Prediction of Post-Operative Delirium in Older Adults from Preoperative Cognition and Alpha Power from Resting-State EEG

Authors: Matthew Ning[a,1], Andrei Rodionov[a,b,c,1], Jessica M. Ross[a,d,e], Recep A. Ozdemir[a,f,g], Maja Burch[a], Shu Jing Lian[a], David Alsop[g, h], Michele Cavallari[h,i], Bradford C. Dickerson[g,j], Tamara G. Fong[f,g,k], Richard N. Jones[l], Towia A. Libermann[g,m], Edward R. Marcantonio[g,n], Emiliano Santarnecchi[a,g,o], Eva M. Schmitt[k], Alexandra Touroutoglou[g,j], Thomas G. Travison[g,k], Leah Acker[p,q,s], Melody Reese[p,q], Haoqi Sun[f,j,r], Brandon Westover[f,j,r], Miles Berger[p,q,s,t,2], Alvaro Pascual-Leone[f,g,k,2], Sharon K. Inouye[g,k,m,2] and Mouhsin M. Shafi[a,f,g,2,*] for the SAGES II Study Group and the INTUIT/PRIME Study Groups

a Berenson-Allen Center for Noninvasive Brain Stimulation, Beth Israel Deaconess Medical Center, Boston, MA, USA

b BioMag Laboratory, HUS Medical Imaging Center, Helsinki University Hospital, Helsinki, Finland

c Faculty of Educational Sciences, University of Helsinki, University of Helsinki, Finland

d Veterans Affairs Palo Alto Healthcare System, and the Sierra Pacific Mental Illness, Research, Education, and Clinical Center, Palo Alto, CA, USA

e Department of Psychiatry and Behavioral Sciences, Stanford Medical School, Stanford, CA, USA

f Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, USA

g Harvard Medical School, Boston, MA, USA

h Department of Radiology, Beth Israel Deaconess Medical Center, Boston, MA, USA

i Center for Neurological Imaging, Department of Radiology, Brigham and Women's Hospital, Boston, MA, USA

j Department of Neurology, Massachusetts General Hospital, Boston, MA, USA

k Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA, USA

l Department of Psychiatry and Human Behavior, Department of Neurology, Warren Alpert Medical School, Brown University, Providence, RI, USA

m Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

n Divisions of General Medicine and Gerontology, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA

o Precision Neuroscience & Neuromodulation Program (PNN), Gordon Center for Medical Imaging, Department of Radiology, Massachusetts General Hospital, Boston, MA, USA

p Department of Anesthesiology, Duke University School of Medicine, Durham, NC, USA

q Duke Center for the Study of Aging and Human Development, Duke University School of Medicine, Durham, NC, USA

r Henry and Allison McCance Center for Brain Health at Mass General

s Center for Cognitive Neuroscience, Duke University, Durham, NC, USA

t Duke-UNC Alzheimer's Disease Research Center, Durham, NC, USA

1 These co-first authors contributed equally to this work.

2 Shared senior authorship

*Corresponding Author:

Berenson-Allen Center for Non-Invasive Brain Stimulation, Beth Israel Deaconess Medical Center, Department of Neurology, Harvard Medical School, Boston, MA, USA;

mshafi@bidmc.harvard.edu

Abstract:

Postoperative Delirium (POD) is the most common complication following surgery among older adults, and has been consistently associated with increased mortality and morbidity, cognitive decline, and loss of independence, as well as markedly increased health-care costs. The development of new tools to identify individuals at high risk for POD could guide clinical decision-making and enable targeted interventions to potentially decrease delirium incidence and POD-related complications. In this study, we used machine learning techniques to evaluate whether baseline (pre-operative) cognitive function and resting-state electroencephalography could be used to identify patients at risk for POD. Pre-operative resting-state EEGs and the Montreal Cognitive Assessment (MoCA) were collected from 85 patients (age = 73 ± 6.4 years) undergoing elective surgery, 12 of whom subsequently developed POD. The model with the highest f1-score for predicting delirium, a linear-discriminant analysis (LDA) model incorporating MoCA scores and occipital alpha-band EEG features, was subsequently validated in an independent, prospective cohort of 51 older adults (age ≥ 60) undergoing elective surgery, 6 of whom developed POD. The LDA-based model, with a total of 7 features, was able to predict POD with area under the receiver operating characteristic curve, specificity and accuracy all >90%, and sensitivity > 80%, in the validation cohort. Notably, models incorporating both resting-state EEG and MoCA scores outperformed those including either EEG or MoCA alone. While requiring prospective validation in larger cohorts, these results suggest that prediction of POD with high accuracy may be feasible in clinical settings using simple and widely available clinical tools.

Highlights:

- Predict postoperative delirium using pre-operative EEG alpha power and MoCA scores.
- Prediction performance improves over cognitive assessment alone.
- ROC-AUC, specificity, accuracy >90%, and sensitivity > 80%, in a validation cohort.
- Abnormalities in baseline EEG are a risk factor for postoperative delirium.

Introduction:

Delirium is a complex neuropsychiatric syndrome that is characterized by an acute, fluctuating disturbance in attention, level of consciousness, and cognition[1]. Post-operative delirium (POD) is a common complication in older adults after surgery, occurring in 19-32% of patients[2], and is associated with longer intensive care unit and hospital stay, increased post-discharge institutionalization, persistent cognitive decline, and increased short- and long-term mortality[3,4,5,6,7]. POD has an estimated annual healthcare cost of over $30 billion in the USA alone[8].

Early and accurate identification of individuals at high risk could enable interventions to reduce the incidence, severity and duration of POD. Such interventions might include a careful evaluation of the risk:benefit ratio for surgery; enabling brain health optimization prior to surgery e.g. via elimination of medications that increase delirium risk or through pre-operative transcranial direct current stimulation[9]; modification of intraoperative anesthesia[10]; postoperative treatment with acetaminophen[11]; and through targeted implementation of more intensive behavioral protocols before and after the surgery (e.g. the ABCDEF Bundle[12], HELP[13,14] and mHELP programs[15], where the latter achieved a 56% reduction of POD incidence rate within the intervention versus control group). Many studies have shown that pre-operative cognitive impairment is a strong predictor of POD[16,17]. Other risk factors include age, history of alcohol abuse, history of smoking, medical comorbidity and pre-existing impairment in activities of daily living[18,19]. However, while these risk factors increase delirium risk at the group level, there are limited clinically useful tools to predict delirium at the individual level. Furthermore, despite studies linking abnormalities in cerebral oscillatory activity with delirium[20], the specific mechanisms by which risk factors for delirium are related to the underlying brain dysfunction and subsequent delirium symptoms are still not fully understood.

As originally proposed in Shafi et al 2017[21], we hypothesized a neurophysiological model of delirium. Specifically, in our model, delirium is the result of a breakdown of cognitive functions in individuals with pre-existing impairments in brain connectivity and plasticity exposed to a stressor, such as surgery. EEG is a candidate neuroimaging technology to measure cortical connectivity and physiology. We have found intraoperative EEG metrics associated with baseline cognitive impairment[20] and up to a 4 fold increase in delirium risk[22]. We also have found preliminary evidence of association between resting-state EEG (rsEEG) power ratios and POD[23]. An independent study on the relationship between preoperative rsEEG and POD showed that a lower median dominant frequency from prefrontal cortex was associated with POD[24]. However, both studies were limited to individuals without pre-operative cognitive impairments.

Based on our conceptual model of delirium[21] and the studies described above, we hypothesized that machine learning techniques applied to preoperative resting-state EEG features and baseline measures of cognitive function can predict individual POD risk with high performance. This hypothesis was tested by developing machine learning models. An initial data exploration across electrodes and frequency bands suggested that occipital alpha power was a promising feature. We subsequently evaluated a range of models utilizing occipital alpha powers and baseline Montreal Cognitive Assessment (MoCA) scores as features in one prospective cohort of 85 older adults age 60+ (of whom 12 developed POD) undergoing elective surgery, and the best-performing model was subsequently validated in a second independent prospective cohort of 51 older adults age 65+ undergoing surgery at another institution, of whom 6 developed POD. We also assessed whether models combining EEG and cognitive function performed better than those using either EEG or cognitive testing alone.

Materials and Method:

Participants:

The dataset from Successful Aging after Elective Surgery renewal (SAGES II, NIH-NIA P01AG031720) study[25], a prospective observational cohort study of older adults scheduled for major elective non-cardiac surgery, was used for the model selection process. SAGES II here will be referred to as SAGES hereafter for brevity. In that study, 420 participants were enrolled. Inclusion criteria were age ≥60 years, English speakers, ability to communicate verbally, scheduled for elective surgery at the following Harvard-affiliated hospitals: Beth Israel Deaconess Medical Center (BIDMC), Brigham and Women's Hospital (BWH), or Brigham and Women's Faulkner Hospital (BWFH), and availability for in-person interviews. Exclusion criteria were delirium at baseline, prior hospitalization within 3 months, current hemodialysis or cancer chemotherapy, legal blindness, severe deafness, terminal condition and history of heavy alcohol abuse or withdrawal. Of the total SAGES cohort of 420 participants, a subgroup (*n*=92), meeting additional criteria (Supplementary material), participated in the EEG study. Of these, two (2.2%) participants were dropped from the EEG study due to a surgery cancellation. The remaining 90 participants (age = 72.5±6.3, range 65-93; 33 males, 57 females) were included in the study sample. Five participants were excluded due to low EEG data quality. Thus, 85 participants remain available for the data analysis. Of these, 12 participants (14% of the analysis sample) developed delirium. Demographic and clinical information for the analytic sample is presented in Table 1 (see Supplementary Table S1 for the excluded cohort). Written informed consent for study participation was obtained from all participants according to procedures approved by the institutional review boards of Beth Israel Deaconess Medical Center, Brigham and Women's Hospital, and the Brigham and Women's Faulkner Hospital — the study hospitals, and Hebrew SeniorLife — the study coordinating center, all located in Boston, Massachusetts.

