Clinical Neurophysiology 153 (2023) 79–87

Contents lists available at ScienceDirect

Clinical Neurophysiology

journal homepage: www.elsevier.com/locate/clinph

Sensor-level MEG combined with machine learning yields robust classification of mild traumatic brain injury patients

Juho Aaltonen [a,b,*], Verna Heikkinen [a,b], Hanna Kaltiainen [b,c], Riitta Salmelin [b], Hanna Renvall [a,b,c]

[a] BioMag Laboratory, HUS Medical Imaging Center, Helsinki University Hospital, Helsinki University and Aalto University School of Science, P.O. Box 340, 00029 HUS Helsinki, Finland
[b] Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Aalto University, P.O. Box 12200, 00760 AALTO, Finland
[c] Department of Neurology, Helsinki University Hospital and Clinical Neurosciences, Neurology, University of Helsinki, P.O. Box 340, 00029 HUS, Helsinki, Finland

HIGHLIGHTS

- Machine learning classified mild traumatic brain injury patients based on MEG data.
- The results were consistent across two independent data sets and sites.
- Sensor-level MEG power spectra provide a feasible measure for clinical follow-up.

ARTICLE INFO

Article history:
Accepted 9 June 2023
Available online 30 June 2023

Keywords:
Mild traumatic brain injury
Magnetoencephalography
Resting-state
Machine learning

ABSTRACT

Objective: Diagnosis of mild traumatic brain injury (mTBI) is challenging despite its high incidence, due to the unspecificity and variety of symptoms and the frequent lack of structural imaging findings. There is a need for reliable and simple-to-use diagnostic tools that would be feasible across sites and patient populations.
Methods: We evaluated linear machine learning (ML) methods' ability to separate mTBI patients from healthy controls, based on their sensor-level magnetoencephalographic (MEG) power spectra in the subacute phase (<2 months) after a head trauma. We recorded resting-state MEG data from 25 patients and 25 age-sex matched controls and utilized a previously collected data set of 20 patients and 20 controls from a different site. The data sets were analyzed separately with three ML methods.
Results: The median classification accuracies varied between 80 and 95%, without significant differences between the applied ML methods or data sets. The classification accuracies were significantly higher with ML than with traditional sensor-level MEG analysis based on detecting pathological low-frequency activity.
Conclusions: Easily applicable linear ML methods provide reliable and replicable classification of mTBI patients using sensor-level MEG data.
Significance: Power spectral estimates combined with ML can classify mTBI patients with high accuracy and have high promise for clinical use.

1. Introduction

Despite the high incidence of mild traumatic brain injuries (mTBI) worldwide (Gardner and Yaffe, 215)), they continue to remain under- and misdiagnosed (Pozzato et al., 2017). Evaluation of mTBI patients is hampered by many confounding factors (Losoi et al., 2016; Iverson et al., 2017), and the prevalence of patients' cognitive complaints does not correlate with possible findings in mere structural imaging (Lee et al., 2008; Jacobs et al., 2010). There is a strong need for reliable, objective biomarkers of neuronal change in mTBI compared with healthy controls, for pinpointing those patients who would most likely benefit from close follow-up during the recovery.

Magnetoencephalography (MEG) and electroencephalography (EEG) directly assess brain electrophysiology, and they can therefore provide means to investigate functional changes at the individual level. Indeed, resting-state MEG and EEG measurements after mTBI have demonstrated increased low-frequency (<7 Hz)

Abbreviations: ML, machine learning; LDA, linear discriminant analysis; SVM, support vector machine; LR, logistic regression.

* Corresponding author at: BioMag Laboratory, HUS Medical imaging Center, Helsinki University Hospital, Haartmaninkatu, 00290 Helsinki, Finland.
E-mail address: juho.aaltonen@aalto.fi (J. Aaltonen).

https://doi.org/10.1016/j.clinph.2023.06.010
1388-2457/

activity (Lewine et al., 1999; Lewine et al., 2007; Huang et al., 2009; Huang et al., 2014), decreased power and peak frequency in the so-called alpha frequency range (8–14 Hz) (Korn et al., 2005; Dunkley, 2015; Koufen and Dichgans, 1978; Tebano et al., 1988), and altered beta-range (14–30 Hz) functional connectivity (Zhang et al., 2020). By themselves such alterations are, however, often subtle, and barely distinguishable from the variation within normal population (Nuwer et al., 2005). The observed changes in oscillatory brain activity seem to normalize during recovery (Koufen and Dichgans, 1978; Kaltiainen et al., 2018), although some residual abnormalities may persist for months or years even after mild injury (Lewine et al., 1999; Lewine et al., 2007; Huang et al., 2009; Kaltiainen et al., 2018).

Oscillatory brain activity at different frequency bands appears to provide promising features for differentiating mTBI patients from control subjects, but their overall effect has rarely been addressed. The utilization of machine learning (ML) techniques for high-dimensional brain imaging data is increasing and may provide such an approach. Using ML combined with connectivity analysis over a wide frequency range, mTBI patients measured with MEG within 24 hours after the trauma were correctly classified in over 90% of cases (Antonakakis et al., 2016; Antonakakis et al., 2017). Recently, Thorpe et al. (2020) differentiated symptomatic mTBI patients from controls with 100% accuracy using connectivity analysis coupled with linear discriminant analysis (LDA) at 7 to 282 months post-injury. Here the results stemmed from rather complicated connectivity analyses which are not easily transformable into clinical use. In their recent review on MEG abnormalities after mTBI, Allen and colleagues highlighted two potential MEG biomarkers, *i.e.*, increased low-frequency power and connectivity alterations across different frequency bands. They also underlined the urgent need for simple-to-use analysis pipelines that would be replicable across different sites, scanners, and patient populations (Allen et al., 2021).

In the present study, we hypothesized that resting-state MEG measurements conducted in sub-acute stage (<2 months) after the trauma, combined with ML techniques over a wide frequency range on sensor-level power spectra, would enable superior accuracy over traditional sensor-level based MEG analysis (Kaltiainen et al., 2018) in differentiating mTBI and healthy control subjects from each other. We focused our analysis on sub-acute mTBI patients rather than chronic mTBI, as they are more likely to have changes in the oscillatory brain activity. We used three different linear classification methods - LDA, support vector machine (SVM), and logistic regression (LR): We hypothesized that with the multi-channel coverage provided by a state-of-the-art MEG device, the classification accuracy would not be highly dependent on the exact ML approach, and would outperform traditional, visual inspection based analysis even for sensor-level data. To confirm the reliability and replicability of the results over different patient populations, we applied the same analysis pipeline on two independent data sets, recorded at two different sites, and compared the results with the traditional sensor-level analysis (Kaltiainen et al., 2018).

