PLOS COMPUTATIONAL BIOLOGY

Using normative modeling and machine learning for detecting mild traumatic brain injury from magnetoencephalography data

Veera Itälinna[1], Hanna Kaltiainen[1,2], Nina Forss[1,2], Mia Liljeström [1,3☯]*,
Lauri Parkkonen [1,4☯]*

1 Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Aalto, Finland, 2 Department of Neurology, Helsinki University Hospital and Clinical Neurosciences, Neurology, University of Helsinki, Helsinki, Finland, 3 BioMag Laboratory, HUS Medical Imaging Center, Helsinki University Hospital, Helsinki, Finland, 4 Aalto NeuroImaging, Aalto University School of Science, Aalto, Finland

☯ These authors contributed equally to this work.
* mia.liljestrom@aalto.fi (ML); lauri.parkkonen@aalto.fi (LP)

Check for updates

Data Availability Statement: The data from the mTBI patients and their healthy controls are not publicly available as it contains information that could compromise the privacy of the research participants. 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. Data can be requested by contacting the ethics committee at HUS at eettinen.toimikunta@hus.fi. The Cam-CAN dataset used for the normative modeling in this

Abstract

New biomarkers are urgently needed for many brain disorders; for example, the diagnosis of mild traumatic brain injury (mTBI) is challenging as the clinical symptoms are diverse and nonspecific. EEG and MEG studies have demonstrated several population-level indicators of mTBI that could serve as objective markers of brain injury. However, deriving clinically useful biomarkers for mTBI and other brain disorders from EEG/MEG signals is hampered by the large inter-individual variability even across healthy people. Here, we used a multivariate machine-learning approach to detect mTBI from resting-state MEG measurements. To address the heterogeneity of the condition, we employed a normative modeling approach and modeled MEG signal features of individual mTBI patients as deviations with respect to the normal variation. To this end, a normative dataset comprising 621 healthy participants was used to determine the variation in power spectra across the cortex. In addition, we constructed normative datasets based on age-matched subsets of the full normative data. To discriminate patients from healthy control subjects, we trained support-vector-machine classifiers on the quantitative deviation maps for 25 mTBI patients and 20 controls not included in the normative dataset. The best performing classifier made use of the full normative data across the entire age and frequency ranges. This classifier was able to distinguish patients from controls with an accuracy of 79%. Inspection of the trained model revealed that low-frequency activity in the theta frequency band (4–8 Hz) is a significant indicator of mTBI, consistent with earlier studies. The results demonstrate the feasibility of using normative modeling of MEG data combined with machine learning to advance diagnosis of mTBI and identify patients that would benefit from treatment and rehabilitation. The current approach could be applied to a wide range of brain disorders, thus providing a basis for deriving MEG/EEG-based biomarkers.

Author summary

Mild traumatic brain injury is extremely common, but no definite diagnostic method is yet available. Objective markers for detecting brain injury are needed to direct care to those who would most benefit from it. Here we present and demonstrate the feasibility of a new approach based on MEG recordings that first explicitly addresses the variability in brain dynamics within the population through normative modeling, and then applies supervised machine-learning to detect pathological deviations related to mTBI. The approach can easily be adapted to other brain disorders as well and could thus provide a basis for an automated tool for analysis of MEG/EEG towards disease-specific biomarkers.

Introduction

Due to its high prevalence and potential long-term adverse health effects, accurate diagnosis of mild traumatic brain injury (mTBI) is of high importance. Objective diagnosis of mTBI remains a challenge, however, as structural imaging methods such as magnetic resonance imaging (MRI) as well as neuropsychological testing often fail to detect clinically significant abnormalities [1,2]. Diagnosis of mTBI after the acute phase is further complicated by post-traumatic symptoms that are nonspecific to mTBI and highly variable across patients [3].

Studies employing noninvasive functional neuroimaging techniques such as functional magnetic resonance imaging (fMRI), electroencephalography (EEG) or magnetoencephalography (MEG) have provided group-level evidence of changes in brain activity following mTBI, even several months after the trauma and in the absence of clinical symptoms [4–7]. New objective measures based on functional brain imaging might prove essential for improving the accuracy and reliability of the diagnosis, and for identifying patients who are at risk of chronic symptoms and would benefit from intervention.

Electrophysiological recordings of brain activity, such as MEG and EEG, provide a range of measures (e.g. amount of low-frequency activity, posterior alpha frequency and power, alterations in functional connectivity, cross-frequency coupling, or network topology measures) that reflect the altered functional state of brain regions and networks after mTBI [5,8–11]. However, the ability to determine the clinical status of individual mTBI patients based on single–or univariate–measures is extremely limited [12], and it is not clear which MEG/EEG measures are most informative of disease pathology. Thus, compound–or multivariate–analysis within a machine-learning framework that jointly exploits multiple measures could potentially increase our ability to accurately detect pathology related to mTBI. Recent studies applying machine learning have shown promise in identifying individuals with mTBI from MEG data using measures of oscillatory activity [13,14], functional connectivity and network topology [11,15,16].

Extraction of objective measures from functional imaging data in mTBI is particularly challenging since the mechanism, location and nature of the head insult, as well as the clinical symptoms are largely heterogeneous. Moreover, individual variation in brain activity is large even in the healthy population and the majority of mTBI patients experience a prompt recovery. Therefore, regarding patients and control subjects as clearly delineated and distinct groups, may not properly reflect the nature of this disorder [17,18]. This variability can partially be addressed by normative modeling [18], where the aim is to map the full range of normal variation within the population and quantify statistical deviations of individual patients. Different normative modeling approaches have recently been applied to neuroimaging data to study disorders such as schizophrenia [19–21], dementia [22,23] and autism [20,24]. The

methods have shown significant promise in providing predictions of disease states at the level of individual subjects.

In this study, we compare source-level power spectra computed from resting-state MEG recordings of mTBI patients and their healthy controls to a large normative reference dataset for the purpose of modeling the pathological features of individual mTBI patients as extreme values or deviations with respect to the normal variation. To discriminate the group of mTBI patients from healthy control subjects, we train a support vector machine (SVM) classifier [25] on the resulting quantitative deviation maps.

A key question in normative modeling is the choice of the reference cohort, which should capture a wide range of variation in the population [26]. An important consideration is therefore the matching of the demographics of the normative reference data to the subject. As there is significant neurophysiological variation across demographic groups, interesting disease-related effects may be diluted if the applied normative data represents the whole population. Here, we explore this question by comparing the results obtained with age-matched and non-matched normative data.

Methods

Datasets

We employed a dataset originally measured by Kaltiainen and colleagues [27,28], comprising resting-state MEG recordings from 25 mild traumatic injury patients and 20 healthy controls. In addition, we employed a large, separate dataset, utilizing MEG recordings from a total of 621 healthy participants [29] as normative data.