	Full sample (n=85)	Delirium (n=12)	No Delirium (n=73)	p-value (t-statistic) (dof=83)
Age, mean year (SD)	73 (6.4)	75 (9.5)	72 (5.8)	--
Female, n (%)	56 (65.9)	10 (83.3)	46 (63.0)	--
MoCA, mean score (SD)	26.1 (3.4)	21.8 (5.4)	26.8 (2.4)	2.09e-6 (5.10)
Surgery type, n (%)				
Total knee replacement	52 (61.2)	8 (66.7)	44 (60.3)	--
Total hip replacement	25 (29.4)	3 (25.0)	22 (30.1)	--
Other	8 (9.4)	1 (8.3)	7 (9.6)	--
Anesthesia type, n (%)				
Spinal	78 (91.8)	11 (91.7)	67 (91.8)	--
General	7 (8.2)	1 (8.3)	6 (8.2)	--
General and Spinal	0 (0.0)	0 (0.0)	0 (0.0)	--

Table 1 Demographics of SAGES analysis sample. Two-tailed two-sample t-test was used to compute the p-value only for the MoCA scores.

The other dataset used for model validation is from Investigating Neuroinflammation Underlying Postoperative Cognitive Dysfunction (INTUIT)[26], also a prospective observational cohort study registered on clinicaltrials.gov (NCT03273335). INTUIT enrolled 201 participants undergoing non-cardiac/non-neurologic surgeries lasting ≥2 hours with a planned postoperative overnight hospitalization. Inclusion criteria were age ≥60 years, English speakers, ability to communicate verbally and scheduled for non-cardiac/non-neurologic surgery at Duke University Medical Center or Duke Regional Hospital, and available for in-person study visits. Unlike SAGES, INTUIT had no exclusion criteria based on

preoperative cognition or delirium status, however, none of the subjects had delirium prior to their scheduled surgery. Exclusion criteria were 1) patients on immunosuppressant (e.g. steroids) or immunomodulatory therapy, chemotherapeutic agents with known cognitive effects, or anticoagulants that would preclude safe lumbar puncture, 2) inmates of correctional facilities, 3) patients who experienced major head trauma or received chemotherapy between the baseline and either postoperative cognitive testing session. A subgroup of INTUIT study participants (n=81) participated in an EEG study funded by the separate PRIME study grant[27]. 63 out of the 81 INTUIT/PRIME EEG participants wore a 32-channel EEG cap with standard international 10-20 montage (Supplementary Figure S1A). The remaining 15 subjects wore a custom 32-channel EEG cap (Supplementary Figure S1B). Due to the lack of overlapping channels with the standard international 10-20 montage, subjects with the custom EEG montage were excluded from further analysis. 1 participant withdrew from the study. EEG data was unable to be collected from 3 participants and 11 participants' EEG data was excluded due to poor EEG quality, leaving 51 INTUIT participants with EEG data available for analysis, of whom 6 developed delirium (12% of the analysis sample). Demographic and clinical information for this 51 patient cohort from the INTUIT/PRIME study is presented in Table 2. Written informed consent was obtained from all participants. The study was approved by Duke University Health System Institutional Review Board.

	Full sample (n=51)	Delirium (n=6)	No Delirium (n=45)	p-value (t-statistic) (dof=49)
Age, mean year (SD)	68.3 (5.2)	71.2 (5.7)	68.0 (5.2)	--
Female, n (%)	25 (49.0)	1 (16.7)	24 (53.3)	--
MMSE, mean score (SD)	27.5 (2.4)	24.0 (2.6)	28.0 (1.9)	4.15e-5 (4.50)
MoCA Converted, mean score (SD)	23.7 (3.6)	18.5 (3.4)	24.4 (3.0)	8.82e-5 (4.27)
Surgery type, n (%)				
Total knee replacement	3 (5.9)	0 (0.0)	3 (6.7)	--
Total hip replacement	1 (2.0)	0 (0.0)	1 (2.2)	--
Other	47 (92.1)	6 (100.0)	41 (91.1)	--
Anesthesia type, n (%)				
General	39 (76.5)	4 (66.7)	35 (77.8)	--
General and Neuraxial Block	8 (15.7)	2 (33.3)	6 (13.3)	--
MAC/Neuraxial Block	3 (5.9)	0 (0.0)	3 (6.7)	--

Table 2 Demographics of INTUIT/PRIME analysis sample. Two-tailed two-sample t-test was used to compute the p-value only for the MMSE and MoCA Converted scores.

Clinical Assessments

For the SAGES Study, delirium was assessed with the Confusion Assessment Method (CAM) long-form daily post-surgery. CAM is a standardized and internationally accepted tool that enables non-psychiatrically trained clinicians to identify and recognize delirium quickly and accurately in both clinical and research settings (sensitivity 94-100%, specificity 90-95%, interrater reliability 84-100%)[28]. Baseline cognitive function was assessed with The Montréal Cognitive Assessment (MoCA, range 0-30, 0=most impaired)[29]. MoCA were conducted either through video call over Zoom or face-to-face in the patient's place of residence prior to surgery. A MoCA score of 26 or above was considered normal while scores less than 26 were considered as indicative of possible cognitive impairment. Because the MoCA was administered after other neuropsychological assessments, the memory subdomain test of MoCA was corrected for possible interference effect from the previous memory tests (See Supplementary materials).

For the INTUIT/PRIME Study, a short form of CAM, called the 3D-CAM, was used to identify delirium. The 3D-CAM has excellent overall agreement with the long-form CAM[30,31]. The Mini-Mental Status Examination (MMSE) was used in the INTUIT/PRIME study for assessment of baseline cognitive function (range 0-30, 0-most impaired). The MMSE was converted to MoCA scores based on an established cross-walk[32,33]. Since one MMSE score can be mapped to multiple MoCA scores, the median of multiple MoCA scores was used for our study following recommended conversion procedures (see Supplementary Table 3 for the conversion table).

EEG Data Collection

For the SAGES EEG Sub-Study, participants attended a baseline EEG visit at least 2-3 days before and up to two months in advance of their scheduled surgery. RsEEG was recorded at 5 kHz while participants were seated in a comfortable chair during a 3-4 hour study visit that included collection of Transcranial Magnetic Stimulation and EEG data. RsEEG was the first neurophysiological measurement taken during the study visit and was recorded for 3-7 minutes each in two conditions: with eyes open (EO) and eyes closed (EC). For the eyes closed condition, the participants were instructed to maintain wakefulness, and were queried intermittently to ensure alertness. RsEEG was recorded with a 64-channel Brainproducts system with AFz as the reference channel and the electrode impedances < 20 kΩ (extended 10-20 international system, reference channel: AFz, Supplementary Figure S1C) and amplifiers (actiCHamp, Brain Products GmbH, Munich, Germany). All participants tolerated the procedure and no significant side effects were reported or noticed.

For the INTUIT/PRIME Study, rsEEG was recorded at 1 kHz from 32 electrodes embedded in a 64-electrode cap custom designed for extended scalp coverage,[34] with Cz as the reference channel and the electrode impedances < 20 kΩ. (BrainAmpMR Plus, Brain Products GmbH, Gilching, Germany). INTUIT/PRIME rsEEG was recorded in the preoperative patient holding area just prior to surgery, for 3 min each with eyes open and with eyes closed. Participants were also instructed and monitored to maintain wakefulness during the eyes closed condition.

EEG Processing and Analysis

EEG processing here is similar to our previous work[23]. In both data sets, EEG signals were first down-sampled to 500 Hz. In order, notch (band stop frequency 57-63 Hz, 4th total order forward-backward Butterworth), high-pass (1 Hz high pass frequency, 4th total order forward-backward Butterworth) and low-pass (50 Hz low-pass frequency, 4th total order forward-backward Butterworth) filters were applied. EEG channels contaminated by artifacts were manually identified and removed (Number of rejected channels reported as average ± SD. SAGES: 2.0 ± 2.0, INTUIT/PRIME: 1.3 ± 1.6. Number of remaining channels reported as average ± SD. SAGES: 60.9 ± 2.0, INTUIT/PRIME: 30.7 ± 1.6). The data were divided into 3-second epochs, visually inspected and bad epochs were manually removed. Number of rejected epochs are reported as average ± SD. SAGES: 2.0 ± 4.7, INTUIT/PRIME: 3.7 ± 4.4. Number of remaining epochs are reported as average ± SD. SAGES: 58.6 ±10.0, INTUIT/PRIME: 56.4 ± 9.6. After that, EEG data were re-referenced to common average reference. Prior to running independent component analysis (ICA) on EEG channels, principal component analysis (PCA) was used to reduce the dimension (30 for SAGES Data Set and 24 for INTUIT/PRIME Data Set). Fast independent component analysis (fICA v2.5, http://research.ics.aalto.fi/ica/fastica/) EEGLAB plugin was used to compute independent components (ICs). Non-brain ICs that represented blink/eye movement, electromyographic activity, single electrode noise, or cardiac beats artifacts were manually identified based on their power spectrum, amplitude, scalp topography, and time course using TMS-EEG Signal Analyser (TESA v1.1.1, http://nigelrogasch.github.io/TESA)[35, 36] EEGLAB toolbox (SAGES: Total number of analyzed ICs – 30, average ±SD rejected ICs = 16.1 ±5.0; total average ±SD remaining ICs = 14.0 ±5.0. INTUIT/PRIME:

Total number of analyzed ICs – 24, average ±SD rejected ICs = 9.4 ±3.2; total average ±SD remaining ICs = 14.6 ±3.2). Any subjects where more than 85% of the EEG signal variance are removed via ICA were excluded. As a result, for the SAGES Data Set, 5 subjects were excluded from further analysis, leaving 85 subjects available for analysis. For the INTUIT/PRIME Data Set, 12 subjects were excluded from further analysis, leaving 51 subjects remaining. Channels rejected during the previous steps were interpolated using spherical interpolation. Individual channel spectrum plots were inspected and remaining bad channels were removed and interpolated again using spherical interpolation (Number of channel rejected, reported as average ± SD: SAGES: 1.0 ± 1.7, INTUIT/PRIME: 0.2 ± 0.6). The data recorded from the channels FT9, FT10, TP9 and TP10 was contaminated by artifacts in most of the SAGES subjects. To ensure high signal to noise ratio, these channels were removed from all SAGES datasets. These channels weren't available in the INTUIT/PRIME datasets.