2. Material and methods

2.1. Subjects

Altogether 45 patients (41 ± 2 years; average ± standard error of the mean, SEM) who suffered from mTBI participated in the study; 22 were females (41 ± 3 years), and 23 males (41 ± 3 years). Patients had no earlier history of traumatic brain injury, neurological or neuropsychological disorders, nor substance abuse. They were all without medication affecting the central nervous system. All patients fulfilled the American Congress of Rehabilitation Medicine (ACRM) criteria for mild traumatic brain injury: loss of consciousness (LOC) ≤ 30 min, post-traumatic amnesia (PTA) ≤ 24 h and 13 to 15 points on the Glasgow Coma Scale evaluation (GCS) 30 min after the head trauma (American Congress of Rehabilitation Medicine Committee, 1993). GCS varies between 3 (deep unconsciousness) and 15 (alert and oriented) points (Teasdale and Jennett, 1974). The predominant classes of injury were bike accident (16 subjects), fall (13), moving vehicle accident (9), sports accident (6), and hit to head (1). All patients underwent MEG measurement within two months after mTBI (6–60 days, average 30 days).

The control group consisted of 45 healthy adults (42 ± 2 years), of whom 22 were females (40 ± 3 years) and 23 males (43 ± 2 years). None of the control subjects had a history of head trauma, neurological or neuropsychological disorders, or hobbies with high susceptibility to brain trauma. All patients and control subjects gave their informed consent to participate in the study, and the study was accepted by the Ethics Committee of Helsinki and Uusimaa Hospital District. This study was carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.

2.2. MEG recordings

Magnetoencephalography recordings were conducted at two sites. The data of twenty patients (42 ± 3 years; 8 females and 12 males) from our previous study (Kaltiainen et al., 2018) and 20 age-matched control subjects (42 ± 3 years; 8 females and 12 males) from earlier studies (Kaltiainen et al., 2016; Kaltiainen et al., 2018; Renvall et al., 2012) was recorded in Aalto NeuroImaging MEG Core, Espoo, Finland, in a magnetically shielded room with a 306-channel whole-head MEG device (Elektra Neuromag™, Elekta Oy, Helsinki, Finland). Data collection from another, previously unpublished, set of 25 patients (41 ± 2 years; 14 females and 11 males) and 25 controls (41 ± 3 years; 14 females and 11 males) was conducted at BioMag Laboratory, Helsinki, Finland, with a newer but otherwise similar 306-channel whole-head MEG device (Elektra Neuromag TRIUX, Megin Oy, Helsinki, Finland). Measurement environments and noise levels, based on frequently conducted empty room measurements and subsequent analysis of channel-level noise, were generally alike between the two sites. Every subject filled the Rivermead Post-Concussion Symptom Questionnaire (RMPCQ). In the following, the data sets recorded at Aalto NeuroImaging MEG Core and BioMag Laboratory are referred to as data sets A and B, respectively. Patient and control demographics are listed in Table 1.

Both MEG devices consist of triplet sensors of two planar gradiometers and one magnetometer at 102 locations, with each gradio/magnetometer coupled to a Superconducting Quantum Interference Device (SQUID). Four to five head-position-indicator (HPI) coils (the standard number of HPI coils changed during the course of the study) were attached to the scalp to enable continuous head position monitoring during the measurement. The head coordinate system was first determined by measuring the locations of the nasion and two preauricular points with a 3D digitizer (Polhemus 3Space® Fastrak™, Colchester, VT, US). The HPI coil locations were then recorded in this head coordinate system. Horizontal and vertical electro-oculogram (EOG) was measured for eye-movement-related artifact removal, and electrocardiogram was available for most of the patients and controls.

We measured spontaneous resting-state activity with eyes closed (EC) and eyes open (EO) from all subjects (3 minutes per condition for the control subjects in data set A, 10 minutes per condition for the patients in data set A, and 5 min per condition for patients and controls in data set B). The MEG signals were band-pass filtered at 0.03–330 Hz and digitized at 1000 Hz, except for

Table 1
Patient and control demographics for both data sets. Age, gender, timing of the magnetoencephalography (MEG) measurement with respect to the trauma (w = week, m = month), possible detectable structural lesion in magnetic resonance image (MRI)/computed tomography scan (CT), and RMPCQ score of subjects in both data sets. RMPCQ = Rivermead Post-Concussion Symptoms Questionnaire; NA = not applicable.

	DATA SET A		DATA SET B	
	Patients	**Controls**	**Patients**	**Controls**
AGE (AVE ± SEM)	42.3 ± 3	42.0 ± 3	40.5 ± 2	41.3 ± 3
GENDER (F/M)	8/12	8/12	14/11	14/11
MEG (≤1W/≤ 1 M/≤ 2 M)	4/9/7	NA	0/12/13	NA
MRI/CT LESION (YES/NO)	12/8	NA	9/16	NA
RMPCQ (AVE ± SD)	16.6 ± 12	NA	16.0 ± 11	NA

the controls in the previous studies (Kaltiainen et al., 2016; Renvall et al., 2012) in which the recording passband was 0.03–200 Hz and sampling frequency 600 Hz.

2.3. MEG preprocessing

We downsampled all the data to 200 Hz to enable comparison between the data sets. The Signal Space Separation (SSS) method (Taulu and Simola, 2006) was applied to exclude external artifacts, and all subjects' measurements were transferred to the same head position using an SSS-based head transformation algorithm (Taulu et al., 2004) in MaxFilter software (Elektra Oy, Helsinki, Finland).

Subsequently, the data was processed using MNE Python software (Gramfort et al., 2013). A high-pass filter of 1 Hz was applied to the data. Independent component analysis (ICA; Hyvärinen and Oja, 2000) was used to identify and remove 1–2 of the most prominent components related to eye blinks or eye movements, as well as cardiac signal (QRS complex). The data was visually inspected to ensure successful artifact removal. Subsequently, a 3-min long time window from both experimental conditions (EC and EO) was used for analysis for each subject. This time window was chosen, as only 3 min of overall measurement data was available for some of the control subjects. The preprocessed data was further divided into three 60-sec time windows. We discarded 20 s at the beginning and end of each data set, and used a maximum of 20 sec (33%) overlap with the previous/following window for all data sets.

The 60-sec long time series at each MEG sensor were converted to frequency space by calculating the power spectral density (PSD) from 1 Hz to 100 Hz with the Welch's method using 2048-point Fast Fourier Transform, 50% overlap, and Hanning windowing in the *Scipy Python* package (Virtanen et al., 2020). The PSDs were converted to decibel scale (dB) to emphasize minor amplitude differences in the data. The processed PSDs were further divided into 21 frequency bands, starting from 1-3 Hz and widening linearly up to 81.8–87.8 Hz; the average of PSD within a given frequency range was used in the subsequent analysis. We disregarded the band with 50 Hz power-line interference.