Ethics statement

The study was approved by the Ethics Committee of Helsinki and Uusimaa Hospital District. Written consent was obtained from the participants in accordance with the Declaration of Helsinki.

mTBI patients and healthy controls

The patient group consisted of 25 mild traumatic brain injury patients (11 females, 14 males) with a mean age of 42 years (range 20–59 years). The control group comprised 20 healthy subjects (8 females, 12 males) with a mean age of 39 years (range 19–58 years). All patients and controls were without neuropsychological disorders, medication affecting the central nervous system, substance abuse or earlier history of TBI. The patients' level of consciousness was assessed with the Glasgow Coma Scale (GCS) [30] shortly after the injury. GCS addresses the level of consciousness, ranging from three (deep unconsciousness) to 15 (alert and awake). The GCS scores varied between 14 and 15, thus fulfilling the criteria for mTBI. All patients maintained TBI symptoms at their first MEG measurement. At their MEG measurement sessions, the patients filled in the Rivermead Post-Concussion Symptoms Questionnaire (RPQ) [31], which measures the severity of post-concussive symptoms after TBI with a five-step scale, compared with the situation before the accident. The maximum score is 64 but answering "no more of a problem as before the accident" yields one point. The scores of the questionnaire varied from 3 to 36 with an average of 17.2. The demographics of the patient group, as well as their GCS and RPQ scores, are presented in Table 1. All patients fulfilled the criteria for mTBI according to the American Congress of Rehabilitation Medicine (ACRM) criteria [32] with loss of consciousness of less than 30 minutes at the time of the accident, GCS varying between 13–15 at 30 min after the accident and the duration of post-traumatic amnesia less than 24h.

Table 1. Demographics of the mTBI patients.

Patient[a]	Age (years)	GCS	RPQ	Delay (weeks)[b]	Lesions[c]
1	43	15	3	17	–
2	50	15	3	9	+
3	42	14	24	22	+
4	46	14	29	20	+
5	37	14	13	15	+
6	32	15	18	17	+
7	59	15	3	3	+
8	54	15	8	9	–
9	39	15	31	9	–
10	20	14	2	4	+
11	44	14	27	7	+
12	43	14	28	26	–
13	36	14	25	7	+
14	39	15	9	3	–
15	29	14	3	4	–
16	37	14	25	4	+
17	50	14	6	9	+
18	28	15	16	1	–
19	29	14	3	3	+
20	59	14	36	1	+
21	53	14	34	3	+
22	51	15	14	1	–
23	23	15	25	1	–
24	40	14	14	4	+
25	56	15	32	3	–
Average	**41.6**	**14.4**	**17.2**	**8.0**	**60%**

[a]The data were obtained from Kaltiainen and colleagues [27,28]. The patients are in the same order as in Table 1 of Kaltiainen et al. [28].

[b]The time between the injury and the MEG measurement. Times expressed in months were converted to weeks using a factor of 4.345 (the average number of weeks in a month).

[c]Lesions observed in comprehensive MRIs.

GCS–Glasgow Coma Scale score; RPQ–Rivermead Post-Concussion Symptoms Questionnaire.

All patients underwent an MEG measurement within 6 months (26 weeks) after the trauma, 19 of which were performed at the subacute stage within 2 months (9 weeks) of the injury. Nine patients underwent a follow-up comprising of both a MEG recording and neuropsychological testing at 6 months after the first recording [28]. The follow-up recordings were not analyzed in the current study due to the small group size. The MEG measurements were performed at Aalto Neuroimaging MEG Core, Aalto University School of Science, Espoo, Finland, using a 306-channel whole-head MEG device (Elekta Neuromag; MEGIN Oy, Helsinki, Finland). During the recordings, data were filtered to 0.03–330 Hz and sampled at 1000 Hz. Electrocardiogram and horizontal and vertical electro-oculograms were measured for managing artifacts caused by heartbeat and eye movement, respectively. Here, from those recordings, we use one MEG session where the subjects rested with eyes closed for 10 minutes. The subjects were instructed to sit relaxed and avoid movement. The measurement was briefly paused twice to confirm that the subjects remained awake and alert.

Anatomical MRI images (Signa HDX 1.5 T, General Electric, Milwaukee, WI, USA) were acquired from all subjects. MRIs from patients were acquired within one week to 16 months after the injury. Trauma lesions were detected in 15 of the 25 (60%) patients (see Table 1).

Normative dataset

A large open neuroimaging dataset by the Cambridge Centre for Ageing and Neuroscience (Cam-CAN) [29], containing MEG and MRI measurements of nearly 700 healthy participants aged 18 to 87, was used for creating the normative reference data. The curated, cross-sectional Cam-CAN dataset contains measurements from approximately 50 men and 50 women in each age decade (18–27, 28–37, 38–47, 48–57, 58–67, 68–77 and 78–87 years). In this study, only the resting-state, eyes-closed MEG and anatomical MRI were used. Further details of the dataset are presented by Taylor and colleagues [29].

Subjects with missing or incomplete resting-state MEG measurements or T1-weighted MRI images were excluded from the analyses, resulting in a set of 621 subjects. The number of subjects in each age group was 56, 92, 104, 93, 95, 102 and 79, from the youngest to oldest.

Data preprocessing

Artifact removal and data segmentation. The temporal extension of the signal space separation (tSSS) [33] method implemented in the MaxFilter software package (MEGIN Oy) was used for reducing external artifacts in the MEG data. To suppress artifacts caused by cardiac activity and eye movement, independent component analysis (ICA) [34] was used for identifying components most prominently related to the aforementioned sources. In most cases, 1–2 components were removed, but for some subjects three or even four components were removed based on manual inspection of the spatial patterns and time courses of the components. The FastICA algorithm available in MNE-Python software [35] was used for the ICA processing.

Data segmentation was applied on the samples by means of a sliding window with a length of 200 seconds and stride of 50 seconds. This resulted in seven time series for each subject and a total sample size of 315 data points.

Source modeling. An automated source-modeling pipeline was applied on the measurements to compute power spectral densities (PSDs) at each cortical location using the MNE-Python software [36].

Reconstruction of each subject's cortical surface was performed using the FreeSurfer software [37,38] from T1-weighted anatomical MRI images. For the forward computation, a surface-based source space with the ico-4 decimation was created, resulting in a set of 5124 cortical locations at which the amplitudes of the current dipoles were estimated. A single-compartment BEM head model was formed based on brain surface tessellations obtained by the FreeSurfer watershed algorithm [39]. The coregistration of the MEG and MRI coordinate frames was performed automatically using MNE-Python and fiducial points calculated using the FieldTrip [40] toolbox in MATLAB.

For the calculation of the inverse operator, noise covariance matrices were computed from recordings without a subject ("empty-room recording") performed during the same measurement session. The ICA solutions computed for each subject's recording were applied also to these empty-room recordings. The noise covariance matrix and the forward solution were used to compute the dSPM inverse solution, which was applied to the complex-valued 8192-point Fourier transform of the Hann-windowed (50% overlap) raw data over a frequency range of 1–40 Hz. A source-level PSD was obtained by taking the magnitude of the estimate at each source point, yielding a matrix X of size 5124 *(number of source locations)* × 319 *(number*