Features and Transformation

We first performed a preliminary data exploration in the SAGES cohort (Figure 1A). Specifically, we systematically examined EEG spectral powers of several different frequency bands as well as ratios of frequency band powers (e.g. [(alpha+beta power)/(delta+theta power)]) from different regions of interest (ROIs) to identify features that might potentially predict delirium (list of frequency bands, ratios and ROIs are in Supplementary Table S4). Additionally, we tested two different data transformations: standardization to Gaussian distribution with 0 mean and unit variance, and log of the distance to the median (Equation 1). Power spectral densities (PSDs), scatter plots of band powers, EEG topographic plots and t-tests all were used to assess potential discriminability between the delirium and non-delirium groups. Through this process, alpha (8-12 Hz) and sub-alpha powers (8-10 Hz, 10-12 Hz) from the occipital region were identified as promising candidates for further analysis and model development. For each of 3 electrodes in the occipital region (O1, O2 and POz), the power spectral density (PSD) was estimated using the multitaper method (using Discrete Prolate Spheroidal (Slepian) Sequences as tapers) and was estimated separately for the eyes-open (EO) and eyes-closed (EC) conditions. 2 different sets of features were tested for the classification performance (Supplementary Table S5). To test whether the resting state occipital alpha powers are predictive of post-operative delirium, the alpha power was computed from the PSD for each electrode and condition and was defined as the average power of the alpha frequency range ([8-12] Hz). In addition, we also tested subsets of the alpha frequency range. Using a bandwidth of 2 Hz, the average band powers were computed using the 8-10 Hz range for the EC condition and the 10-12 Hz range for the EO condition. In both feature sets, the powers were normalized to the SAGES median as follows:

$$\tilde{p}_{i,j,k} = \left|\log_{10} p_{i,j,k} - \log_{10} median(\mathbf{p}_{j,k})\right| \qquad \text{Equation 1}$$

where i represents an individual, j represents a channel, k represents an eye condition and $\mathbf{p}$ represents the SAGES sample of powers for each channel and condition. In both SAGES and INTUIT/PRIME Data Sets, the median is defined as the median of the entire SAGES Data Set for each electrode and condition. The MoCA scores were transformed using the z-score. The z-scores for both SAGES and INTUIT/PRIME Data Sets were computed using the sample mean and standard deviation of the SAGES Data Set. MoCA was used as a feature in both feature sets. Thus, in both feature sets tested, a total of 7 features were used: 6 EEG powers and 1 cognitive score (Supplementary Table S5). The feature set with the alpha powers and MoCA will be referred to throughout this paper as the Principal Feature Set (left column of Supplementary Table S5). The feature set containing the subsets of the alpha powers and MoCA will be referred to as the Sub-Alpha Feature Set (right column of Supplementary Table S5). The performance of both feature sets was benchmarked against the performance of the models trained on

MoCA alone, referred to as MoCA-Alone. Two other reference sets of features for comparison are Alpha Powers-Alone and Sub-Alpha Powers-Alone, which consist of only EEG features (no MoCA).

ML Models and Cross-Validation

The SAGES cohort was used for the model selection step (Figure 1B). The following models were tested during the model selection step: logistic regression with L2 regularization with inner cross-validation (CV) for hyperparameter tuning; linear discriminant analysis (LDA) with inner CV for hyperparameter tuning; linear discriminant analysis (LDA) using Ledoit-Wolf estimator (LDA LW); LDA using Oracle Shrinkage Approximation estimator (LDA OA); nearest shrunken centroid (NSC) using Manhattan distance metric; NSC using Euclidean distance metric; Gaussian Naïve Bayes (GNB) using empirical priors; GNB using a priori-defined class priors (80-20 ratio) and decision tree. During the model selection step, the SAGES Data Set was cross-validated using 10 repetitions of 5-fold stratified CV (Figure 1D). The 95% confidence intervals were estimated using the student's t-distribution using the sample mean and standard deviation of 50 folds. Due to class imbalance, the performances of the models were assessed with 8 metrics: accuracy, sensitivity, specificity, f1-score, area under the curve of the receiver operating characteristic curve (ROC-AUC), precision-recall curve (PR-AUC) and positive predictive value (PPV, also known as precision) and negative predictive value (NPV). To prioritize the minority class (delirium), f1-score was chosen as the deciding factor due to it being the harmonic mean of sensitivity and precision.

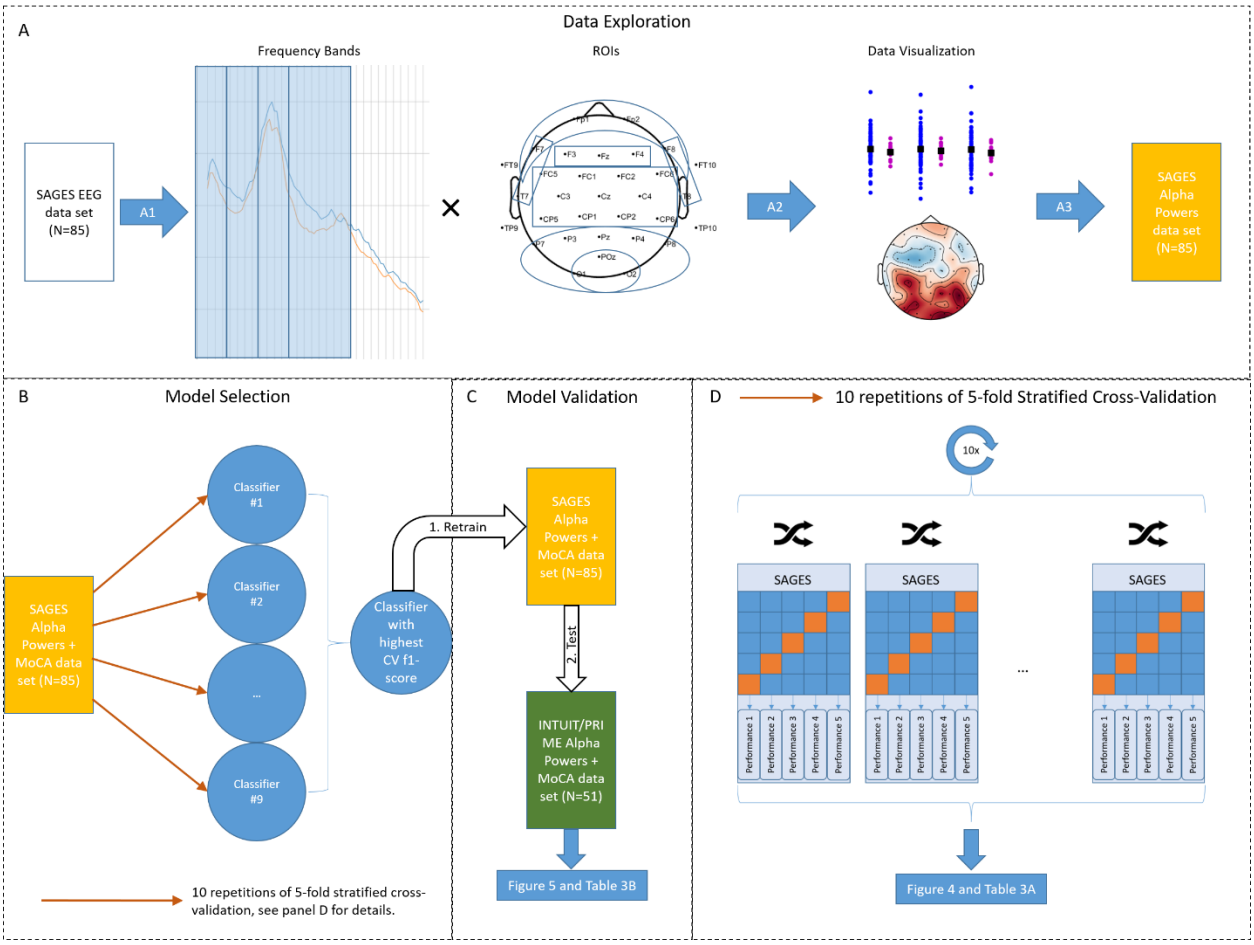

Figure 1 ***Schematic diagram of machine learning framework.*** *A) Data exploration. A1) different band powers of different regions of interested (ROIs) were extracted from SAGES EEG Data Set. A2) the results are visualized using scatter plots as well as EEG topographic plots. A3) occipital alpha powers (and sub-alpha powers) from both eyes-open and eyes-closed condition*

was selected. Baseline cognition (MoCA) was selected a priori. B) Model Selection. The feature sets consisting of EEG alpha (and sub-alpha) powers determined from data exploration (A), along with MoCA selected a priori, were tested using 9 different classifiers using cross-validation (red arrows). The details of the cross-validation is shown in (D). C) After the classifier with the highest f1-score is determined in model selection step, the classifier was re-trained on the entire SAGES Data Set, then the parameters of the classifier were held fixed and independently validated on INTUIT/PRIME Data Set. The results are plotted in Figure 5 and tabulated in Table 3B. D) Details of the cross-validation used in model selection step. The CV performance are plotted in Figure 4 and Table 3A.

