The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics, 95*, Article 103208. https://doi.org/10.1016/j.jbi.2019.103208

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics, 42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Hartmann, K. G., Schirrmeister, R. T., & Ball, T. (2018). Hierarchical internal representation of spectral features in deep convolutional networks trained for EEG decoding. *2018 6th International Conference on Brain-Computer Interface (BCI)* (pp. 1–6). IEEE. https://doi.org/10.1109/IWW-BCI.2018.8311493

Hosseini, M.-P., Tran, T. X., Pompili, D., Elisevich, K., & Soltanian-Zadeh, H. (2020). Multimodal data analysis of epileptic EEG and rs-fMRI via deep learning and edge computing. *Artificial Intelligence in Medicine, 104*, Article 101813. https://doi.org/10.1016/j.artmed.2020.101813

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

Jiang, M., Zhao, Q., Li, J., Wang, F., He, T., Cheng, X., Yang, B. X., Ho, G. W. K., & Fu, G. (2024). *A generic review of integrating artificial intelligence in cognitive behavioral therapy* [Preprint]. arXiv. https://doi.org/10.48550/arXiv.2407.19422

Karpiak, C. P., & Zaboski, B. A. (2013). Lifetime prevalence of mental disorders in the general population. In G. P. Koocher, J. C. Norcross, & B. A. Greene (Eds.), *Psychologists' desk reference* (3rd ed., pp. 3–16). Oxford University Press.

Kessler, R. C., Petukhova, M., Sampson, N. A., Zaslavsky, A. M., & Wittchen, H.-U. (2012). Twelve-month and lifetime prevalence and lifetime morbid risk of anxiety and mood disorders in the United States. *International Journal of Methods in Psychiatric Research, 21*(3), 169–184. https://doi.org/10.1002/mpr.1359

Kingdon, B. L., Egan, S. J., & Rees, C. S. (2012). The illusory beliefs inventory: A new measure of magical thinking and its relationship with obsessive compulsive disorder. *Behavioural and Cognitive Psychotherapy, 40*(1), 39–53. https://doi.org/10.1017/S135246581100063X

Klepl, D., Wu, M., & He, F. (2024). Graph neural network-based EEG classification: A survey. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 32*, 493–503. https://doi.org/10.1109/TNSRE.2023.3348801

Krohn, J., Beyleveld, G., & Bassens, A. (2019). *Deep learning illustrated*. Addison-Wesley Professional.

Kuruvila, I., Muncke, J., Fischer, E., & Hoppe, U. (2021). Extracting the auditory attention in a dual-speaker scenario from EEG using a joint CNN-LSTM model. *Frontiers in*

Physiology, 12, Article 700655. https://doi.org/10.3389/fphys.2021.700655

Kwapil, T. R., Miller, M. B., Zinser, M. C., Chapman, J., & Chapman, L. J. (1997). Magical ideation and social anhedonia as predictors of psychosis proneness: A partial replication. *Journal of Abnormal Psychology, 106*(3), 491–500. https://doi.org/10.1037/0021-843X.106.3.491

Law, C., & Boisseau, C. L. (2019). Exposure and response prevention in the treatment of obsessive-compulsive disorder: Current perspectives. *Psychology Research and Behavior Management, 12*, 1167–1174. https://doi.org/10.2147/PRBM.S211117

Lee, T.-W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Computation, 11*(2), 417–441. https://doi.org/10.1162/089976699300016719

Lopes, M., Cassani, R., & Falk, T. H. (2023). Using CNN saliency maps and EEG modulation spectra for improved and more interpretable machine learning-based Alzheimer's disease diagnosis. *Computational Intelligence and Neuroscience, 2023*, Article 3198066. https://doi.org/10.1155/2023/3198066

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering, 4*(2), R1–R13. https://doi.org/10.1088/1741-2560/4/2/R01

Martin, E. A., Becker, T. M., Cicero, D. C., Docherty, A. R., & Kerns, J. G. (2011). Differential associations between schizotypy facets and emotion traits. *Psychiatry Research, 187*(1–2), 94–99. https://doi.org/10.1016/j.psychres.2010.12.007

Mattera, E. F., Ching, T. H., Zaboski, B. A., & Kichuk, S. A. (2024). Suicidal obsessions or suicidal ideation? A case report and practical guide for differential assessment. *Cognitive*

and Behavioral Practice, 31(2), 259-271.

McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 51–56). SciPy. https://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf

Mennes, M., Wouters, H., Vanrumste, B., Lagae, L., & Stiers, P. (2010). Validation of ICA as a tool to remove eye movement artifacts from EEG/ERP. *Psychophysiology, 47*(6), 1142–1150. https://doi.org/10.1111/j.1469-8986.2010.01015.x

Metin, S. Z., Balli Altuglu, T., Metin, B., Erguzel, T. T., Yigit, S., Arıkan, M. K., & Tarhan, K. N. (2020). Use of EEG for predicting treatment response to transcranial magnetic stimulation in obsessive compulsive disorder. *Clinical EEG and Neuroscience, 51*(4), 246–253. https://doi.org/10.1177/1550059419879569

Mevlevioğlu, D., Tabirca, S., & Murphy, D. (2024). Real-time classification of anxiety in virtual reality therapy using biosensors and a convolutional neural network. *Biosensors, 14*(3), Article 131. https://doi.org/10.3390/bios14030131

Naidu, G., Zuva, T., & Sibanda, E. M. (2023). A review of evaluation metrics in machine learning algorithms. In R. Silhavy & P. Silhavy (Eds.), *Artificial intelligence application in networks and systems* (Lecture Notes in Networks and Systems, Vol. 724, pp. 15–25). Springer. https://doi.org/10.1007/978-3-031-35314-7_2

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830. http://www.jmlr.org/papers/v12/pedregosa11a.html

Pérez-Vigil, A., de la Cruz, L. F., Brander, G., Isomura, K., Jangmo, A., Feldman, I., Hesselmark, E., Serlachius, E., Lázaro, L., & Rück, C. (2018). Association of obsessive-compulsive disorder with objective indicators of educational attainment: A nationwide register-based sibling control study. *JAMA Psychiatry, 75*(1), 47–55. https://doi.org/10.1001/jamapsychiatry.2017.3603

Pittenger, C. (2017). *Obsessive-compulsive disorder: Phenomenology, pathophysiology, and treatment*. Oxford University Press.

Python Software Foundation. (2021). *Python language reference* (Version 3.10). https://docs.python.org/3.10/reference/index.html

Rahul, J., Sharma, D., Sharma, L. D., Nanda, U., & Sarkar, A. K. (2024). A systematic review of EEG based automated schizophrenia classification through machine learning and deep learning. *Frontiers in Human Neuroscience, 18*, Article 1347082. https://doi.org/10.3389/fnhum.2024.1347082

Rance, M., Zhao, Z., Zaboski, B., Kichuk, S. A., Romaker, E., Koller, W. N., . . . & Hampson, M. (2023). Neurofeedback for obsessive compulsive disorder: A randomized, double-blind trial. *Psychiatry Research*, *328*, Article 115458.

Ren, G., Kumar, A., Mahmoud, S. S., & Fang, Q. (2024). A deep neural network and transfer learning combined method for cross-task classification of error-related potentials. *Frontiers in Human Neuroscience, 18*, Article 1394107. https://doi.org/10.3389/fnhum.2024.1394107

Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2010). The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Molecular Psychiatry, 15*(1), 53–63. https://doi.org/10.1038/mp.2008.94

Salehzadeh, R., Soylu, F., & Jalili, N. (2023). A comparative study of machine learning methods for classifying ERP scalp distribution. *Biomedical Physics & Engineering Express, 9*(4), Article 045010. https://doi.org/10.1088/2057-1976/acdbd0

Salomoni, G., Grassi, M., Mosini, P., Riva, P., Cavedini, P., & Bellodi, L. (2009). Artificial neural network model for the prediction of obsessive-compulsive disorder treatment response. *Journal of Clinical Psychopharmacology, 29*(4), 343-349.