of frequency values). Finally, the subject-specific cortical PSDs were morphed to a reference brain (the "fsaverage" brain provided by FreeSurfer) to enable comparison of the power spectra across subjects (see Fig 1A).
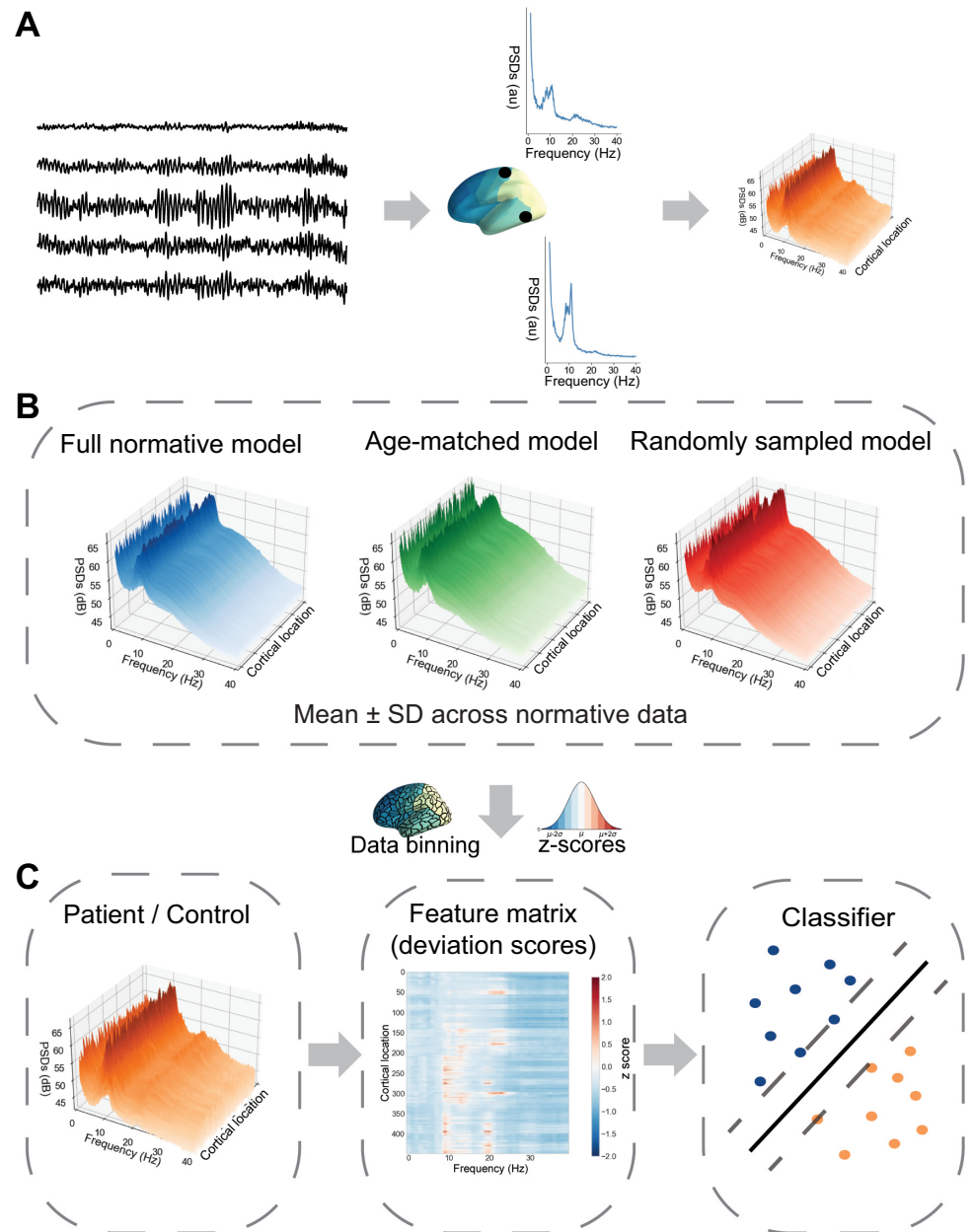

Fig 1. Analysis pipeline. (**A**) **Data preprocessing.** Source-level power spectra were first calculated for each cortical region and morphed to a common brain template. (**B**) **Construction of normative models.** The mean μ and standard deviation σ were calculated across subjects within the reference dataset to obtain normative data for each location and frequency. Three different types of normative models were constructed: a full normative model containing all participants from the reference dataset across the entire age range (depicted in blue, mean values across all frequencies and cortical locations are shown), age-matched models containing a subset of participants within the same age range as the patient/control (exemplar depicted in green), and for comparison a random model with a subset of reference data of random ages (shown in red). (**C**) **Classification procedure.** The normative models were used for converting the power spectra of the patients and controls into deviation scores (z-scores). The deviation scores, binned into 448 cortical parcels and six frequency bands, were then entered into the classification procedure.

https://doi.org/10.1371/journal.pcbi.1011613.g001

Feature engineering.   To obtain normative models, the mean μ and standard deviation σ of the power spectra were calculated across the subjects from the normative dataset (Fig 1B). These statistics were then used for converting the power spectrum matrices for the patients and their healthy controls, into *deviations maps* of Z-scores over the entire cortex, calculated as

$$Z = \frac{X - \mu}{\sigma,} \tag{1}$$

where $X \in \mathbf{R}^{5124 \times 319}$ is the power spectrum of an individual subject.

In addition, data binning was performed both over the spatial and the frequency dimensions to reduce dimensionality. The brain sources were automatically grouped together according to an anatomical parcellation scheme with 448 cortical regions [41], where the activity of a cortical region was represented by the mean power of the source points within that region (see Fig 1C).

In the frequency dimension, six frequency bands were defined: delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), low beta (13–17 Hz), high beta (17–30 Hz), and gamma (30–40 Hz). The average was taken over the power values corresponding to each frequency interval. Binning the data into cortical parcels and into canonical frequency bands significantly reduced the dimensionality of the data: the resulting power spectra of size 448×6 were only 0.16% the size of the original data. Finally, the deviation matrices were flattened into feature vectors containing the six frequency features for each cortical region, resulting in 2688 features per subject.

Model training and validation

A support vector machine with a radial basis-function kernel was selected for classifying the measurements of mTBI patients and controls due to its ability to perform well in high-dimensional settings, even when the size of the dataset is smaller than the number of features [42].

A nested cross-validation strategy was selected for evaluating model performance. In the inner 5-fold cross-validation loop, the best values for the regularization hyperparameters C and γ were chosen based on the best average accuracy across the folds. In the outer 7-fold loop, the model was re-trained using the chosen hyperparameters and evaluated using the independent validation set of the fold. To reduce possible bias of a single cross-validation split, the nested procedure was repeated 5 times with a different split each round, resulting in a total of 35 folds in the outer loop. The splits were stratified by the target labels.

The hyperparameter values tested in the inner cross-validation loop were 1, 5 and 10 for C and 0.1, 0.01 and 0.001 for γ. The penalty parameter C was weighted inversely proportional to class frequencies to avoid bias towards the positive class (patients) which had a small majority.

The data were centered by subtracting the median and scaled to the range between the 1st and 3rd quartile of the data. The median and the interquartile range were calculated from the training data at each cross-validation fold and applied to both training and testing data before fitting and evaluating the model.

The employed data segmentation approach resulted in multiple samples corresponding to the same subject, which leads to the samples being dependent. To avoid leaking information, the cross-validation splits were constructed so that the samples of any single subject were included only in the training or only in the validation set, but never both. The predicted class label for each subject, positive (patient) or negative (control), was determined by the label given to the majority of samples from that subject.

To test the effect of using normative data from a specific age group, the model was trained and evaluated on an age-matched subset of the normative data, i.e., only the power spectra belonging to the same age decade as the subject were used. The age groups were defined according to the normative dataset as 18–27, 28–37, 38–47, 48–57 and 58–67 years. To ensure

that the possible difference in the results was not due to the smaller size of the normative dataset, the model was also evaluated on a dataset where the normative data used for each subject were randomly subsampled so that the size of the normative dataset was equal to the number of normative samples in the subject's age group. The results of the randomized procedure were averaged over three repetitions to reduce the effect of a single random sampling.

In addition to the above three ways of employing the large normative dataset, we trained and evaluated the model on a dataset where the features were created from only the power spectra of the mTBI patients and their healthy controls, applying the same binning scheme as described earlier but without computing the Z-scores using normative data. This was done to assess the performance of the proposed normative modeling approach compared to classifying the power spectra as such.