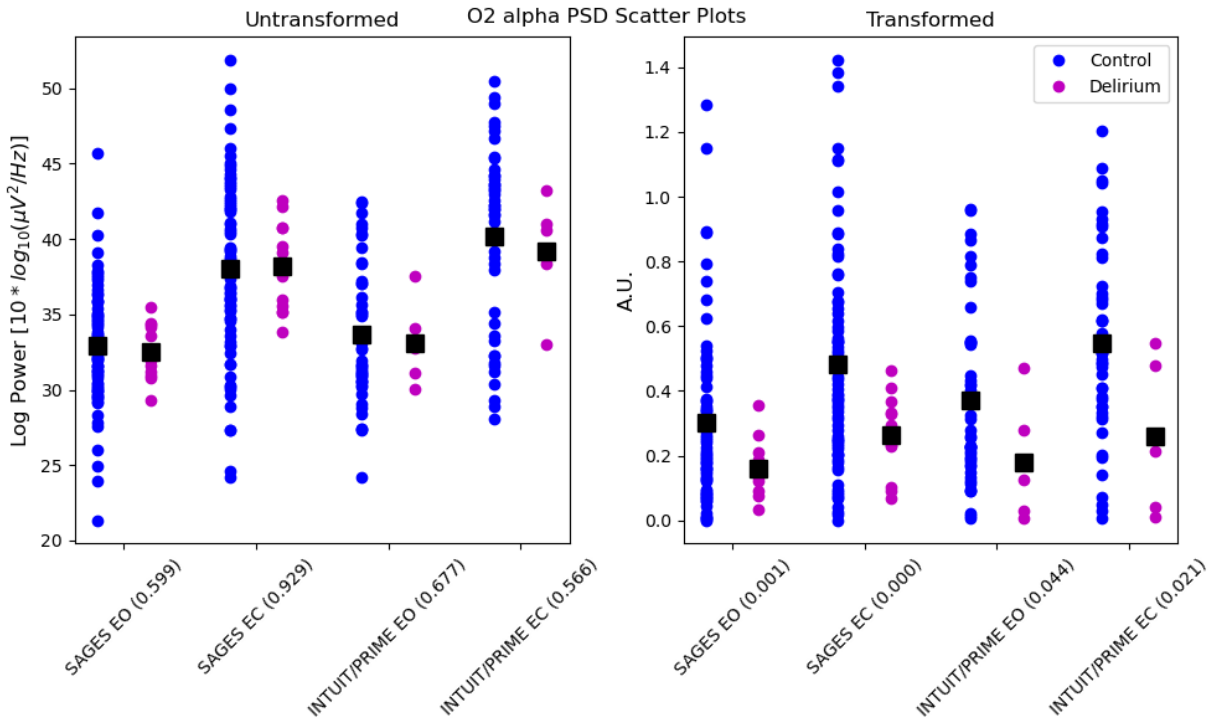

Figure 2 ***Scatter plots of individual EEG alpha powers in O2 channel.*** *Scatter plot showing distributions of alpha powers in participants with post-operative delirium (purple dots) and without delirium (blue dots). Black squares represent the mean of their respective groups. Left panel shows the untransformed alpha powers as seen in PSD plots (Fig. 3) and right panel shows transformed alpha powers. Values enclosed in parentheses in the x axis represent the p-values of two-sample two-tailed Welch's t-test. This channel is representative of all channels in the occipital region and their scatter plots can be found in the Supplementary Fig. S2 & S3. In addition, scatter plots of individuals' sub-alpha powers are shown in Supplementary Fig. S4 & S5.*

The INTUIT/PRIME cohort was used for the model validation (Figure 1C). For the model validation step, the best performing model from the model selection step was chosen based on f1-score, retrained on the entire SAGES Data Set, and model parameters were then fixed and tested on the entire INTUIT/PRIME Data Set. Unlike the model selection step, no cross-validation was done here. Bootstrapping was used to compute the 95% confidence intervals, where the test set (INTUIT/PRIME Data Set) was re-sampled 2000 times, with each re-sampling set stratified to the class proportions of the original sample.

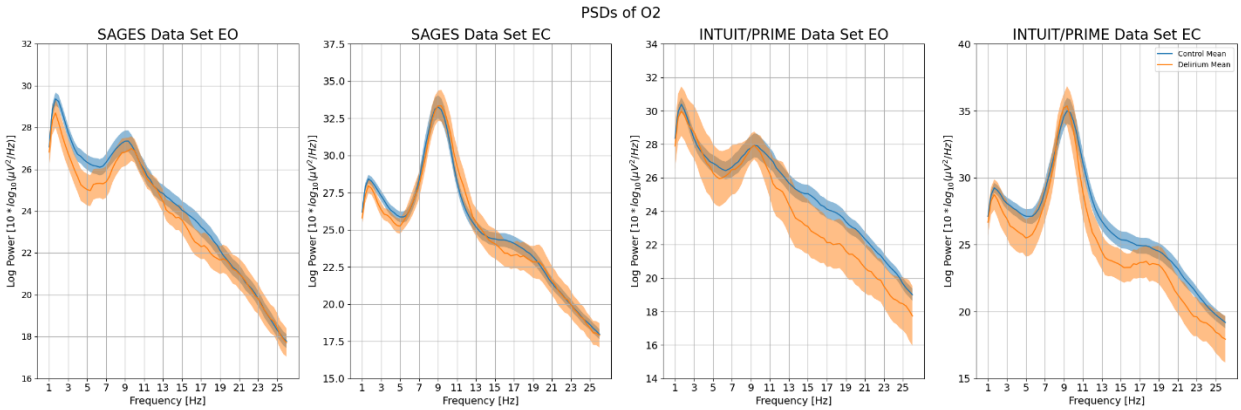

Figure 3 ***Power spectral densities of O2 channel.*** *Power spectral densities of O2 channel from both SAGES and INTUIT/PRIME Data Sets in both eyes-open (EO) and eyes-closed (EC) conditions for [1, 26] Hz frequency range. Shaded regions represent standard error of the means. Blue represents control group and orange represents delirium group. PSDs for O1 and POz are shown in Supplementary Figure S6 & S7, respectively.*

Result

Scatterplots of the untransformed (left panel) and transformed (right panel) resting state alpha powers from the O2 channel are shown in Figure 2. In the untransformed samples, the control group (blue) has similar means but much larger variances than the delirium group (purple) in both data sets and eye conditions, with no significant differences between groups. Despite difference in sample sizes, the variances of both SAGES and INTUIT/PRIME samples are remarkably similar within the same eye conditions and groups. After transformation, all of the group contrasts are significant (two-sample two-tailed Welch's t-test, $\alpha<0.05$). Findings are similar for the resting state alpha powers in the other channels in the occipital region (Supplementary Figure S2 and S3) as well as sub-alpha powers (Supplementary Figure S4 and S5). The PSD plots for the O2 channel in both SAGES and INTUIT/PRIME Data Sets are shown in Figure 3. Remarkably, the alpha peaks are consistent across both data sets within the same eye conditions and groups. However, the distributions of the powers within the delta, theta and beta frequency ranges are not consistent across the data sets within the same eye conditions and groups. Similar findings arise in O1 and POz channels (Supplementary Figure S6 and S7, respectively).

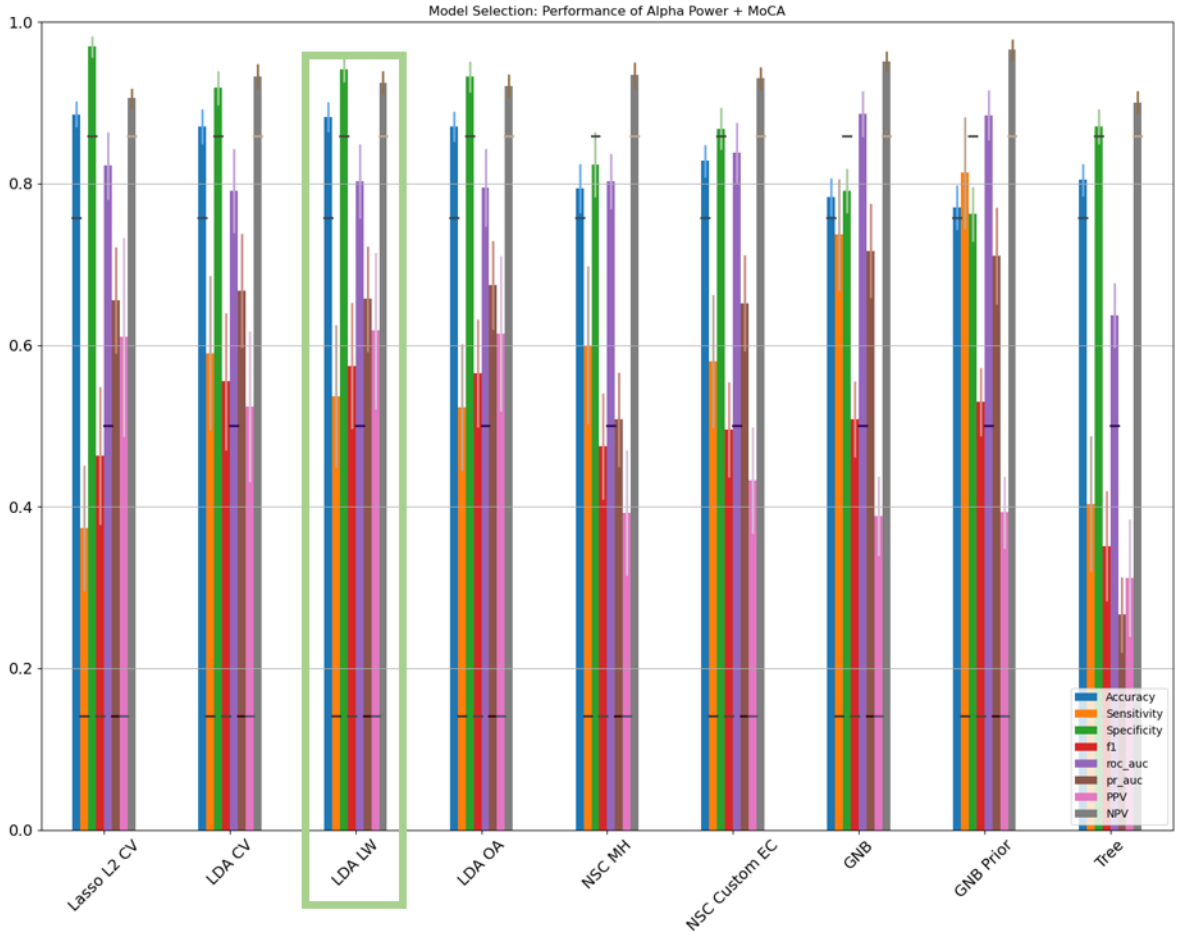

Figure 4 ***Model selection: cross-validation performances using the Principal Feature Set (Alpha Powers + MoCA).*** *The performances of 9 models assessed with 8 different metrics are shown. 95% confidence intervals are shown as thin vertical bars. Chance levels for each metric are shown as dark horizontal lines (details of chance levels in Supplementary Materials). Model with the highest f1-score is enclosed by light green box. Blue: accuracy, orange: sensitivity, green: specificity, red: f1 score, purple: AUC of ROC curve, brown: AUC of precision-recall curve, pink: positive predictive value (PPV, also known as precision), grey: negative predictive value (NPV). Values are tabulated in Table 3A.*