2.4. Classification of the MEG data

We analyzed the data from the two measurement environments separately to avoid any possible bias related to subtle differences in the noise levels between devices or environments. As gradiometers and magnetometers give qualitatively different (although correlated) data, the sensor types were first included both separately and then together in the analysis. Thus we had a total of 2142, 4284, or 6428 spatial-frequency features, *i.e.*, 21 frequency bands on 102 (magnetometers), 204 (gradiometers), or 306 (both) sensors per subject.

We compared three different widely used linear classifiers implemented in the *Scikit-learn Python* package (Pedregosa et al., 2011). In Linear Discriminant Analysis (LDA), a set of uncorrelated linear functions, known as discriminant functions, that maximize the distance between the mean vectors of classes of objects are learnt from the training data and applied on the test data. Support Vector Machines (SVM) aim to maximize the margin of the hyperplane that separates two classes in a linear or non-linear manner, while logistic regression (LR) predicts the probability of a binary target variable using logit transformation of the data. On LDA, prior probabilities between subject groups were set equal, and least-squares solution, optimal for binary classification task, combined with Oracle Shrinkage Approximating (OAS) covariance estimator was used as a solver (Chen et al., 2010). OAS and the least-squares solver allow the data to diverge from Gaussian distribution. For SVM, we used a linear kernel with a regularization parameter of 0.025. On LR, maximum iterations were raised to 2000 to ensure proper convergence.

Each classifier was evaluated through leave-one-out-cross-validation (LOOCV) protocol. In LOOCV, one subject represents the test group at each time, while the other subjects form the training group. The number of true (T) and false (F) positive (P) and negative (N) cases (TP/N and FP/N) were evaluated and used to compute the classifier's accuracy, and the results were summarized visually with receiver operating characteristic (ROC) curves.

The training data were normalized so that each spatial-frequency feature, *i.e.,* frequency band per channel over the subjects had zero mean and one standard deviation (SD). Subsequently, the normalization was applied to the test data. The data set for each LOOCV evaluation was formed by using one randomly chosen power spectrum (*i.e.,* one 60-sec interval) per each subject. This procedure was repeated 50 times for each classifier to decrease the possibility of selection bias. The final prediction label was assigned based on the repeated runs.

2.5. Statistical analysis

To examine the effect of each classifier (LDA, SVM, LR), experimental condition (EC, EO) and sensor type (planar gradiometer, magnetometer) on the accuracy results, we fitted a binomial mixed-effects model. Here subjects, using the within-subject accuracies over 50 repetitions per condition (see above), were treated as random effects and experimental factors as fixed effects. The significance of each factor was determined using the likelihood ratio test. The same binomial mixed-effect model was used for comparing the effect of classifier and experimental condition between the two data sets. The statistical analysis was conducted using R programming language (R Core Team, 2020). The significance level was set to 0.01.

Mann-Whitney U test was used for comparing the RMPCQ scores between patients classified correctly and incorrectly within the data sets. Magnetic resonance images (MRI), available for all patients, were used for comparing whether the presence of structural trauma lesions (detectable vs. non-detectable lesion) correlated with the patients' average classification accuracy, using Chi-square statistics and Pearson correlation coefficient.

2.6. Comparison with traditional sensor-level analysis

In addition to the ML analyses above, sensor-level analysis was conducted in all patients similarly to Kaltiainen et al. (2018). Each patients' power spectra were compared with the average spectra calculated over a control group of 139 healthy adults (Renvall et al., 2012): Spectral power exceeding that of the control subjects by 2 SD in the low-frequency range (3–7 Hz) of the EC condition was considered pathological. Fig. 1 depicts an example of EC condition on a demonstrative channel in one patient, compared with the mean (+2 SD) activity over the control group. We then studied whether ML-based analyses recognized more patients than traditional analysis using binomial test. Chi-square statistics were used to evaluate if the classification using traditional analysis correlated with the ML results.

2.7. Comparison of narrow vs. wide band activity

Previous electrophysiological studies have emphasized the presence of low-frequency activity (<7 Hz) as a biomarker for mTBI (Lewine et al., 1999; Lewine et al., 2007; Huang et al., 2009; Huang et al., 2014; Kaltiainen et al., 2018), but higher frequencies may also show changes after brain injury (Korn et al., 2005; Dunkley, 2015; Koufen and Dichgans, 1978; Tebano et al., 1988; Zhang et al., 2020). Therefore, we compared the performance of the present classifiers when the full power spectrum (1–87.8 Hz) was included *vs.* when the analysis was limited to low frequencies (1–8 Hz) and to the alpha–beta band (~8–30 Hz); the Wilcoxon signed-rank test was applied on the individual classification accuracies.

The ML methods applied do not give direct information about which frequency bands are the most significant for the classification performance. The contribution of different bands was tested by performing the above analysis by adding one band at a time up to the full 21 bands (1–87.8 Hz) in a random order 50 times.

3. Results

3.1. Effect of sensor type, experimental condition and classifier

Fig. 2 demonstrates the distribution of discriminant function values for mTBI patients and control subjects for both data sets in the EC condition when using gradiometer data and LDA classifier (50 repetitions per subject). Fig. 3 and Table 2 depict the classification results for all three classifiers, separately in each experimental condition and for each data set.

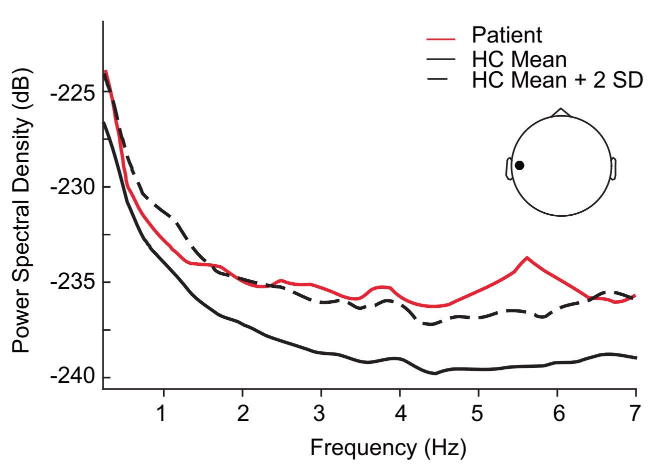

Fig. 1. Example of one patient (from data set A) exhibiting excessive low-frequency activity compared with a healthy control data set (HC, N = 139), peaking at 5–6 Hz over the left temporal area. The insert refers to the location of the depicted channel.

The experimental condition was a major factor for the classification accuracy: The probability for the subjects to be classified correctly was significantly higher in the EC than EO condition (data set A: $Z = -4.60$, $P < 0.001$; data set B: $Z = -3.80$, $P < 0.001$). However, the effect of the classifier was not significant (likelihood test, data set A: $\chi^2 = 0.723$, $P = 0.70$; data set B: $\chi^2 = 2.58$, $P = 0.28$), and there was no significant interaction between the experimental condition and classifier (data set A: $\chi^2 = 0.064$, $P = 1.0$; data set B: $\chi^2 = 0.53$, $P = 0.77$). The classification accuracies of the data sets did not differ from each other in this respect (effect of experimental condition $\chi^2 = 0.94$, $P = 0.33$; classifier $\chi^2 = 2.14$, $P = 0.34$). The mean classification accuracies varied from 63% (median 80%) to 67% (median 95%) for data set A, and from 63% (median 85%) to 66% (median 88%) for data set B when using gradiometer data on the EC condition.