The statistical significance of the results was explored using permutation tests, where the model was cross-validated 1000 times with randomly permuted group labels for subjects. A p-value < 0.05 was considered significant.

Estimating feature significance

After verifying the predictive capabilities of the model, permutation feature importance [43] was used for estimating how much the model relies on each individual feature for aiding the classification. The method is defined as the decrease in prediction performance (in this case, accuracy) when the values of the feature are randomly permuted while keeping other features intact. When calculating the permutation feature importance, correlated features may lead to artificially low features importance. We expected that many of the features within the dataset would be highly correlated: for example, the alpha-band power values in two neighbouring regions are probably very similar. To reduce multicollinearity of the features, the number of features was reduced with a hierarchical clustering approach, where the Spearman rank correlations between features were clustered using Ward's method. A threshold of 2 was manually selected to form the clusters, the first feature of each cluster was picked, and the resulting set of features was used to train the model with the cross-validation approach described before. The selection of features to be removed was performed only on the training data of each fold to avoid leaking information to the test set. The permutation importance was calculated on the test data of each fold for 5 permutations within each fold and averaged across folds.

Correlation of patient demographics and classification

To evaluate the effects of the timing of the MEG recording with respect to injury and the RPQ score on the classifier's performance (Table 1) we report these values for each patient that was incorrectly classified. In addition, we used a Mann Whitney U-test to assess whether the decision function scores differed significantly between patients with and without visible trauma lesions in their MRIs. The average decision function values of each patient were calculated over the five repeats of the 7-fold cross validation. The decision function values are proportional to the distance of the samples from the hyperplane separating the classes, and so they are indicators of the classifier's confidence regarding a particular sample. The predicted class corresponds to the sign of the decision function output.

Results

Deviation scores

Fig 2A shows the average deviation scores for the patient and control groups after binning the data in the spatial dimension. The patient group shows higher average activation compared to
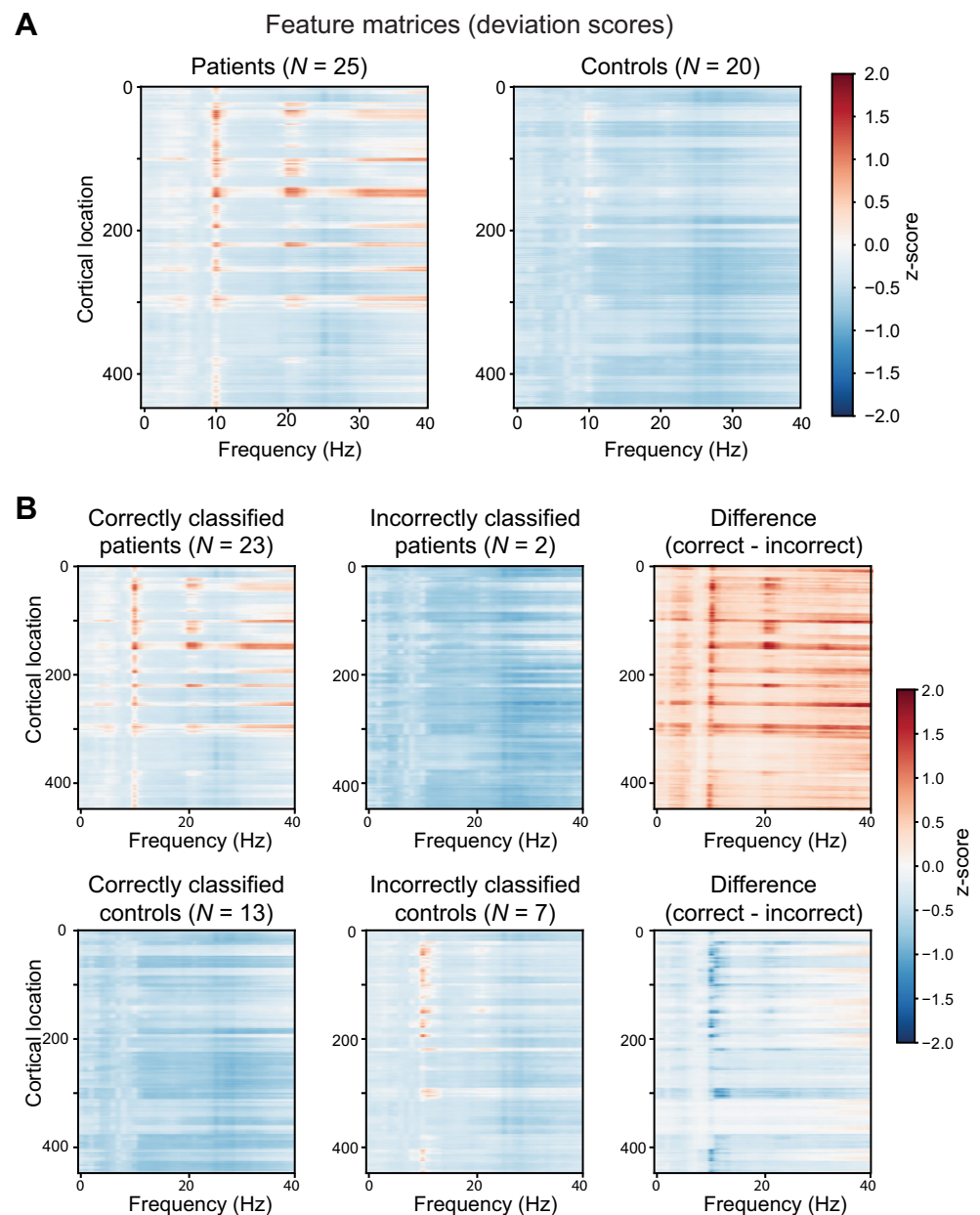
Fig 2. (A) Group-average, relative power spectra of mTBI patients and healthy controls. Horizontal axis is frequency (Hz), vertical axis the cortical location (indices of the 448 cortical regions ordered alphabetically), and the color indicates the Z-score with respect to the normative data at that frequency and cortical location. **(B) Relative power spectra associated with different classification results**. The average Z-score maps for patients (first row) and controls (second row) by classification output, from left to right: correctly classified samples, incorrectly classified samples and the difference between the correctly and incorrectly classified samples.

https://doi.org/10.1371/journal.pcbi.1011613.g002

the normative dataset, mainly around 10 and 20 Hz as well as at frequencies over 30 Hz. The deviation values for the control subjects are overall slightly lower compared to the normative dataset, which might be an effect of different measurement sites.

Classification performance

The mean and standard deviation of the accuracy, sensitivity, and specificity across the cross-validation folds are reported in Table 2 for the three different approaches that were used for

Table 2. Classification of mTBI patients and healthy controls.

Normative data	Accuracy[a]	Sensitivity[a]	Specificity[a]
Full	**0.790** (±0.154)	0.912 (±0.176)	**0.638** (±0.277)
Age-matched	0.761 (±0.150)	0.907 (±0.166)	0.581 (±0.277)
Random	0.711 (±0.153)	**0.914** (±0.154)	0.467 (±0.318)
None	0.786 (±0.162)	0.912 (±0.154)	0.629 (±0.293)

[a]The average (± standard deviation) accuracy, sensitivity and specificity over a 5×7-fold repeated nested cross-validation with hierarchical clustering of correlated features.

https://doi.org/10.1371/journal.pcbi.1011613.t002

selecting the normative data (using the whole normative dataset, an age-group-matched subset, or a random subset of the same size as the age-matched set), as well as for the results obtained without using normative data.