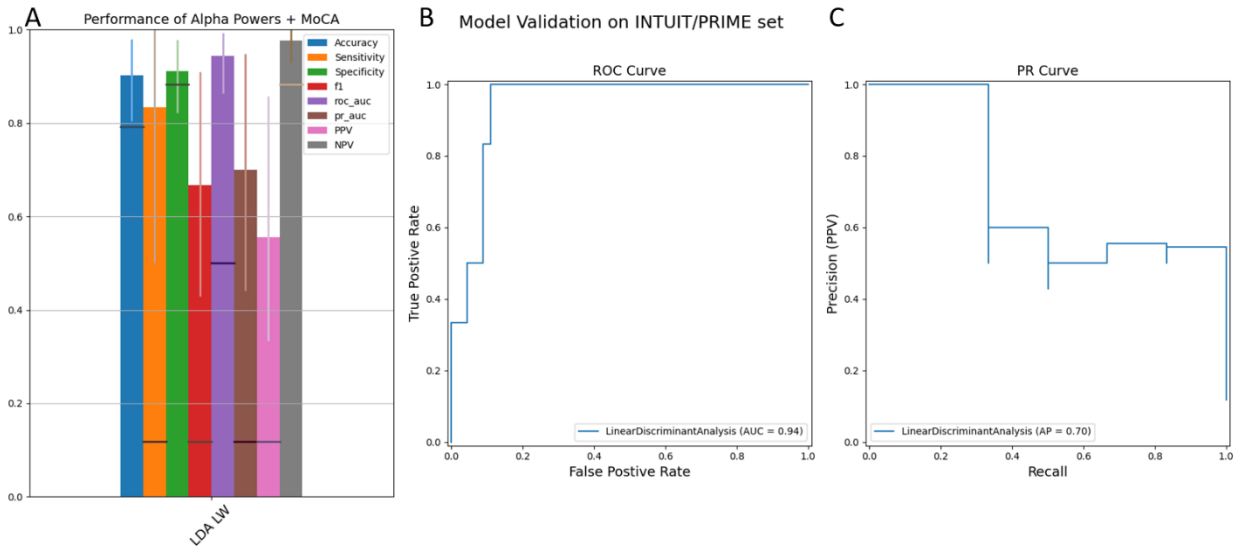

Figure 5 ***Model validation: performances using the Principal Feature Set (Alpha Powers + MoCA).*** *A) the performance of the model (LDA LW) selected from the model selection step. Same legend as in Fig. 3. Values are tabulated in Table 3B. ROC (B) and PR (C) curves are shown for the selected LDA LW model. The AUCs reported as mean and 95% CI are 0.94 [0.86, 0.99] for the ROC curve and 0.70 [0.44, 0.95] for the PR curve.*

The model selection results of the classification performance of the Principal Feature Set cross-validated on SAGES Data Set are shown in Figure 4 as well as in Table 3A. Out of the 9 models tested, LDA LW model has the highest f1-score (mean ± 95% confidence interval: 0.57 ± 0.07, sensitivity: 0.54 ± 0.08, specificity: 0.94 ± 0.02, ROC-AUC: 0.80 ± 0.04). When retraining and then testing the LDA LW model on the entire SAGES Data Set (whole-SAGES model), the f1-score, sensitivity, specificity and ROC-AUC are 0.61, 0.58, 0.95 and 0.76, respectively. For the model validation, when the whole-SAGES LDA LW model is tested on the INTUIT/PRIME Data Set (Figure 5A, Table 3B), the f1-score is 0.67 (95% CI [0.43, 0.92]), sensitivity and specificity are 0.83 (95% CI [0.50, 1.00]) and 0.91 (95% CI [0.82, 0.98]), respectively. The ROC and precision-recall curves are shown in Figure 5B-C. Their AUCs are 0.94, 95% CI [0.87, 0.99] and 0.70, 95% CI [0.46, 0.95], respectively. When compared to MoCA-Alone (Table 4B), the Principal Feature Set outperformed MoCA alone in all metrics except for sensitivity (tied at 0.83). MoCA-Alone yielded sensitivity, specificity and ROC-AUC of 0.83 (95% CI [0.50, 1.00]), 0.80 (95% CI [0.67, 0.91]), 0.91 (95% CI [0.81, 0.98]) respectively. The biggest improvements from MoCA-Alone to the Principal Feature Set lie in PPV (0.20 increase) and sensitivity (0.11 increase). When compared to Alpha-Powers-Alone (Table 4C), Alpha-Powers-Alone has very skewed performance between sensitivity (0.17, 95% CI [0, 0.50]) and specificity (1.00, 95% CI [1.00, 1.00]). Alpha-Powers-Alone's ROC-AUC is 0.77, 95% CI [0.57, 0.94]. The same analyses are done on Sub-Alpha Feature Set, with findings similar to the Principal Feature Set (Table S6, S7, Supplementary Figure S8, S9).

A) SAGES Data Set

Model Name	Accuracy	Sensitivity	Specificity	F1	ROC AUC	PR AUC	PPV	NPV
Lasso L2 CV	0.89 [0.87, 0.90]	0.37 [0.30, 0.45]	0.97 [0.96, 0.98]	0.46 [0.38, 0.55]	0.82 [0.78, 0.86]	0.66 [0.59, 0.72]	0.61 [0.49, 0.73]	0.91 [0.89, 0.92]
LDA CV	0.87 [0.85, 0.89]	0.59 [0.49, 0.69]	0.92 [0.90, 0.94]	0.56 [0.47, 0.64]	0.79 [0.74, 0.84]	0.67 [0.60, 0.74]	0.52 [0.43, 0.62]	0.93 [0.92, 0.95]
LDA LW	0.88 [0.86, 0.90]	0.54 [0.45, 0.62]	0.94 [0.93, 0.96]	0.57 [0.50, 0.65]	0.80 [0.76, 0.85]	0.66 [0.59, 0.72]	0.62 [0.52, 0.71]	0.93 [0.91, 0.94]
LDA OA	0.87 [0.85, 0.89]	0.52 [0.44, 0.60]	0.93 [0.91, 0.95]	0.57 [0.50, 0.63]	0.80 [0.75, 0.84]	0.67 [0.62, 0.73]	0.61 [0.52, 0.71]	0.92 [0.91, 0.94]
NSC MH	0.79 [0.76, 0.82]	0.60 [0.50, 0.70]	0.82 [0.78, 0.86]	0.47 [0.41, 0.54]	0.80 [0.77, 0.84]	0.51 [0.45, 0.57]	0.39 [0.32, 0.47]	0.93 [0.92, 0.95]
NSC EC	0.83 [0.81, 0.85]	0.58 [0.50, 0.66]	0.87 [0.84, 0.89]	0.50 [0.44, 0.55]	0.84 [0.80, 0.88]	0.65 [0.59, 0.71]	0.43 [0.37, 0.50]	0.93 [0.92, 0.94]
GNB	0.78 [0.76, 0.81]	0.74 [0.67, 0.81]	0.79 [0.76, 0.82]	0.51 [0.46, 0.56]	0.89 [0.86, 0.92]	0.72 [0.66, 0.78]	0.39 [0.34, 0.44]	0.95 [0.94, 0.96]
GNB Prior	0.77 [0.74, 0.80]	0.81 [0.74, 0.88]	0.76 [0.73, 0.80]	0.53 [0.49, 0.57]	0.88 [0.85, 0.92]	0.71 [0.65, 0.77]	0.39 [0.35, 0.44]	0.97 [0.95, 0.98]
Tree	0.80 [0.78, 0.82]	0.40 [0.32, 0.49]	0.87 [0.85, 0.89]	0.35 [0.28, 0.42]	0.64 [0.60, 0.68]	0.27 [0.22, 0.31]	0.31 [0.24, 0.38]	0.90 [0.89, 0.91]

B) INTUIT/PRIME Data Set

	Accuracy	Sensitivity	Specificity	F1	ROC AUC	PR AUC	PPV	NPV
LDA LW	0.90 [0.80, 0.98]	0.83 [0.50, 1.00]	0.91 [0.82, 0.98]	0.67 [0.43, 0.91]	0.94 [0.86, 0.99]	0.70 [0.44, 0.95]	0.56 [0.33, 0.86]	0.98 [0.93, 1.00]

Table 3 Performances of Principal Feature Set (Alpha Powers + MoCA). Ranges enclosed in brackets represent 95% confidence intervals. A) Model Selection Results. Green highlighted row represents the model with the highest f1-score. Note: the f1-score is 0.574 for LDA LW and 0.565 for LDA OA. B) Model Validation Results of LDA LW, the model with the highest f1-score.

A) Principal Feature Set (Alpha-Powers + MoCA)

	Accuracy	Sensitivity	Specificity	F1	ROC AUC	PR AUC	PPV	NPV
LDA LW	0.90 [0.80, 0.98]	0.83 [0.50, 1.00]	0.91 [0.82, 0.98]	0.67 [0.43, 0.91]	0.94 [0.86, 0.99]	0.70 [0.44, 0.95]	0.56 [0.33, 0.86]	0.98 [0.93, 1.00]

B) MoCA Alone

	Accuracy	Sensitivity	Specificity	F1	ROC AUC	PR AUC	PPV	NPV
LDA LW	0.80 [0.69, 0.90]	0.83 [0.50, 1.00]	0.80 [0.67, 0.91]	0.50 [0.31, 0.71]	0.91 [0.81, 0.98]	0.54 [0.30, 0.88]	0.36 [0.21, 0.56]	0.97 [0.92, 1.00]

C) Alpha-Powers Alone

	Accuracy	Sensitivity	Specificity	F1	ROC AUC	PR AUC	PPV	NPV
LDA LW	0.90 [0.88, 0.94]	0.17 [0.00, 50]	1.00 [1.00, 1.00]	0.29 [0.00, 0.67]	0.77 [0.57, 0.94]	0.53 [0.21, 0.87]	1.00 [0.00, 1.00]	0.90 [0.88, 0.94]

Table 4 Performances of A) Principal Feature Set (Alpha-Powers + MoCA), B) MoCA-Alone and C) Alpha-Powers-Alone on INTUIT/PRIME Data Set. Ranges enclosed in brackets represent 95% confidence intervals.