When we compared the classification accuracies with and without including the sensor type in the model, the difference was not significant ($P > 0.1$), suggesting that the choice of sensor combination did not play a significant role in the results. As the responses over the sensors and of their combination are, by nature, highly correlated, the rest of the analysis was applied on gradiometer data alone, taken the generally lower noise-sensitivity of gradiometers compared with magnetometers.

3.2. Diagnostic value of the ML methods

The diagnostic ability of the applied ML methods was moderate, as shown by the receiver operating characteristic (ROC) curves in Fig. 4. The area under the curve values (AUCs) varied between 0.6 and 0.75. For data set A, the LDA classifier had the highest AUC of 0.74, whereas for data set B the corresponding AUC was 0.65. Notably, the algorithms performed very similarly within both data sets, and the AUC values did not differ significantly from each other between the data sets for any of the ML classifiers ($Z_{LDA} = 1.233$, $P = 0.11$; $Z_{LR} = 0.852$, $P = 0.20$; $Z_{SVM} = 0.255$, $P = 0.40$).

3.3. Comparison of the ML-based MEG results with the patients' behavioral and structural data

The RMPCQ scores of the patients varied from 1 to 36 (17 ± 3; ave ± SEM, data set A), and from 0 to 38 (16 ± 2, data set B): the most prominent symptoms reported by the subjects were fatigue (A: 16/20 subjects, B: 22/25 subjects), poor concentration (A: 16/20, B: 17/25), and forgetfulness (A: 12/20, B: 18/25). The RMPCQ score was not a significant discriminating factor between the patients classified correctly and incorrectly in either of the data sets: For example, in the EC condition using the LDA classifier, the average RMPCQ scores were 17 (A) and 14 (B) for patients classified correctly, and 17 (A) and 18 (B) for patients classified incorrectly (data set A: $U = 45.5$, $P = 1.0$; data set B: $U = 61.0$, $P = 0.45$).

Visible MRI trauma lesions were observed in 12 of the 20 patients in data set A and in 9 of the 25 patients in data set B. Based on the proportion of patients with detectable lesions in each data set, we expected, for the LDA classifier in the EC condition, on average eight subjects (~12/20 × 0.67 × 20 subjects) with detectable lesions to be found among the correctly classified patients in data set A and five in data set B. We observed seven (A) and six (B) subjects with detectable lesions among the correctly classified subjects, but taken the small size of the patient groups, statistical evaluation cannot be reliably conducted here. There was no obvious correlation between the within-subject classification accuracies and visible MRI trauma lesions (Pearson correlation coefficient −0.07 and 0.28 for data set A and B, respectively).

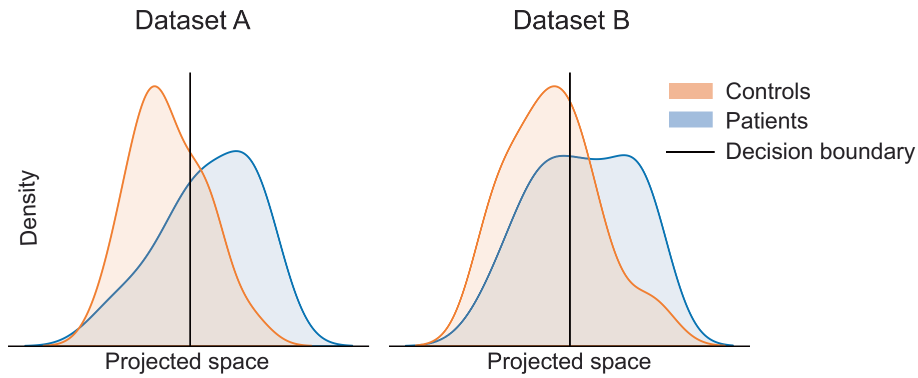

Fig. 2. The distribution of LDA relative log posterior values, projected here to one dimension, for control subjects and mTBI patients (50 repetitions each) for the eyes-closed condition and gradiometer data.

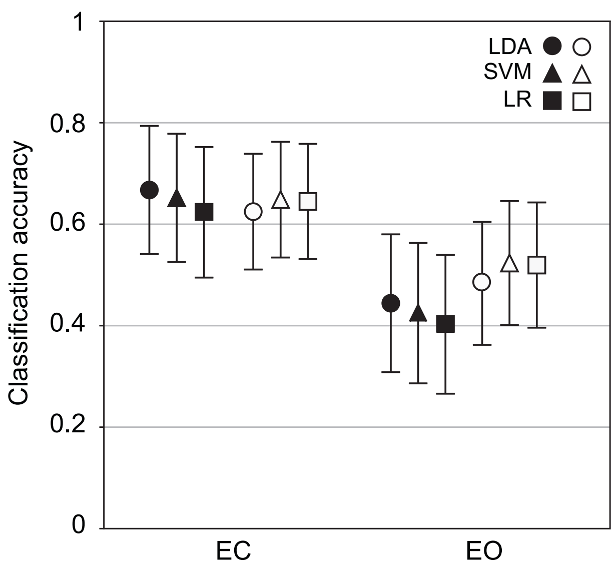

Fig. 3. Mean classification accuracies for data sets A (black) and B (white) for the three classifiers and the two experimental conditions (EC = eyes closed, EO = eyes open). LDA = linear discriminant analysis; SVM = support vector machine; LR = logistic regression.

3.4. Comparison of the ML-based and traditional MEG analysis

In our previous study on data set A, MEG sensor-level analysis revealed abnormal low-frequency activity (LFA, 3–7 Hz) in six patients out of 20 measured within two months after the trauma (Kaltiainen et al., 2018). Using the same criteria, data set B here revealed abnormal activity in 3/25 patients. All the present ML methods outperformed the traditional LFA analysis. For example, LDA in the EC condition correctly classified 13 (15) patients in data set A (B), performing thus significantly better than the traditional sensor-level analysis (data set A: P = 0.001; data set B: P < 0.001).

Interestingly, the ML methods recognized partly different patients than the traditional sensor-level analysis. Patients found by the ML methods did not correlate with the observed LFA (data set A χ^2 = 0.95, P = 0.36; data set B χ^2 = 1.01, P = 0.31): three out of six (one out of three) subjects with visible LFA in data set A(B) were found with the ML algorithms.