The largest accuracy (0.790) was achieved by using all available normative data. Using age-matched normative data yielded slightly lower results: the accuracy was 0.761 with feature selection by clustering. When comparing the results using age-matched normative data to a random sample of normative data of the same size, the randomly selected normative data yielded a notably lower accuracy of 0.711. Classification without the use of normative data yielded an accuracy of 0.786, which is only marginally lower than the highest value obtained. Permutation tests indicated that the accuracy of the classifier was significantly higher than chance level at $p < 0.05$ for all classification tasks.

In all cases the classifier had a high sensitivity, with the largest value (0.914) obtained for the random normative dataset. The specificity of the classifier was notably lower, at most 0.638 with full normative data. Using age-matched normative data yielded a decrease in sensitivity and an increase in specificity when compared to a random selection of normative data.

Model interpretation

To gain insight to the decision function of the classifier, averaged spectral Z-score maps of both patients and controls were plotted for correctly and incorrectly classified subjects together with their difference; see Fig 2B. In this analysis, the correct vs. incorrect classification was based on the best performing model, which used all available normative data without age-group matching.

Similarly to the observations from Fig 2A, the correctly classified patients seem to be characterized by larger Z-scores around the alpha (~ 10 Hz) and beta (~ 20 Hz) frequencies and also in the high end of the spectrum–the gamma band. Higher activation can also be seen in the slower waves of the theta band. The mTBI patients incorrectly classified as controls appear to be lacking these features at least at the group level, which is likely the reason for their misclassification. For the control group, the most notable difference between the correct and incorrect classifications is in the 10-Hz frequency band, which shows higher values for the false positives. The difference plots show that the values of the correctly classified patients are overall slightly higher than those of the incorrectly classified patients, while the inverse is true for the controls.

Fig 3A shows the permutation importance of the features selected by the hierarchical clustering method. Only 30 features with the largest mean importance are shown. A clear majority of the most significant features correspond to the theta frequency band. The list also includes a few features from the alpha, delta and low beta frequency bands. Cortical areas prominently present among the most important features are mostly located in the parietal lobe, such as the supramarginal gyrus, postcentral gyrus, superior and inferior parietal lobule and precuneus.
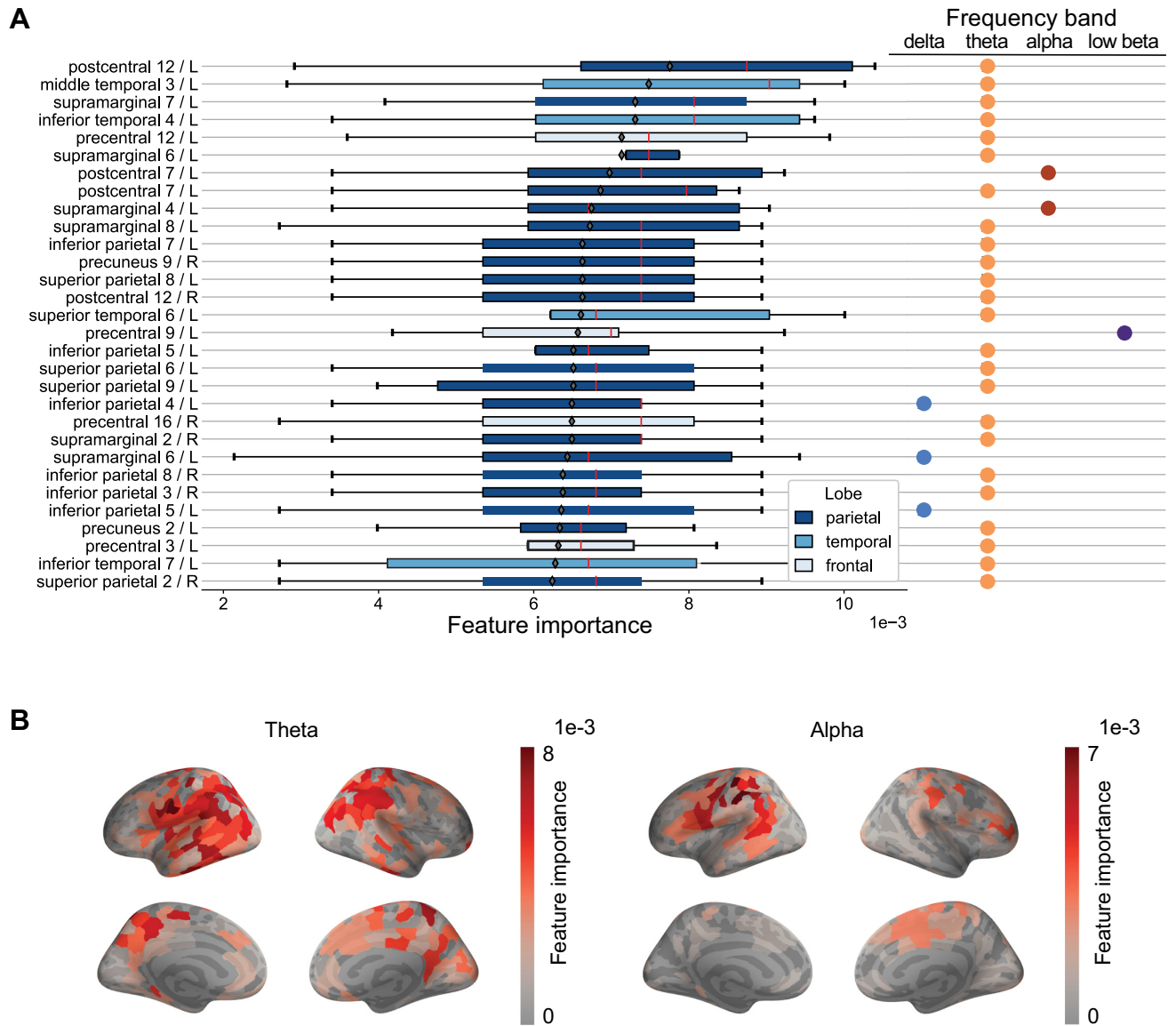
Fig 3. (A) Feature importance. The feature importance (horizontal axis) is defined as the reduction in accuracy when the feature is randomly permuted. The labels of the features indicate the cortical region, the index of the subarea within the subdivided region, and the hemisphere (L for left, R for right) that the feature corresponds to. Only the 30 features with the largest mean importance are shown. Values are sorted according to mean importance across folds, and the median values are shown in red. **(B) Cortical sources contributing to the classification of patients and controls at two frequency bands.** The spatial distribution of the average feature importance across folds, shown for the theta and alpha frequency bands. The values were calculated as the permutation feature importance.

https://doi.org/10.1371/journal.pcbi.1011613.g003

Other featured areas include the precentral gyrus in the posterior frontal lobe and the middle and inferior temporal cortex in the temporal lobe.

The mean values of the estimated feature importance for the theta and alpha bands are visualized superimposed on the cortical surface in Fig 3B. In line with the results in Fig 3A, the most significant features are concentrated in the parietal lobe while there are also some in the temporal and occipital lobes and in parts of the frontal lobe.