Discussion:

Using preoperative rsEEG and baseline cognitive functions collected from 2 independent cohorts of older adults undergoing elective surgeries, we trained and tested machine learning models to assess their performance in predicting post-operative delirium. Specifically, the feature set includes 6 EEG features comprised of resting state alpha power measures from the occipital region in eyes-open and eyes-closed conditions, along with baseline MoCA scores. With this feature set, multiple models were trained on the SAGES Data Set using 5-fold stratified cross-validation. LDA LW, the model with the highest f1-score was selected and retrained on the entire SAGES Data Set and with the model parameters fixed, tested on an independently collected cohort, the INTUIT/PRIME Data Set. In the INTUIT/PRIME Data Set, LDA LW yielded ≥ 0.9 in accuracy, specificity and ROC-AUC and ≥ 0.8 in sensitivity. Importantly, it substantially outperformed MoCA-Alone, with the biggest improvements in PPV (0.20 increase) and sensitivity (0.11 increase). This finding has two major implications: 1) rsEEG can be used in conjunction with the MoCA to preoperatively screen for patients at high risk for post-operative delirium, and 2) occipital alpha powers represent a candidate risk factor that can be studied further to understand the possible neurobiological etiology of delirium[37].

The strong predictive performance seen above can be attributed to the difference in the variances of the resting state occipital alpha and sub-alpha powers. One possible interpretation of the differences in the variances is that individuals within the control group exhibit greater fluctuation in the EEG powers over time relative to the delirium group and the greater fluctuation could be an indirect indication of cortical connectivity and plasticity[38,39,40]. Thus, greater fluctuation over time from one healthy individual could be translated to greater fluctuation across multiple healthy individuals from a fixed time point, relative to individuals with POD. The converse is equally appealing and insightful; limited fluctuation of alpha and sub-alpha powers over time could be an early warning sign and a possible biomarker of impairment in cortical connectivity and plasticity and thus presents greater risk and vulnerability to POD. While the variances of alpha power over time within one resting state recording session didn't show strong predictive performance for delirium (results not shown), we hypothesized that the variance over much longer range of time might better measure the integrity of cortical circuits. This could be tested with repeated measurements from the same individuals. Alternatively, exploring functional connectivity could reveal more direct information about the cortical connectivity[,39,40].

There are only a few studies on the relationship between preoperative EEG and POD and all of them used regressions/statistical tests to establish association. Kim et al.[24] used logistic regression to show association between the median dominant frequency of the prefrontal cortical region (Fp1 and Fp2) and POD. Similarly, while Schuβler et al[,41] recorded EEG using 10-20 international system, they limited their report to F1 and F2 channels. They used statistical tests to show a significant decrease in the power of the high beta and low gamma bands. One study that was done on the same INTUIT/PRIME dataset used here, but analyzed independently of this study, used multivariate logistic and proportion-odds regression analysis to show that preoperative alpha power attenuation, the reduction in alpha power from eyes-closed to eyes-opened condition, is inversely associated with postoperative attention[37]. However, based on our internal testing, the alpha power attenuation, either taken as the difference or ratio between eyes-closed and eyes-opened conditions, doesn't have the same predictive performance of delirium (not shown) as our proposed models. In contrast, one recurring and consistent finding of EEG during episodes of delirium is the alpha slowing in the occipital region,[42,43,44,45,46] suggesting that alpha slowing is a state marker of delirium. The connection between alpha slowing during delirium and restricted range of occipital alpha powers observed preoperatively (and hence before delirium) here is unclear, but one possibility is that the

thalamocortical connectivity responsible for the generation of the alpha rhythm is particularly susceptible to breakdown during stressors; more research is needed to better understand the connection.

While the performance of our optimal machine learning model was quite good in the test INTUIT/PRIME cohort (ROC-AUC > 0.90), there is room for improvement, particularly with regard to sensitivity. As seen in the scatterplots of the resting-state occipital alpha powers in Figure 2, even after the transformation, there remains significant overlaps between control and delirium groups, making perfect or nearly-perfect prediction difficult. Thus, extracting these group-level statistics to make prediction on single individuals remains challenging. Additional features that are generally independent of the resting state occipital alpha and sub-alpha powers could potentially address that pitfall, suggesting that more data exploration is needed. One important class of EEG measures that could potentially fill that gap but weren't explored in this study is functional connectivity.

Improvement in predictive performance of POD over pre-operative cognitive assessment alone is a major finding in this study. Past studies on pre-operative prediction of POD have largely focused on different measures of cognitive functioning, including MoCA[16,17]. While lower cognitive scores have been independently associated with higher risk of developing POD, the relationship and especially the pathophysiology and etiology between cognitive functioning and POD are still not fully understood. Additionally, cognitive functioning is not the sole risk factor for POD[47]. Instead, risk is assumed to be multifactorial but is primarily determined by both patient-related predisposing factors (which include cognitive impairment) and treatment-associated precipitating factors[18,19]. Our analysis suggests that incorporation of neurophysiological features such as resting-state EEG measures improve model performance. This in turn provides support for our conceptual model linking POD to underlying changes in cerebral physiology[21].

One limitation of this study is the difference in the predictive performance in the SAGES and INTUIT/PRIME cohorts, especially in the sensitivity metric: the LDA LW model performed substantially better on the INTUIT/PRIME cohort than on SAGES cohort, even though the training was done on the SAGES cohort. One possible explanation is that the rsEEG for the INTUIT/PRIME cohort was recorded on the same day of the surgery, whereas the rsEEG for the SAGES cohort was recorded at least 2-3 days and up to 2 months in advance of scheduled surgeries. Another factor is that the INTUIT/PRIME cohort originally used MMSE in lieu of MoCA. While crosswalks between MoCA and MMSE are available, this introduced a source of uncertainty. Furthermore, the MMSE to MoCA conversion is harsh. As seen in Supplementary Table S3, a perfect MMSE score of 30 would be penalized during conversion to a MoCA score of 28.5 and a MMSE score of 29 would be converted to a MoCA score of 25.5, which would be considered as possible mild cognitive impairment. Another potential source of variability is that the SAGES cohort used the CAM long form and consensus review for delirium detection, whereas the INTUIT/PRIME cohort used the 3D-CAM approach.

Another limitation is the relatively small sizes of the training set (SAGES) and especially the test set (INTUIT/PRIME). In particular, the low number of participants with POD (n=6) in the INTUIT/PRIME cohort means that one misclassification of the actual delirium case could drastically alter the performance metrics by almost 20%. This is corroborated by rather wide confidence intervals of sensitivity and precision (Table 3B). Thus, a larger sample size is needed to confirm our findings.

In conclusion, machine learning techniques utilizing preoperative rsEEG and MoCA performance features can predict POD with strong performance across multiple metrics. These findings suggest that pre-operative rsEEG can identify subclinical neurophysiological changes that may play a crucial role in the neuropathology of postoperative delirium. Independent confirmation in a larger sample size and with

prospective application is needed to validate our proposals, and to potentially improve model performance by incorporation of features from more frequencies and brain regions. If successful, translation into clinical setting would be convenient and straightforward as EEG is a widely available, inexpensive and well-tolerated neuroimaging technology, and assessment of cognitive performance using the MoCA is relatively fast and simple. The application of such ML approaches to preoperative data thus has the potential to identify individuals at high risk of POD, enabling testing of targeted interventions to reduce the risk and morbidity of this common problem.

Summary:

Previously, we proposed a neurophysiological model of delirium in which delirium is the result of a breakdown of brain functions in individuals with impairments in brain connectivity and plasticity exposed to a stressor, including surgery. Based on this conceptual model, in this study, we applied machine learning techniques to investigate whether accurate individual prediction of POD can be obtained from relatively simple preoperative rsEEG and cognitive performance measures. We also hypothesized that preoperative rsEEG can substantially improve the predictive performance of POD relative to pre-operative cognitive assessments alone. This was evaluated by developing machine learning models on one prospective cohort of 85 older adults (12 cases of delirium) undergoing elective surgery (SAGES). Models were developed using occipital alpha-band power and MoCA task performance as key features. LDA LW, the model with the highest f1-score was selected, retrained on the entire SAGES sample and the model parameters held fixed. That model was subsequently validated on an independent prospective cohort of 51 older adults (6 cases of delirium) undergoing surgery (INTUIT/PRIME). On that validation set, it was able to predict POD with accuracy, ROC-AUC, specificity all ≥ 90% and sensitivity > 80%. With only 7 features, the risk of overfitting remains low. The strong performances seen here are due to the differences in the variances of the alpha powers between control and post-operative delirium groups. Specifically, we showed that while the means of the two groups are similar, the variance of the alpha powers in the control group is much larger than those of post-operative delirium group. While prospective validation with a larger sample is needed, our findings suggest that individuals can be screened for POD risk using simple preoperative measures. Our findings also suggested that occipital alpha powers as a group is a candidate for risk factor that can be studied further for possible neurobiological etiology.

Author Contribution:

Matthew Ning: Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Andrei Rodionov:** Data curation, Formal analysis, Investigation, Methodology, Writing - review & editing. **Jessica M. Ross:** Data curation, Formal analysis, Investigation, Metholodogy, Writing - review & editing. **Recep A. Ozdemir:** Investigation, Metholodogy, Writing - review & editing. **Maja Burch:** Data curation, Writing - review & editing. **Shu Jing Lian:** Data curation, Writing - review & editing. **David Alsop:** Conceptualization, Writing - review & editing. **Michele Cavallari:** Conceptualization, Writing - review & editing. **Bradford C. Dickerson:** Conceptualization, Writing - review & editing. **Tamara G. Fong:** Conceptualization, Writing - review & editing. **Richard N. Jones:** Data curation, Conceptualization, Writing - review & editing. **Towia A. Libermann:** Conceptualization, Writing - review & editing. **Edward R. Marcantonio:** Conceptualization, Supervision, Writing - review & editing. **Emiliano Santarnecchi:** Conceptualization, Writing - review & editing. **Eva M. Schmitt:** Conceptualization, Supervision, Writing - review & editing. **Alexandra Touroutoglou:** Conceptualization, Writing - review & editing. **Thomas G. Travison:** Conceptualization, Writing - review & editing. **Leah Acker:** Data curation, Writing - review & editing. **Melody Reese:** Data curation, Writing - review & editing. **Haoqi Sun:** Investigation,

Methodology, Writing - review & editing. **Brandon Westover:** Investigation, Methodology, Writing - review & editing. **Miles Berger:** Conceptualization, Funding acquisition, Investigation, Resources, Supervision, Writing - review & editing. **Alvaro Pascual-Leone:** Conceptualization, Investigation, Funding acquisition, Resources, Supervision, Writing - review & editing. **Sharon K. Inouye:** Conceptualization, Data curation, Investigation, Resources, Project administration, Supervision, Writing - review & editing. **Mouhsin M. Shafi:** Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing - review & editing.