3.5. Comparison of wide vs. narrow band classifier performance

When the analysis was restricted to low frequencies alone (1–8 Hz), the median classification accuracy for LDA was 59 % for data set A, and 91 % for data set B, while the wide-band (1–88 Hz) spectra yielded the accuracies 95 % (data set A) and 84 % (data set B). When the analysis was restricted to the alpha–beta band (8–30 Hz), the median accuracies were 93 % (data set A) and 79

Table 2
Summary statistics of the classification results for each classifier in each experimental condition. 25th and 75th refer to the corresponding percentiles. LDA = linear discriminant analysis; SVM = support vector machine; LR = logistic regression; SEM = standard error of the mean.

DATA SET A CLASSIFIER	EXP CONDITION	MEAN ACCURACY	MEDIAN ACCURACY	SEM	25TH	75TH
LDA	EC	0.67	0.95	0.06	0.22	1.00
LDA	EO	0.44	0.34	0.07	0.00	0.97
SVM	EC	0.66	0.88	0.07	0.17	1.00
SVM	EO	0.42	0.17	0.07	0.00	0.96
LR	EC	0.63	0.80	0.06	0.19	1.00
LR	EO	0.40	0.12	0.07	0.00	0.96
DATA SET B CLASSIFIER	EXP CONDITION	MEAN ACCURACY	MEDIAN ACCURACY	SEM	25TH	75TH
LDA	EC	0.63	0.85	0.06	0.14	1.00
LDA	EO	0.48	0.29	0.06	0.02	1.00
SVM	EC	0.65	0.88	0.06	0.18	1.00
SVM	EO	0.52	0.59	0.06	0.00	1.00
LR	EC	0.65	0.88	0.06	0.23	1.00
LR	EO	0.52	0.45	0.06	0.00	1.00

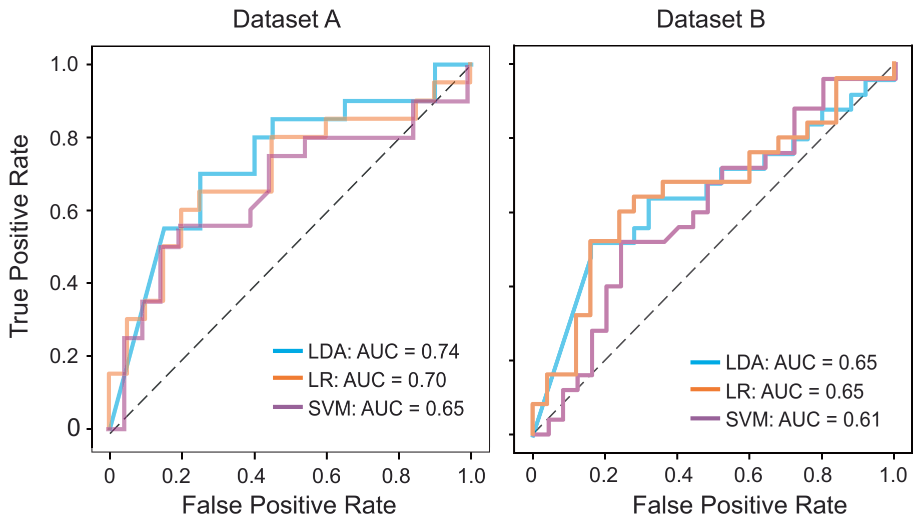

Fig. 4. Receiver operating characteristic curves (ROCs) and corresponding areas under the curve (AUCs) for different classifiers based on the leave-one-out-cross-validation (LOOCV) performance. For both data sets, the curves were estimated using the eyes-closed condition and gradiometer data. LDA = linear discriminant analysis; SVM = support vector machine; LR = logistic regression.

% (data set B). The accuracies of the narrow-band analyses did not differ from those obtained with wide-band analysis for either of the data sets ($P > 0.06$). Permutation analysis showed that the mean classification accuracy appeared to increase as more bands were included in the analysis, but this did not reach statistical significance (see Fig. 5).

4. Discussion

In the present study, we compared the performance of three widely applied linear machine learning (ML) methods (linear discriminant analysis (LDA), Support Vector Machine (SVM), and logistic regression (LR)) for separating mTBI patients from healthy controls based on their resting-state MEG power spectra. We used two independent data sets recorded with different devices and in different measurement environments. All three classifiers separated the patients from controls clearly better than traditional MEG analysis that is based on detecting excessive low-frequency activity. Between the three classifiers, the performance did not significantly differ. However, the eyes-closed (EC) data yielded significantly better classification accuracies than eyes-open (EO) data. The results were comparable in two different patient populations recorded at two different sites, highlighting the robustness of our analysis method.

4.1. Three common machine learning classifiers produced corresponding results irrespective of measurement site and outperformed traditional sensor-level analysis

Traumatic brain injuries, even mild, form a heterogenous group of patients when it comes to trauma mechanisms, impact sites and, therefore, quality and location of possible changes in brain activity. This variety pinpoints the need to simultaneously explore a wide selection of potentially significant features in the data, which could be achieved with the help of ML models. Here we tested the performance of three commonly used ML classifiers and hypothesized that the classification accuracy would not be highly dependent on the exact ML approach. The similar performance of the classifiers can be expected given our dichotomous outcome variable (patients *vs.* controls), and the analogous linear decision boundaries between the used algorithms. The posterior probabilities for both LR and LDA are linear functions of the data, differing only in their

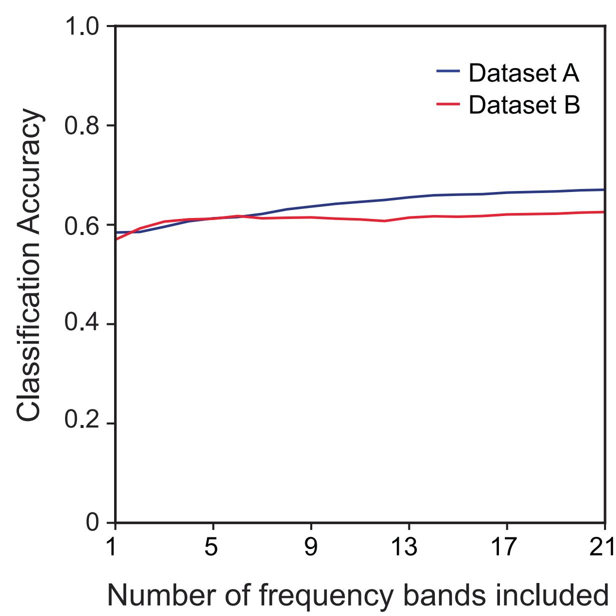

Fig. 5. Change in the mean classification accuracy as a function of the number of bands included in the analysis over 50 permutations (linear discriminant analysis, eyes closed condition).

parameter estimation techniques and exact model assumptions. Furthermore, when applicable, LR and SVM tend to find the same separating hyperplane (Hastie et al., 2001).