Schirrmeister, R. T., Springenberg, J. T., Fiederer, L. D. J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., & Ball, T. (2017). Deep learning with convolutional neural networks for EEG decoding and visualization. *Human Brain Mapping, 38*(11), 5391–5420. https://doi.org/10.1002/hbm.23730

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., & Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *The Journal of Clinical Psychiatry, 59*(Suppl. 20), 22–33.

Stein, D. J., Costa, D. L. C., Lochner, C., Miguel, E. C., Reddy, Y. C. J., Shavitt, R. G., van den Heuvel, O. A., & Simpson, H. B. (2019). Obsessive–compulsive disorder. *Nature Reviews Disease Primers, 5*, Article 52. https://doi.org/10.1038/s41572-019-0102-3

Tamburro, G., Croce, P., Zappasodi, F., & Comani, S. (2021). Is brain dynamics preserved in the EEG after automated artifact removal? A validation of the fingerprint method and the automatic removal of cardiac interference approach based on microstate analysis. *Frontiers in Neuroscience, 14*, Article 577160. https://doi.org/10.3389/fnins.2020.577160

Tolin, D. F., Gilliam, C., Wootton, B. M., Bowe, W., Bragdon, L. B., Davis, E., Hannan, S. E.,

Steinman, S. A., Worden, B., & Hallion, L. S. (2018). Psychometric properties of a structured diagnostic interview for DSM-5 anxiety, mood, and obsessive-compulsive and related disorders. *Assessment, 25*(1), 3–13. https://doi.org/10.1177/1073191116638410

Wang, Y.-P., & Gorenstein, C. (2013). Psychometric properties of the Beck Depression Inventory-II: A comprehensive review. *Brazilian Journal of Psychiatry, 35*(4), 416–431. https://doi.org/10.1590/1516-4446-2012-1048

Waskom, M. L. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software, 6*(60), Article 3021. https://doi.org/10.21105/joss.03021

Wolpert, D. H. (1992). Stacked generalization. *Neural Networks, 5*(2), 241–259. https://doi.org/10.1016/S0893-6080(05)80023-1

Xu, B., Zhang, L., Song, A., Wu, C., Li, W., Zhang, D., Xu, G., Li, H., & Zeng, H. (2018). Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification. *IEEE Access, 7*, 6084–6093. https://doi.org/10.1109/ACCESS.2018.2889094

Yao, X., Li, T., Ding, P., Wang, F., Zhao, L., Gong, A., Nan, W., & Fu, Y. (2024). Emotion classification based on Transformer and CNN for EEG spatial–temporal feature learning. *Brain Sciences, 14*(3), Article 268. https://doi.org/10.3390/brainsci14030268

Zaboski, B. A., & Bednarek, L. (2025). Precision psychiatry for obsessive-compulsive disorder: Clinical applications of deep learning architectures. *Journal of Clinical Medicine, 14*(7), Article 2442.

Zaboski, B. A., Stern, E. F., Skosnik, P. D., & Pittenger, C. (2021). Electroencephalographic correlates and predictors of treatment outcome in OCD: A brief narrative review. *Frontiers in Psychiatry, 12*, Article 703398. https://doi.org/10.3389/fpsyt.2021.703398

Zaboski, B. A., Wilens, A., McNamara, J. P., & Muller, G. N. (2024). Predicting OCD severity from religiosity and personality: A machine learning and neural network approach. *Journal of Mood & Anxiety Disorders, 8*, Article 100089.

Zhang, L., Xiao, F., & Cao, Z. (2023). Multi-channel EEG signals classification via CNN and multi-head self-attention on evidence theory. *Information Sciences, 642*, Article 119107. https://doi.org/10.1016/j.ins.2023.119107