To further assess the reliability of the results, the Z-score map of each patient was visually compared to the findings of Kaltiainen and colleagues [27]. In that study, theta band activity
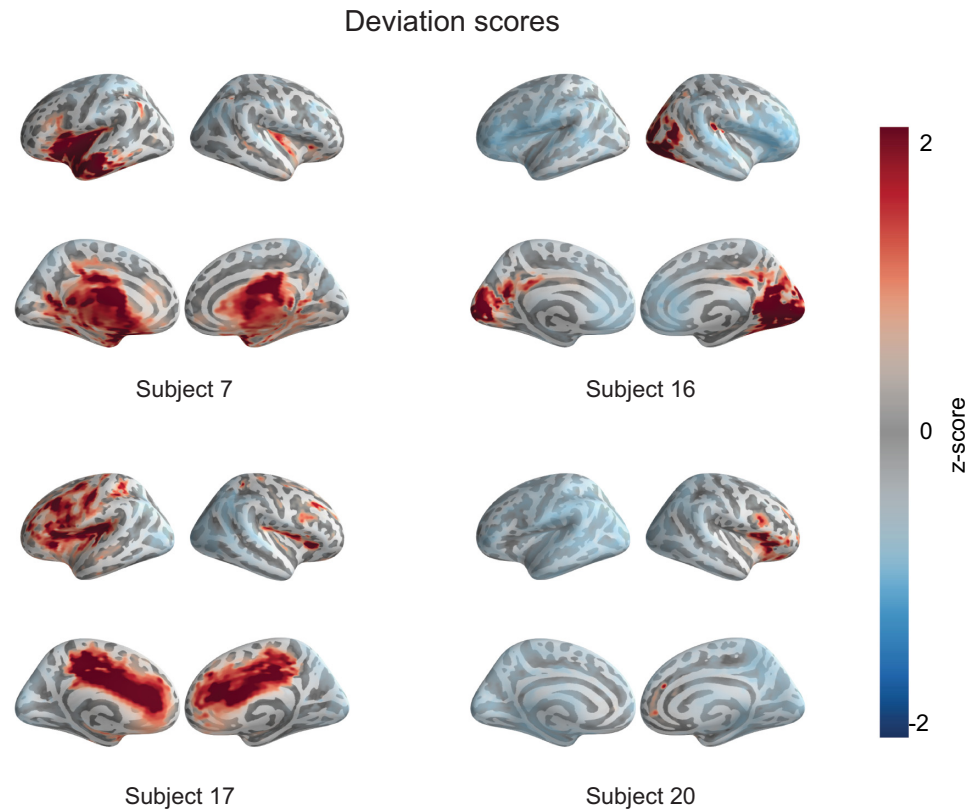
Deviation scores

Fig 4. Deviation score maps for theta-band power in four patients. The color indicates the Z-score with respect to the level of 4–8-Hz activity in the full normative data.

https://doi.org/10.1371/journal.pcbi.1011613.g004

exceeding two standard deviations from the healthy subjects' average was found in seven of the 26 mTBI patients, 25 of which were also analyzed in the current study where the Z-maps (computed with the full normative dataset) revealed such aberrant low-frequency activity in eight patients, five of which were the same as the ones identified in that earlier study. With age-matched normative data, the abnormality was found in one additional patient. Representative examples of the Z-score maps of patients with abnormal theta activity are shown in Fig 4. As seen in the figure, the locations of this abnormal low-frequency activity are highly variable.

The eight patients with abnormal theta activity (or nine in the age-matched case) were classified by the model with an accuracy of 0.950 (age-matched: 0.911). On the other hand, a similar increase of the theta frequency band was observed in three out of the 20 healthy controls with non-matched data and in four with age-matched data. These subjects were classified incorrectly as patients without exception.

Overall, patients were classified correctly with high sensitivity: with the full normative model, only two patients were consistently misclassified (Patients 1 and 5 in Table 2). Both of these patients were measured relatively late with respect to the time of injury (4 and 3.5 months, respectively), and had RPQ scores below the average of the patient cohort (3 and 13, respectively). In the two misclassified patients, the earlier study identified abnormal oscillatory activity over frontal areas in the gamma and theta frequency range [27]. One of these patients had a lesion in their MRI; however, the lesion was deep in the brain and not associated with low-frequency activity in the earlier study [27]. There was no significant difference in the

decision function output between those patients who had a lesion in their MRI and those who did not (p = 0.52, Mann Whitney U-test).

Discussion

Classification accuracy and significant features

Finding objective diagnostic biomarkers for mTBI is challenging due to the high variability and nonspecificity of posttraumatic symptoms. Combining measurements of brain electric activity with machine learning could aid in identification of mTBI, and thus help clinical decision making. Here we show that mTBI patients can be separated from healthy controls with 79.0% accuracy using quantified deviations from normative power spectra combined with supervised machine learning. Activity in the theta frequency band (4–8 Hz) provided the most significant discriminative features for the classification, with additional contributions from the delta, alpha and low beta frequencies. Increased neural oscillatory activity below 8 Hz, previously associated with axonal injury [44], is the most frequent finding in mTBI patients even in the chronic stage of the injury [4,5,10,44–47]. The obtained results are thus in line with previous literature.

Considering that 10 out of 25 mTBI patients did not have any structural lesions visible in MRI scans and that more than half of the MEG recordings were conducted over a month post-injury, the achieved accuracy can be regarded as satisfactory. Furthermore, it is likely that the patient group includes subjects whose brain activity and function could be considered normal by all relevant metrics despite the earlier trauma, so achieving an accuracy close to 100% may be an unrealistic goal.

Recent studies utilizing machine learning combined with MEG or EEG data to predict mTBI at the subacute or chronic stage have yielded promising results. The accuracy of the methods presented in this paper slightly exceed the accuracy reported by Cao and colleagues [48], where 61 subjects were classified with a 77.1% accuracy using task-related EEG measurements, and those by Lewine and colleagues [12] who achieved a 75% accuracy in classifying 153 subjects using five global features calculated from resting-state EEG data. It is notable that we reach a comparable accuracy with only 45 subjects compared to 61 and 153 subjects in these earlier studies, respectively. In a recent MEG study, Huang and colleagues [13] reached a sensitivity of 95.5% and a specificity of 90% in pediatric mTBI when combining delta and gamma band spectral features. Importantly, Aaltonen and colleagues [14] also demonstrated that several common classifiers applied to power spectral features derived from MEG data produce corresponding results on two different patient cohorts, recorded at two different sites. Taken together, these studies show that power spectral features derived from MEG or EEG data can provide a neural signature of brain injury that can differentiate between mTBI patients and control subjects. The current approach that employs features quantifying the deviation from a normative sample yields a more interpretable result compared to typical machine-learning models, which is a significant advantage over previous models, despite its lower accuracy in some cases. The approach can be further extended to measures derived from functional connectivity that have proven sensitive in identifying patients with mTBI, reaching in some cases over 90% classification accuracy [11,15,16,49].

Signal features within the theta frequency band contributed significantly to the classification accuracy. The low-frequency abnormalities were also a major finding for the current dataset in the original study [27]. Interestingly, features from the alpha, beta and gamma frequency bands were not among the most significant features, even though these bands showed visible differences between patients and controls at the group level. Previously, Zhang and colleagues detected reduced beta power in frontotemporal regions [50], but for changes in alpha and

gamma oscillatory power after mTBI, the reports of alterations in neural oscillatory power are contradictory [11,45,51,52]. A possible explanation for the lack of significant features is that there is large physiological variability in these frequency bands between and even within individuals e.g. according to their vigilance and attention. Another explanation might be the outlier values in these bands that affect the average but do not contain valuable information for the training of the SVM classifier, as the decision boundary of the SVM is robust to outliers. While this and many other studies have detected abnormalities in spectral features in the lower frequency bands related to mTBI with high sensitivity, such deviations from the normal range can occur also from other causes (ie. the deviations are not specific to mTBI). Focusing on a single aspect, such as theta frequency abnormalities, may thus be too limiting for the highest classification accuracy.