The performance of all three classifiers was similar for both sites and data sets suggesting that the approach is reliable and replicable across data sets. Poor reproducibility of the results is a well-known problem in neurotrauma research (Huie et al., 2018), which needs to be solved before any novel method can be taken into clinical practice. Several ML approaches have been applied on neurophysiological data obtained from mTBI patients, but the results have been mixed and difficult to replicate (Thatcher et al., 1989; Trudeau et al., 1998), with high false-positive rates up to 52% (Thornton 1999). Recently, a ML paradigm correctly classified 75% of the mTBI patients at 5–60 months after trauma (Lewine et al., 2019) but the authors suggested the need for earlier evaluation with respect to the trauma. We tested here a simple, straightforward analysis pipeline that could be easily used at different sites and could be applicable also for clinical use. The majority of earlier MEG studies using ML methods have relied on rather complicated connectivity analyses (Antonakakis et al., 2016; Vakorin et al., 2016; Vergara et al., 2017, 2018), which might generate problems when applied to independent data sets with, *e.g.,* different signal-to-noise characteristics (Allen et al., 2021).

All the applied methods yielded significantly better classification accuracy using resting-state EC data compared with EO data. The EC condition may contain less blink and muscle artifacts than the EO condition even after careful artifact management, and thus include less features unrelated to neuronal activity. The EO condition, as a state of higher arousal (*e.g.,* Barry and Blasio 2017), may also introduce more variance within and between individuals that is harder to capture with the models. The EC condition promoting drowsiness may additionally boost low-frequency activity both in healthy subjects (Geller et al., 2014) and especially in the patients, and modulate higher frequency (>30 Hz) activity (Geller et al., 2014), which has been demonstrated to be increased in mTBI patients (Mišić et al., 2016; Huang et al., 2019 and 2021).

4.2. Mild traumatic brain injury causes changes on the oscillatory activity at multiple frequency bands

Previous studies have revealed excess low-frequency activity early after the trauma, with a tendency for the activity to normalize during the successive months (Kaltiainen et al., 2018), but detectable also later after trauma in still symptomatic patients (Lewine et al., 1999; Lewine et al., 2007; Huang et al., 2009; Huang et al., 2014). Besides low frequencies, mTBI-related alterations in oscillatory activity have been reported also, *e.g.,* at beta (15–30 Hz) and gamma (30–90 Hz) bands (Huang et al., 2017, 2019 and 2020). Decrease in alpha peak frequency as well as alpha and beta power have also been reported (Dunkley, 2015; Popescu et al., 2016; Zhang et al., 2020).

Here, wide-band (1–88 Hz) spectral analysis yielded similar results to using only a low-frequency band (1–8 Hz) or a combined alpha–beta band (8–30 Hz), suggesting that changes in the oscillatory brain activity after mTBI occurred over a wide frequency range. Indeed, a deep-learning neural network approach using a wide frequency band yielded better classification performance compared to any narrow band, even if the most significant features were found at low frequency (<8 Hz) and gamma bands (>30 Hz; Huang et al., 2021).

The ML methods presented here significantly outperformed traditional sensor-level spectral analysis (Kaltiainen et al., 2018) in separating patients from controls. However, some of the patients recognized on traditional sensor-level analysis based on low-frequency activity (3–7 Hz) alone were incorrectly classified by the ML method, suggesting that the classification indeed relied on wide spectral content.

4.3. Machine learning approaches may provide increased sensitivity to clinical measurement practice

Our study demonstrates that ML applied on two different data sets obtained from two different measurement sites produced comparable and robust results, suggesting a possibility to utilize such methods also in clinical practice. The straightforward sensor-level analysis pipeline is easily replicable at different sites by different analysts. Information on detectable functional changes after mTBI can provide tools for selecting those patients who are more likely to benefit from closer follow-up, allowing longer recuperation period and gradual return to productive activities.

We recruited our patients in the sub-acute phase quite early after injury. As the recovery after mTBI is often rapid and complete, also possible abnormalities in oscillatory brain activity can be expected to be very mild and transient. Therefore, we did not presume perfect classification performance in this sub-acute measurement setting. Here, as well as in some other studies assessing mTBI patients, there was no clear association between subjective symptom score and classification results (Lewine et al., 2007; Kaltiainen et al., 2018). This emphasizes the various possible etiologies of the post-traumatic symptoms after mTBI, and the difficulties in trying to objectively assess their constellation. The presence or absence of MRI lesions did not significantly correlate with the correct classification performance either, further highlighting the added value of functional imaging in the follow-up after mTBI. In the future, assessing recordings during cognitive tasks might aid in combining patients' cognitive complaints with objective test metrics (Kaltiainen et al., 2019).

4.4. Limitations of the study

Our patient population was heterogeneous: Some of the patients were already well recuperated at the time of the measurement session, thus lowering the probability for correct classification performance. Furthermore, measuring a vast amount of mTBI patients with MEG early after injury is not feasible considering the availability of MEG instrumentation, cost of measurements, and generally good prognosis of the patient population. Therefore, measuring patients with a high symptom score at sub-acute stage might help to distinguish the patients with need for further follow-up, rehabilitation and support. EEG could provide a cost-effective and readily available method for screening purposes in patient groups with a risk for long-term complications. Data from a small cohort (N = 6) suggested that MEG may provide superior sensitivity over EEG (Li et al., 2015), but this issue should be addressed with simultaneous MEG and EEG measurements in larger patient cohorts.

5. Conclusion

Conventional ML approaches classified mTBI patients on the basis of MEG resting-state power spectra at subacute stage (<2 months after the trauma) with median accuracy of 80–95%, without significant differences between three different classifiers. The classifiers performed similarly for two independent data sets, and drastically outperformed the earlier expert-based analysis. The applied methods yielded significantly better classification accuracy on EC than EO data which should be taken into account when developing clinical measurement pipelines.

Our present results suggest that - when measured early after the trauma - rather simple power spectral estimates combined with ML approach can classify mTBI patients with high accuracy, with high promise also for clinical use.

Data and code availability statement

Finnish data protection laws do not allow for MEG data to be made publicly available, as the data cannot be fully anonymized. Data can however be shared with scientific collaborators after an amendment to the research ethics permit via the hospital's ethics committee and a data transfer agreement.

Analysis was carried out using standard software packages as outlined in the methods section.

Funding information & role of the funding source

Hanna Renvall was supported by the Academy of Finland (grant number 321460), Finnish Cultural Foundation and Paulo Foundation, Hanna Kaltiainen by the Finnish Medical Foundation, and Riitta Salmelin by the Academy of Finland (grant number 315553) and the Sigrid Jusélius Foundation.

Declaration of Competing Interest

The authors declare no competing interests.

Acknowledgements

We acknowledge the computational resources provided by the Aalto Science-IT project and the skilled help in the MEG measurements by Mr. Jari Kainulainen. Part of the present work has previously been submitted as a partial fulfillment of the first author's Master of Science Degree (Aaltonen J: Application of linear machine learning methods for the diagnosis of mild traumatic brain injuries, Aalto University 2022).