Conflict of interest statement

Dr. E. Santarnecchi serves on the scientific advisory boards for BottNeuro, which has no overlap with present work; and is listed as an inventor on several issued and pending patents on brain stimulation solutions to diagnose or treat neurodegenerative disorders and brain tumors.

Dr. A. Pascual-Leone is a co-founder of Linus Health and TI Solutions AG which have no overlap with present work. He serves on the scientific advisory boards for the ACE Foundation and the IT'IS Foundation, Neuroelectrics, TetraNeuron, Skin2Neuron, MedRhythms, and Magstim Inc; and is listed as an inventor on several issued and pending patents on the real-time integration of noninvasive brain stimulation with electroencephalography and magnetic resonance imaging, applications of noninvasive brain stimulation in various neurological disorders, as well as digital biomarkers of cognition and digital assessments for early diagnosis of dementia.

Dr. M Berger has received private legal consulting fees related to perioperative neurocognitive disorders.

None of the other authors report any conflicts of interest. All the other co-authors fully disclose they have no financial interests, activities, relationships and affiliations. The other co-authors also declare they have no potential conflicts in the three years prior to submission of this manuscript.

Data Availability

The pre-processed and de-identified EEG data is anticipated to be available on a public repository in late 2025. The link to the repository for the EEG data will be posted on the Github repository linked in the next section.

Code Availability

The source codes (Python and Matlab) are available at the following Github repository: (https://github.com/bacnbs/sages2_repo).

Acknowledgment

The authors gratefully acknowledge the contributions of the patients, family members, nurses, physicians, staff members, and members of the Executive Committee who participated in the Successful Aging after Elective Surgery (SAGES) Study and the patients and family members who participated in the INTUIT/PRIME study.

Funding

This manuscript was funded by grants P01AG031720 (SKI), R33AG071744 (SKI/RNJ), K76-AG057022 (MB), R01- AG073598 (MB), UH2-AG056925 (HEW, CCE) and a NIDUS pilot grant to MMS, MB and BW from R24-AG054259 (SKI).

Dr. Ross was supported during manuscript preparation by the Department of Veterans Affairs Office of Academic Affiliations Advanced Fellowship Program in Mental Illness Research and Treatment, the Medical Research Service of the Veterans Affairs Palo Alto Health Care System and the Department of Veterans Affairs Sierra-Pacific Data Science Fellowship.

Dr. Santarnecchi was partially supported by the NIH (P01 AG031720,) and ADDF (ADDF-FTD GA201902–2017902).

Dr. Inouye holds the Milton and Shirley F. Levy Family Chair at Hebrew SeniorLife/Harvard Medical School, and is supported in part by grants P01AG031720 and R33AG071744 from the National Institutes of Health. She is the Editor in Chief of JAMA Internal Medicine.

Dr. Shafi was partly supported by the Football Players Health Study at Harvard University, and the National Institutes of Health (R01MH115949, R01AG060987, and P01 AG031720.)

Dr. A. Pascual-Leone was partly supported by the National Institutes of Health (R01AG076708, R01AG059089, R03AG072233, and P01 AG031720), the Bright Focus Foundation, and the Barcelona Brain Health Initiative (Institute Guttmann).

Dr. Marcantonio was partially supported by the following grants from the National Institute on Aging (P01 AG031720, R01AG051658, K24 AG035075.

Dr. Berger was partially supported by the National Institutes of Health K76AG057022 and R01AG073598, and received additional supports from the National Institutes of Health (P30-AG028716, P30-AG072958 and UH2-AG056925).

Appendix 1

SAGES II STUDY GROUP

Overall Principal Investigator: Sharon K. Inouye, MD, MPH (Overall PI, Administrative Core, Project 1; HSL, BIDMC, HMS).

Project and Core Leaders: Bradford Dickerson, MD (Project 3; BIDMC, HMS); Richard Jones, ScD (Data Core, Project 4; Brown University); Thomas Travison, PhD (Data Core, HSL, HMS); Edward R. Marcantonio, MD, SM (Overall Co-PI, Epidemiology Core, Project 2; BIDMC, HMS), Towia Libermann, PhD (Project 2, HMS, BIDMC); Alvaro Pascual-Leone, MD, PhD (Project 5, HMS, HSL); Mouhsin M. Shafi, MD, PhD (Project 5, HMS, BIDMC).

Executive Committee: Michele Cavallari, MD, PhD (BWH); Tamara Fong, MD, PhD (HMS, HSL, BIDMC); Eva M. Schmitt, PhD (Overall Project Director, HSL); Emiliano Santarnecchi, PhD (HMS, BIDMC); Alexandra Touroutoglou, PhD (HMS,MGH).

Other Co-investigators: Steven Arnold, MD, (MGH); Simon T. Dillon, PhD (HMS, BIDMC); Charles Guttmann, MD (BWH, HMS); Tammy Hshieh, MD, MPH (BWH); Long Ngo, PhD (HMS, BIDMC); Andrei Rodionov, PhD (BIDMC), Jessica M. Ross, PhD (BIDMC), Sarinnapha Vasunilashorn, PhD (HMS, BIDMC), Pia Webb Kivisakk,M.D.,Ph.D.(MGH, HMS).

Clinical Consensus Panel: Franchesca Arias, PhD (HMS, BIDMC); Eyal Kimchi, MD (HMS, MGH), Jason Strauss, MD (Cambridge Health Alliance); Bonnie Wong, PhD (HMS, MGH).

Surgical and Anesthesia Leaders: Ayesha Abdeen, MD (HMS, BIDMC); Brandon Earp, MD (HMS, BWFH); Lisa Kunze, MD (HMS, BIDMC); Jeffrey Lange, MD (HMS, BWH); Marc Schermerhorn, MD (HMS, BIDMC); David Shaff, MD (HMS, BWFH); Kamen Vlassakov, MD (HMS, BWH).

Epidemiology Core: Sheena Baratono (BIDMC), Rejoice Dhliwayo (BIDMC), Amanda Gallagher (HSL); Maja Burch (BIDMC), Brenna Hagan (BIDMC), Yonah Joffe (HSL), Sofia Kirkman (BIDMC), Shu Jing Lian (BIDMC), Julianna Liu (HSL); Molly Mackler (HSL); Madeleine Martine (HSL); Jacqueline Nee (HSL); Kerry Palihnich (BIDMC), Fotini Papadopoulou (BIDMC), Lauren Phung (BIDMC), Christopher Ramirez, (MGH); Meghan Shanahan (HSL), Louis Shaevel (BIDMC), Bianca Trombetta, (MGH); Stephanie Waldman (BIDMC), Michelle Ward (BIDMC), Guoquan Xu (HSL).

Data Management and Statistical Analysis Core: Fan Chen (HSL) Yun Gou, MA (HSL); Benjamin Helfand, MSc, MD/PhD (University of Massachusetts Medical School); Yoojin Jung, PhD (BIDMC); Douglas Tommet, MPH (Brown University).

Abbreviations: BIDMC, Beth Israel Deaconess Medical Center; BWH, Brigham and Women's Hospital; BWFH, Brigham and Women's Faulkner Hospital; HMS, Harvard Medical School; HSL, Hebrew SeniorLife; MGH, Massachusetts General Hospital; PI, principal investigator.

The INTUIT/PRIME study groups include: Cindy Louise Amundsen, Oke Anakwenze, Harel Anolick, David Attarian, Sandra Au, Chakib Ayoub, Matthew Barber, Rachel Beach, Andrew Berchuck, Dan G. Blazer III, Michael Bolognesi, Rachele Brassard, Brian Brigman, Jeffrey N. Browndyke, W. Michael Bullock, Thomas Bunning, Yee Ching Vanessa Cheong, Soren Christensen, Brian Colin, Cathleen Colon-Emeric, Mitchell Wayne Cox, Thomas D'Amico, Brittany Anne Davidson, James Keith Deorio, Michael J. Devinney, Mark E. Easley, Sarada Eleswarpu, Detlev Erdmann, Mariana Feingold, Michael Ferrandino, Jeffrey Gadsden, Mark Gage, Arun Ganesh, Grant Edward Garrigues, Charles M. Giattino, Rachel Adams Greenup, Ashraf Habib, Ashley Hall, Rhett K. Hallows, David Harpole Jr., Matthew Hartwig, Laura Havrilesky, Courtney Holland, Scott Thomas Hollenbeck, Thomas Hopkins, Edward Ross Houser II, Samuel Huang, Ehimemen Iboaya, Brant Inman, William Jiranek, Russel Kahmke, Amie Kawasaki, Brendan Kelleher, Jay Han Kim, Jacob Klapper, Christopher Klifto, Rebecca Klinger, Stuart Knechtle, Sandhya A. Lagoo-Deenadayalan, Billy Lan, Walter Lee, Howard Levinson, Brian Lewis, Michael Lipkin, Lewis A. Lipsitz, Christopher Mantyh, Hector Martinez-Wilson, John Migaly, Eugene Moretti, Judd Moul, David Murdoch, Thomas L. Novick, Kathryn Odom, Brian Ohlendorf, Steven Olson, Deborah Oyeyemi, Shannon Page, Theodore Pappas, John Park, Andrew Peterson, Andreea Podgoreanu, Thomas J. Polascik, Dana Portenier, Glenn M. Preminger, Rebecca Ann Previs, Edward Nandlal Rampersaud Jr., Kenneth C. Roberts, Cary N. Robertson, Sanziana Alina Roman, Jason Rothman, Aaron Sandler, Siddharth Sata, Charles Scales Jr., Randall Scheri, Thorsten Seyler, Keri Anne Seymour, Nazema Y. Siddiqui, Shayan Smani, Michael Stang, Samuel David Stanley, Katherine Sweeney, Martin V. Taormina, Julie Thacker, Jake Thomas, Betty Tong, Yanne Toulgoat-Dubois, Keith Vandusen, Nathan Waldron, Alison Weidner, Kent Weinhold, Samuel Wellman, Heather E. Whitson, David Williams, Marty G. Woldorff, Megan K. Wong, Mary C. Wright, Rosa Yang, Christopher Young, Sabino Zani, and Mimi Zhang.