The most significant features of the theta frequency band were found to be located in the parietal, temporal and occipital regions of the brain, whereas the most significant features from the delta, alpha and beta band were detected in parietal and the Rolandic cortex. The paucity of frontal features is notable, however, since frontobasal areas are among the most frequent lesion sites after mTBI [53,54]. The locations generating mTBI-related low-frequency activity are typically highly variable [46], and likely to be influenced by the location of the impact to the head. This spatial variance was confirmed with visual inspection of the individual Z-score maps presented here. This heterogeneity highlights the need to focus on individual-level abnormalities rather than a "typical mTBI patient", as in a traditional case-control paradigm.

Normative modeling: Advantages and considerations

Interpretability of machine-learning models in a medical context is important due to safety and ethical concerns: clinicians should be able to identify possible errors in the model's predictions [55]. Normative modeling, an intuitive approach familiar from children's growth charts, together with a supervised classifier has the potential to help place confidence in the model's predictions and detect possible errors. In addition to the prediction of the model, the decisions can be aided with visualizations of the patients' individual cortical Z-score maps, possibly limited to the theta frequency band.

Many of the earlier studies utilizing a normative modeling approach have relied on Gaussian process regression [56] for modeling the healthy variation [18,21,24], which has the benefit of quantifying the uncertainty of the model. The method could be explored in the context of our proposed approach in future research. In this study, a relatively simple and straightforward calculation of Z-scores was adapted, as robust statistical inference was not of concern: the normative modeling provided features for supervised machine learning rather than being directly used for discriminating patients from controls. A larger normative dataset of thousands of measurements might also enable the use of state-of-the-art deep learning methods for building the normative model, such as deep autoencoders as in Pinaya and colleagues [20,22], which have performed well in complex tasks but require a large number of samples to learn effectively.

The deviation values for our control subjects were slightly below zero, probably reflecting technical differences in the MEG recording sites. To alleviate these effects, multi-site calibration approaches that have been developed for normative modeling of MRI data could be adapted to MEG data as well [57]. Here, since the same normative model was applied to both patient and control data, we do not expect the effect on the classification results to be notable.

As neurophysiological patterns vary significantly with age, patients should be compared to their own age group in normative comparisons [58]. In this study, selecting the normative data

by each subject's own age group for calculating the Z-score maps yielded better results compared to using an equally sized random sample of normative data. This suggests that true abnormalities are indeed more reliably identified if a subject is compared with their own age group: selecting the normative data by age may lead to better detection of phenomena normal for some age groups but possibly pathological for others. However, the highest accuracy was achieved by using all available normative data, which suggests that a sufficiently large normative dataset is needed to capture enough individual variation. A normative database integrated with a future clinical application should thus ideally aggregate thousands of M/EEG measurements of healthy subjects across different studies, sites and demographic variables to enable selecting a sufficiently sized subset of normative data with suitable properties for each task.

The current study demonstrates the feasibility of employing a normative modeling approach applied to MEG data to identify patients with mTBI. Further studies should be conducted using a larger clinical dataset. Large functional imaging datasets from clinical populations are, however, rarely available, and drawing reliable conclusions about classification results is often hindered by the small size of the datasets. In this study, the effect of the dataset size was most clearly seen in the large standard deviations of the accuracy, sensitivity and specificity scores, which ranged from 0.15 to 0.32. Larger clinical datasets would increase the robustness of the predictions and identification of the most significant features but collecting such large-scale patient data remains a challenge.

One limiting factor in translating research on candidate biomarkers for detecting mTBI from MEG data to clinical settings are the sometimes complex analysis pipelines, which may be time-consuming or require extensive expertise. Here, the analysis was automated with the exception of artefact removal, which required manual inspection of the ICA components selected for removal. Using novel automated artefact removal algorithms that utilize machine-learning algorithms [59–61] could help clean the data further and provide a fully automated analysis pipeline.

We employed an SVM classifier, as they typically perform well on data with high dimensionality [42]. The nonlinearity of the SVM classifier does, however, introduce some limitations to the interpretability of the results, as the weights of the individual features are not directly interpretable. Importantly, the results of such methods should not be thought of as revealing exactly the location and frequency of the most discriminative features, but as giving a general sense about brain activation patterns associated with mTBI.

Conclusions

We introduced a normative-modeling and machine-learning approach capable of discriminating mild traumatic brain injury patients from healthy controls with up to 79.0% accuracy. Most of the features that were significant for the classification corresponded to the theta frequency band, the excess activity of which has been associated with pathological phenomena, including mTBI, in earlier studies. The approach could help differentiate mTBI-type symptoms in patients who exhibit prolonged symptoms suggestive of TBI and/or problems with vocational performance. The approach could also help detect patients in need of a neuropsychological intervention.

We demonstrated how normative modeling enables an intuitive interpretation of the predictions of the classifier even at the level of individual patients, a framework that could be recycled for other clinical diagnostic needs. This framework would enable building a system where, for example, a brain scan of an individual patient could be automatically checked for different pathological patterns against a large normative database. The present study acts as an example use case for such a system, with a preprocessing and classification pipeline that can be made fully automatic from the raw measurement data to the final results.

Acknowledgments

Author Contributions

Conceptualization: Veera Itälinna, Hanna Kaltiainen, Nina Forss, Mia Liljeström, Lauri Parkkonen.

Data curation: Veera Itälinna, Hanna Kaltiainen.

Formal analysis: Veera Itälinna, Mia Liljeström.

Funding acquisition: Lauri Parkkonen.

Investigation: Veera Itälinna, Hanna Kaltiainen, Nina Forss, Mia Liljeström, Lauri Parkkonen.

Methodology: Lauri Parkkonen.

Project administration: Lauri Parkkonen.

Supervision: Mia Liljeström, Lauri Parkkonen.

Visualization: Veera Itälinna, Mia Liljeström.

Writing – original draft: Veera Itälinna, Hanna Kaltiainen, Nina Forss, Mia Liljeström, Lauri Parkkonen.

Writing – review & editing: Veera Itälinna, Hanna Kaltiainen, Nina Forss, Mia Liljeström, Lauri Parkkonen.

References

1. Bigler ED, Abildskov TJ, Goodrich-Hunsaker NJ, Black G, Christensen ZP, Huff T, et al. Structural neuroimaging findings in mild traumatic brain injury. Sports Medicine and Arthroscopy Review. 2016; 24(3): e42–52. https://doi.org/10.1097/JSA.0000000000000119 PMID: 27482782

2. Dikmen S, Machamer J, Temkin N. Mild traumatic brain injury: Longitudinal study of cognition, functional status, and post-traumatic symptoms. Journal of Neurotrauma. 2017; 34(8):1524–30. https://doi.org/10.1089/neu.2016.4618 PMID: 27785968

3. Wäljas M, Iverson GL, Lange RT, Hakulinen U, Dastidar P, Huhtala H, et al. A prospective biopsychosocial study of the persistent post-concussion symptoms following mild traumatic brain injury. Journal of Neurotrauma. 2015; 32(8):534–47. https://doi.org/10.1089/neu.2014.3339 PMID: 25363626

4. 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: Clinical. 2014; 5:109–19.