References

Allen CM, Halsey L, Topcu G, Rier L, Gascoyne LE, Scadding JW, et al. Magnetoencephalography abnormalities in adult mild traumatic brain injury: A systematic review. Neuroimage Clin 2021;31:102697. https://doi.org/10.1016/j.nicl.2021.102697.

American Congress of Rehabilitation Medicine Committee on Social, Ethical and Rehabilitation, Environmental Aspects of (1993). Addressing the post-rehabilitation health care needs of persons with disabilities. Arch Phys Med Rehabil, 74(12 Spec No): S8–14.

Antonakakis M, Dimitriadis SI, Zervakis M, Micheloyannis S, Rezaie R, Babajani-Feremi A, et al. Altered cross-frequency coupling in resting-state Meg after mild traumatic brain injury. Int J Psychophysiol 2016;102:1–11. https://doi.org/10.1016/j.ijpsycho.2016.02.002.

Antonakakis M, Dimitriadis SI, Zervakis M, Papanicolaou AC, Zouridakis G. Altered rich-club and frequency-dependent subnetwork organization in mild traumatic brain injury: A MEG resting-state study. Front Hum Neurosci 2017;11:416. https://doi.org/10.3389/fnhum.2017.00416.

Barry RJ, De Blasio FM. EEG differences between eyes-closed and eyes-open resting remain in healthy ageing. Biol Psychol 2017;129:293–304. https://doi.org/10.1016/j.biopsycho.2017.09.010.

Chen Y, Wiesel A, Eldar YC, Hero AO. Shrinkage algorithms for MMSE covariance estimation. IEEE Trans Signal Process 2010;58(10):5016–29. https://doi.org/10.1109/tsp.2010.2053029.

Dunkley BT. Differential intrinsic coupling modes in psychological and physical trauma. Front Psychiatry 2015;6:140. https://doi.org/10.3389/fpsyt.2015.00140.

Gardner RC, Yaffe K. Epidemiology of mild traumatic brain injury and neurodegenerative disease. Mol Cell Neurosci 2015;66:75–80. https://doi.org/10.1016/j.mcn.2015.03.001.

Geller AS, Burke JF, Sperling MR, Sharan AD, Litt B, Baltuch GH, et al. Eye closure causes widespread low-frequency power increase and focal gamma attenuation in the human electrocorticogram. Clin Neurophysiol 2014;125(9):1764–73. https://doi.org/10.1016/j.clinph.2014.01.021.

Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MEG and EEG data analysis with MNE-Python. Front Neurosci 2013;7:267. https://doi.org/10.3389/fnins.2013.00267.

Hastie T, Tibshirani R, Friedman J. The elements of statistical learning 2001;vol. 1. New York: Springer Series in Statistics; 2001.

Huang MX, Huang CW, Harrington DL, Robb-Swan A, Angeles-Quinto A, Nichols S, et al. Resting-state magnetoencephalography source magnitude imaging with deep-learning neural network for classification of symptomatic combat-related mild traumatic brain injury. Hum Brain Mapp 2021;42(7):1987–2004. https://doi.org/10.1002/hbm.25340.

Huang MX, Huang CW, Harrington DL, Nichols S, Robb-Swan A, Angeles-Quinto A, et al. Marked increases in resting-state MEG gamma-band activity in combat-related mild traumatic brain injury. Cereb Cortex 2020;30(1):283–95. https://doi.org/10.1093/cercor/bhz087.

Huang MX, Nichols S, Robb-Swan A, Angeles-Quinto A, Harrington DL, Drake A, et al. MEG working memory N-back task reveals functional deficits in combat-related mild traumatic brain injury. Cereb Cortex 2019;29(5):1953–68. https://doi.org/10.1093/cercor/bhy075.

Huang MX, Harrington DL, Robb Swan A, Angeles Quinto A, Nichols S, Drake A, et al. Resting-state magnetoencephalography reveals different patterns of aberrant functional connectivity in combat-related mild traumatic brain injury. J Neurotrauma 2017;34(7):1412–26. https://doi.org/10.1089/neu.2016.4581.

Huang MX, Nichols S, Baker DG, Robb A, Angeles A, Yurgil KA, et al. Single-subject-based whole-brain meg slow-wave imaging approach for detecting abnormality in patients with mild traumatic brain injury. Neuroimage Clin 2014;5:109–19. https://doi.org/10.1016/j.nicl.2014.06.004.

Huang MX, Theilmann RJ, Robb A, Angeles A, Nichols S, Drake A, et al. Integrated Imaging Approach with Meg and DTI to detect mild traumatic brain injury in military and civilian patients. J Neurotrauma 2009;26(8):1213–26. https://doi.org/10.1089/neu.2008.0672.

Huie JR, Almeida CA, Ferguson AR. Neurotrauma as a big-data problem. Curr Opin Neurol 2018;31(6):702–9. https://doi.org/10.1097/WCO.0000000000000614.

Hyvärinen A, Oja E. Independent component analysis: algorithms and applications. Neural Netw 2000;13(4–5):411–30.

Iverson GL, Gardner AJ, Terry DP, Ponsford JL, Sills AK, Broshek DK, et al. Predictors of clinical recovery from concussion: A systematic review. Br J Sports Med 2017;51(12):941–8. https://doi.org/10.1136/bjsports-2017-097729.

Jacobs B, Beems T, Stulemeijer M, van Vugt AB, van der Vliet TM, Borm GF, et al. Outcome prediction in mild traumatic brain injury: Age and clinical variables are stronger predictors than CT abnormalities. J Neurotrauma 2010;27(4):655–68. https://doi.org/10.1089/neu.2009.1059.

Kaltiainen H, Helle L, Liljeström M, Renvall H, Forss N. Theta-band oscillations as an indicator of mild traumatic brain injury. Brain Topogr 2018;31(6):1037–46. https://doi.org/10.1007/s10548-018-0667-2.

Kaltiainen H, Liljeström M, Helle L, Salo A, Hietanen M, Renvall H, et al. Mild traumatic brain injury affects cognitive processing and modifies oscillatory brain activity during attentional tasks. J Neurotrauma 2019;36(14):2222–32. https://doi.org/10.1089/neu.2018.6306.

Kaltiainen H-L, Helle LM, Renvall H-M-L, Forss NH. Slow-wave oscillations in awake healthy subjects. J Clin Neurophysiol 2016;33(4):367–72. https://doi.org/10.1097/wnp.0000000000000251.

Korn A, Golan H, Melamed I, Pascual-Marqui R, Friedman A. Focal cortical dysfunction and blood-brain barrier disruption in patients with postconcussion syndrome. J Clin Neurophysiol 2005;22(1):1–9. https://doi.org/10.1097/01.wnp.0000150973.24324.a7.