Reference

1. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. (American Psychiatric Association Publishing, 2022). doi:10.1176/appi.books.9780890425787.
2. Yan, E. *et al.* Association between postoperative delirium and adverse outcomes in older surgical patients: A systematic review and meta-analysis. *Journal of Clinical Anesthesia* **90**, 111221 (2023).

3. Dolan, M. M. *et al.* Delirium on Hospital Admission in Aged Hip Fracture Patients: Prediction of Mortality and 2-Year Functional Outcomes. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences* **55**, M527–M534 (2000).
4. Ely, E. W. Delirium as a Predictor of Mortality in Mechanically Ventilated Patients in the Intensive Care Unit. *JAMA* **291**, 1753 (2004).
5. Ouimet, S., Kavanagh, B. P., Gottfried, S. B. & Skrobik, Y. Incidence, risk factors and consequences of ICU delirium. *Intensive Care Med* **33**, 66–73 (2007).
6. Ní Chróinín, D. *et al.* Older trauma patients are at high risk of delirium, especially those with underlying dementia or baseline frailty. *Trauma Surg Acute Care Open* **6**, e000639 (2021).
7. Sanchez, D. *et al.* Frailty, delirium and hospital mortality of older adults admitted to intensive care: the Delirium (Deli) in ICU study. *Crit Care* **24**, 609 (2020).
8. Gou, R. Y. *et al.* One-Year Medicare Costs Associated With Delirium in Older Patients Undergoing Major Elective Surgery. *JAMA Surg* **156**, 462 (2021).
9. Tao, M. *et al.* Efficacy of transcranial direct current stimulation on postoperative delirium in elderly patients undergoing lower limb major arthroplasty: A randomized controlled trial. *Brain Stimulation* **16**, 88–96 (2023).
10. Weinstein, S. M. *et al.* Postoperative delirium in total knee and hip arthroplasty patients: a study of perioperative modifiable risk factors. *British Journal of Anaesthesia* **120**, 999–1008 (2018).
11. Subramaniam, B. *et al.* Effect of Intravenous Acetaminophen vs Placebo Combined With Propofol or Dexmedetomidine on Postoperative Delirium Among Older Patients Following Cardiac Surgery: The DEXACET Randomized Clinical Trial. *JAMA* **321**, 686 (2019).
12. Marra, A., Ely, E. W., Pandharipande, P. P. & Patel, M. B. The ABCDEF Bundle in Critical Care. *Critical Care Clinics* **33**, 225–243 (2017).
13. Inouye, S. K. *et al.* A Multicomponent Intervention to Prevent Delirium in Hospitalized Older Patients. *N Engl J Med* **340**, 669–676 (1999).
14. Reuben, D. B. *et al.* MODELS OF GERIATRICS PRACTICE; The Hospital Elder Life Program: A Model of Care to Prevent Cognitive and Functional Decline in Older Hospitalized Patients. *J American Geriatrics Society* **48**, 1697–1706 (2000).
15. Chen, C. C.-H. *et al.* Effect of a Modified Hospital Elder Life Program on Delirium and Length of Hospital Stay in Patients Undergoing Abdominal Surgery: A Cluster Randomized Clinical Trial. *JAMA Surg* **152**, 827 (2017).
16. Wong, C. K. *et al.* Head-to-head comparison of 14 prediction models for postoperative delirium in elderly non-ICU patients: an external validation study. *BMJ Open* **12**, e054023 (2022).
17. Cizginer, S. *et al.* The Cognitive Reserve Model in the Development of Delirium: The Successful Aging After Elective Surgery Study. *J Geriatr Psychiatry Neurol* **30**, 337–345 (2017).
18. Inouye, S. K. Predisposing and Precipitating Factors for Delirium in Hospitalized Older Patients. *Dement Geriatr Cogn Disord* **10**, 393–400 (1999).
19. Bellelli, G., Brathwaite, J. S. & Mazzola, P. Delirium: A Marker of Vulnerability in Older People. *Front. Aging Neurosci.* **13**, 626127 (2021).
20. Giattino, C. M. *et al.* Intraoperative Frontal Alpha-Band Power Correlates with Preoperative Neurocognitive Function in Older Adults. *Front. Syst. Neurosci.* **11**, 24 (2017).
21. Shafi, M. M. *et al.* Advancing the Neurophysiological Understanding of Delirium. *J American Geriatrics Society* **65**, 1114–1118 (2017).
22. Cooter Wright, M. *et al.* A Processed Electroencephalogram–Based Brain Anesthetic Resistance Index Is Associated With Postoperative Delirium in Older Adults: A Dual Center Study. *Anesthesia & Analgesia* **134**, 149–158 (2022).
23. Ross, J. M. *et al.* Neurophysiologic predictors of individual risk for POST-OPERATIVE delirium after elective surgery. *J American Geriatrics Society* **71**, 235–244 (2023).
24. Kim, J. *et al.* Resting-state prefrontal EEG biomarker in correlation with postoperative delirium in elderly patients. *Front. Aging Neurosci.* **15**, 1224264 (2023).

25. Ward, M. *et al.* Successful aging after elective surgery II : Study cohort description. *J American Geriatrics Society* **72**, 209–218 (2024).
26. Berger, M. *et al.* The INTUIT Study: Investigating Neuroinflammation Underlying Postoperative Cognitive Dysfunction. *J American Geriatrics Society* **67**, 794–798 (2019).
27. Whitson, H. E. *et al.* A template for physical resilience research in older adults: Methods of the PRIME-KNEE study. *J American Geriatrics Society* **69**, 3232–3241 (2021).
28. Inouye, S. K. Clarifying Confusion: The Confusion Assessment Method: A New Method for Detection of Delirium. *Ann Intern Med* **113**, 941 (1990).
29. Nasreddine, Z. S. *et al.* The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment. *J American Geriatrics Society* **53**, 695–699 (2005).
30. Marcantonio, E. R. *et al.* 3D-CAM: Derivation and Validation of a 3-Minute Diagnostic Interview for CAM-Defined Delirium: A Cross-sectional Diagnostic Test Study. *Ann Intern Med* **161**, 554 (2014).
31. Vasunilashorn, S. M. *et al.* A New Severity Scoring Scale for the 3-Minute Confusion Assessment Method (3D-CAM). *J American Geriatrics Society* **68**, 1874–1876 (2020).
32. David-Bercholz, J. *et al.* Conserved YKL-40 changes in mice and humans after postoperative delirium. *Brain, Behavior, & Immunity - Health* **26**, 100555 (2022).
33. Saczynski, J. S. *et al.* The Montreal Cognitive Assessment: Creating a Crosswalk with the Mini-Mental State Examination. *J American Geriatrics Society* **63**, 2370–2374 (2015).
34. Woldorff, M. G. *et al.* The temporal dynamics of the effects in occipital cortex of visual-spatial selective attention. *Cognitive Brain Research* **15**, 1–15 (2002).
35. Rogasch, N. C. *et al.* Removing artefacts from TMS-EEG recordings using independent component analysis: Importance for assessing prefrontal and motor cortex network properties. *NeuroImage* **101**, 425–439 (2014).
36. Mutanen, T. P., Biabani, M., Sarvas, J., Ilmoniemi, R. J. & Rogasch, N. C. Source-based artifact-rejection techniques available in TESA, an open-source TMS–EEG toolbox. *Brain Stimulation* **13**, 1349–1351 (2020).
37. Acker, L. *et al.* Preoperative electroencephalographic alpha-power changes with eyes opening are associated with postoperative attention impairment and inattention-related delirium severity. *British Journal of Anaesthesia* **132**, 154–163 (2024).
38. Deco, G., Jirsa, V. K. & McIntosh, A. R. Emerging concepts for the dynamical organization of resting-state activity in the brain. *Nat Rev Neurosci* **12**, 43–56 (2011).
39. McIntosh, A. R., Kovacevic, N. & Itier, R. J. Increased Brain Signal Variability Accompanies Lower Behavioral Variability in Development. *PLoS Comput Biol* **4**, e1000106 (2008).
40. Woltering, S., Jung, J., Liu, Z. & Tannock, R. Resting state EEG oscillatory power differences in ADHD college students and their peers. *Behav Brain Funct* **8**, 60 (2012).
41. Schüßler, J. *et al.* Preoperative characterization of baseline EEG recordings for risk stratification of post-anesthesia care unit delirium. *Journal of Clinical Anesthesia* **86**, 111058 (2023).
42. Palanca, B. J. A., Wildes, T. S., Ju, Y. S., Ching, S. & Avidan, M. S. Electroencephalography and delirium in the postoperative period. *British Journal of Anaesthesia* **119**, 294–307 (2017).
43. Kinoshita, H. *et al.* The Perioperative Frontal Relative Ratio of the Alpha Power of Electroencephalography for Predicting Postoperative Delirium After Highly Invasive Surgery: A Prospective Observational Study. *Anesthesia & Analgesia* **137**, 1279–1288 (2023).
44. Guay, C. S. *et al.* Postoperative Delirium Severity and Recovery Correlate With Electroencephalogram Spectral Features. *Anesthesia & Analgesia* **136**, 140–151 (2023).
45. Gutierrez, R. *et al.* Intraoperative Low Alpha Power in the Electroencephalogram Is Associated With Postoperative Subsyndromal Delirium. *Front. Syst. Neurosci.* **13**, 56 (2019).
46. Wiegand, T. L. T., Rémi, J. & Dimitriadis, K. Electroencephalography in delirium assessment: a scoping review. *BMC Neurol* **22**, 86 (2022).

47. Menzenbach, J. *et al.* PRe-Operative Prediction of postoperative DElirium by appropriate SCreening (PROPDESC) development and validation of a pragmatic POD risk screening score based on routine preoperative data. *Journal of Clinical Anesthesia* **78**, 110684 (2022).