Koufen H, Dichgans J. Frequency and course of posttraumatic EEG-abnormalities and their correlations with clinical symptoms: A systematic follow up study in 344 adults. Fortschr Neurol Psychiatr Grenzgeb 1978;46(4):165–77.

Lee H, Wintermark M, Gean AD, Ghajar J, Manley GT, Mukherjee P. Focal lesions in acute mild traumatic brain injury and neurocognitive outcome: CT versus 3T MRI. J Neurotrauma 2008;25(9):1049–56. https://doi.org/10.1089/neu.2008.0566.

Lewine JD, Davis JT, Bigler ED, Thoma R, Hill D, Funke M, et al. Objective documentation of traumatic brain injury subsequent to mild head trauma: Multimodal Brain Imaging With MEG, SPECT, and MRI. J Head Trauma Rehabil 2007;22(3):141–55. https://doi.org/10.1097/01.htr.0000271115.29954.27.

Lewine JD, Davis JT, Sloan JH, Kodituwakku PW, Orrison Jr WW. Neuromagnetic assessment of pathophysiologic brain activity induced by minor head trauma. AJNR Am J Neuroradiol 1999;20(5):857–66.

Lewine JD, Plis S, Ulloa A, Williams C, Spitz M, Foley J, et al. Quantitative EEG biomarkers for mild traumatic brain injury. J Clin Neurophysiol 2019;36(4):298–305. https://doi.org/10.1097/WNP.0000000000000588.

Li L, Pagnotta MF, Arakaki X, Tran T, Strickland D, Harrington M, et al. Brain activation profiles in mTBI: Evidence from combined resting-state EEG and MEG activity. Annu Int Conf IEEE Eng Med Biol Soc 2015:6963–6. https://doi.org/10.1109/EMBC.2015.7319994.

Losoi H, Silverberg ND, Wäljas M, Turunen S, Rosti-Otajärvi E, Helminen M, et al. Recovery from mild traumatic brain injury in previously healthy adults. J Neurotrauma 2016;33(8):766–76. https://doi.org/10.1089/neu.2015.4070.

Mišić B, Dunkley BT, Sedge PA, Da Costa L, Fatima Z, Berman MG, et al. Post-traumatic stress constrains the dynamic repertoire of neural activity. J Neurosci 2016;36(2):419–31. https://doi.org/10.1523/JNEUROSCI.1506-15.2016.

Nuwer MR, Hovda DA, Schrader LM, Vespa PM. Routine and quantitative EEG in mild traumatic brain injury. Clin Neurophysiol 2005;116(9):2001–25. https://doi.org/10.1016/j.clinph.2005.05.008.

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. J Mach Learn Res 2011;12(85):2825–30.

Popescu M, Hughes JD, Popescu E-A, Riedy G, DeGraba TJ. Reduced prefrontal Meg alpha-band power in mild traumatic brain injury with associated posttraumatic stress disorder symptoms. Clin Neurophysiol 2016;127(9):3075–85. https://doi.org/10.1016/j.clinph.2016.06.004.

Pozzato I, Cameron ID, Meares S, Kifley A, Vu KV, Liang A, et al. A surveillance study to determine the accuracy of mild traumatic brain injury diagnosis in an emergency department: protocol for a retrospective cohort study. BMJ Open 2017;7(8):e016222.

R Core Team, R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria; 2020. https://www.R-project.org/ [accessed 1 March 2022].

Renvall H, Salmela E, Vihla M, Illman M, Leinonen E, Kere J, Salmelin R. Genome-wide linkage analysis of human auditory cortical activation suggests distinct loci on chromosomes 2, 3, and 8. J Neurosci 2012;32(42):14511–8. https://doi.org/10.1523/jneurosci.1483-12.2012.

Taulu S, Simola J. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Phys Med Biol 2006;51(7):1759–68. https://doi.org/10.1088/0031-9155/51/7/008.

Taulu S, Kajola M, Simola J. Suppression of interference and artifacts by the Signal Space Separation Method. Brain Topogr 2004;16(4):269–75. https://doi.org/10.1023/b:brat.0000032864.93890.f9.

Teasdale G, Jennett B. Assessment of coma and impaired consciousness: A Practical Scale. Lancet 1974;304(7872):81–4. https://doi.org/10.1016/s0140-6736(74)91639-0.
Tebano MT, Cameroni M, Gallozzi G, Loizzo A, Palazzino G, Pezzini G, et al. EEG spectral analysis after minor head injury in man. Electroencephalogr Clin Neurophysiol 1988;70(2):185–9. https://doi.org/10.1016/0013-4694(88)90118-6.
Thatcher R, Walker R, Gerson I, Geisler F. EEG discriminant analyses of mild head trauma. Electroencephalogr Clin Neurophysiol 1989;73(2):94–106. https://doi.org/10.1016/0013-4694(89)90188-0.
Thornton KE. Exploratory investigation into mild brain injury and discriminant analysis with high frequency bands (32–64 Hz). Brain Inj 1999;13(7):477–88. https://doi.org/10.1080/026990599121395.
Thorpe D, Engdahl B, Leuthold A, Georgopoulos A. Assessing recovery from mild traumatic brain injury (Mtbi) using magnetoencephalography (MEG): An application of the Synchronous Neural Interactions (SNI) test. J Neurol Neuromedicine 2020;5(3):28–34. https://doi.org/10.29245/2572.942x/2020/3.1274.
Trudeau DL, Anderson J, Hansen LM, Shagalov DN, Schmoller J, Nugent S, et al. Findings of mild traumatic brain injury in combat veterans with PTSD and a history of blast concussion. J Neuropsychiatry Clin Neurosci 1998;10(3):308–13. https://doi.org/10.1176/jnp.10.3.308.
Vakorin VA, Doesburg SM, da Costa L, Jetly R, Pang EW, Taylor MJ. Detecting mild traumatic brain injury using resting state magnetoencephalographic connectivity. PLoS Comput Biol 2016;12(12):e1004914.
Vergara VM, Mayer AR, Kiehl KA, Calhoun VD. Dynamic functional network connectivity discriminates mild traumatic brain injury through machine learning. Neuroimage Clin 2018;19:30–7. https://doi.org/10.1016/j.nicl.2018.03.017.
Vergara V, Mayer A, Kiehl K, Calhoun V. Detection of mild traumatic brain injury by machine learning classification using resting state functional network connectivity and fractional anisotropy. J Neurotrauma 2017;34(5):1045–53. https://doi.org/10.1089/neu.2016.4526.
Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020;17(3):261–72. https://doi.org/10.1038/s41592-019-0686-2.
Zhang J, Safar K, Emami Z, Ibrahim GM, Scratch SE, da Costa L, et al. Local and large-scale beta oscillatory dysfunction in males with mild traumatic brain injury. J Neurophysiol 2020;124(6):1948–58. https://doi.org/10.1152/jn.00333.2020.