5. 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. Journal of Head Trauma Rehabilitation. 2007; 22(3):141–55. https://doi.org/10.1097/01.HTR. 0000271115.29954.27 PMID: 17510590

6. Pontifex MB, O'Connor PM, Broglio SP, Hillman CH. The association between mild traumatic brain injury history and cognitive control. Neuropsychologia. 2009; 47(14):3210–6. https://doi.org/10.1016/j. neuropsychologia.2009.07.021 PMID: 19664646

7. Rogers JM, Fox AM, Donnelly J. Impaired practice effects following mild traumatic brain injury: An event-related potential investigation. Brain Injury. 2015; 29(3):343–51. https://doi.org/10.3109/ 02699052.2014.976273 PMID: 25436891

8. Castellanos NP, Paúl N, Ordóñ Ez VE, Demuynck O, Bajo R, Campo P, et al. Reorganization of functional connectivity as a correlate of cognitive recovery in acquired brain injury. Brain. 2010; 133 (8):2365–81. https://doi.org/10.1093/brain/awq174 PMID: 20826433

9. Dimitriadis SI, Zouridakis G, Rezaie R, Babajani-Feremi A, Papanicolaou AC. Functional connectivity changes detected with magnetoencephalography after mild traumatic brain injury. NeuroImage: Clinical. 2015; 9:519–31.

10. Haneef Z, Levin HS, Frost JD, Mizrahi EM. Electroencephalography and quantitative electroencephalography in mild traumatic brain injury. Journal of Neurotrauma. 2013; 30(8):653–6. https://doi.org/10.1089/neu.2012.2585 PMID: 23249295

11. 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. International Journal of Psychophysiology. 2016; 102:1–11. https://doi.org/10.1016/j.ijpsycho.2016.02.002 PMID: 26910049

12. Lewine JD, Plis S, Ulloa A, Williams C, Spitz M, Foley J, et al. Quantitative EEG biomarkers for mild traumatic brain injury. Journal of Clinical Neurophysiology. 2019; 36(4):298–305. https://doi.org/10.1097/WNP.0000000000000588 PMID: 31094883

13. Huang MX, Angeles-Quinto A, Robb-Swan A, De-la-Garza BG, Huang CW, Cheng CK, et al. Assessing Pediatric Mild Traumatic Brain Injury and Its Recovery Using Resting-State Magnetoencephalography Source Magnitude Imaging and Machine Learning. Journal of Neurotrauma [Internet]. 2023 Jun 1 [cited 2023 Aug 29]; Available from: https://www.liebertpub.com/doi/ https://doi.org/10.1089/neu.2022.0220 PMID: 36884305

14. Aaltonen J, Heikkinen V, Kaltiainen H, Salmelin R, Renvall H. Sensor-level MEG combined with machine learning yields robust classification of mild traumatic brain injury patients. Clinical Neurophysiology. 2023 Sep 1; 153:79–87. https://doi.org/10.1016/j.clinph.2023.06.010 PMID: 37459668

15. Thorpe DR, Engdahl BE, Leuthold A, Georgopoulos AP. Assessing Recovery from Mild Traumatic Brain Injury (Mtbi) using Magnetoencephalography (MEG): An Application of the Synchronous Neural Interactions (SNI) Test. Journal of Neurology & Neuromedicine [Internet]. 2020 Sep 3 [cited 2023 Aug 29]; 5(3). Available from: https://www.jneurology.com/articles/assessing-recovery-from-mild-traumatic-brain-injury-mtbi-using-magnetoencephalography-meg-an-application-of-the-synchronous-neural-interactions-sni-test.html.

16. 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. Frontiers in Human Neuroscience [Internet]. 2017 [cited 2023 Aug 29]; 11. Available from: https://www.frontiersin.org/articles/10.3389/fnhum.2017.00416. https://doi.org/10.3389/fnhum.2017.00416 PMID: 28912698

17. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it. Molecular Psychiatry. 2012; 17(12):1174–9. https://doi.org/10.1038/mp.2012.105 PMID: 22869033

18. Marquand AF, Rezek I, Buitelaar J, Beckmann CF. Understanding heterogeneity in clinical cohorts using normative models: beyond case-control studies. Biological Psychiatry. 2016; 80(7):552–61. https://doi.org/10.1016/j.biopsych.2015.12.023 PMID: 26927419

19. Kia SM, Marquand AF. Neural Processes Mixed-Effect Models for Deep Normative Modeling of Clinical Neuroimaging Data. In: Cardoso MJ, Feragen A, Glocker B, Konukoglu E, Oguz I, Unal G, et al., editors. Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning [Internet]. PMLR; 2019. p. 297–314. (Proceedings of Machine Learning Research; vol. 102). Available from: https://proceedings.mlr.press/v102/kia19a.html.

20. Pinaya WHL, Mechelli A, Sato JR. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study. Human Brain Mapping. 2019; 40(3):944–54. https://doi.org/10.1002/hbm.24423 PMID: 30311316

21. Wolfers T, Doan NT, Kaufmann T, Alnæs D, Moberget T, Agartz I, et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. JAMA Psychiatry. 2018; 75(11):1146–55. https://doi.org/10.1001/jamapsychiatry.2018.2467 PMID: 30304337

22. Pinaya WHL, Scarpazza C, Garcia-Dias R, Vieira S, Baecker L, F da Costa P, et al. Using normative modelling to detect disease progression in mild cognitive impairment and Alzheimer's disease in a cross-sectional multi-cohort study. Scientific Reports. 2021 Aug 3; 11(1):15746. https://doi.org/10.1038/s41598-021-95098-0 PMID: 34344910

23. Ziegler G, Ridgway GR, Dahnke R, Gaser C. Individualized Gaussian process-based prediction and detection of local and global gray matter abnormalities in elderly subjects. NeuroImage. 2014; 97:333–48. https://doi.org/10.1016/j.neuroimage.2014.04.018 PMID: 24742919

24. Zabihi M, Oldehinkel M, Wolfers T, Frouin V, Goyard D, Loth E, et al. Dissecting the heterogeneous cortical anatomy of autism spectrum disorder using normative models. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2019; 4(6):567–78.

25. Vapnik VN. The Nature of Statistical Learning Theory. New York: Springer; 1995.

26. Marquand AF, Kia SM, Zabihi M, Wolfers T, Buitelaar JK, Beckmann CF. Conceptualizing mental disorders as deviations from normative functioning. Molecular Psychiatry. 2019; 24(10):1415–24. https://doi.org/10.1038/s41380-019-0441-1 PMID: 31201374

27. Kaltiainen H, Helle L, Liljeström M, Renvall H, Forss N. Theta-band oscillations as an indicator of mild traumatic brain injury. Brain Topography. 2018; 31(6):1037–46. https://doi.org/10.1007/s10548-018-0667-2 PMID: 30097835

28. 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. Journal of Neurotrauma. 2019; 36(14):2222–32. https://doi.org/10.1089/neu.2018.6306 PMID: 30896274

29. Taylor JR, Williams N, Cusack R, Auer T, Shafto MA, Dixon M, et al. The Cambridge Centre for Ageing and Neuroscience (Cam-CAN) data repository: Structural and functional MRI, MEG, and cognitive data from a cross-sectional adult lifespan sample. NeuroImage. 2017; 144:262–9. https://doi.org/10.1016/j.neuroimage.2015.09.018 PMID: 26375206

30. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. The Lancet. 1974; 304(7872):81–4. https://doi.org/10.1016/s0140-6736(74)91639-0 PMID: 4136544

31. King NS, Crawford S, Wenden FJ, Moss NEG, Wade DT. The Rivermead Post Concussion Symptoms Questionnaire: a measure of symptoms commonly experienced after head injury and its reliability. Journal of Neurology. 1995; 242(9):587–92. https://doi.org/10.1007/BF00868811 PMID: 8551320

32. Kay T, Harrington D, Adams Anderson, Berrol Cicerone, et al. Definition of mild traumatic brain injury. The Journal of Head Trauma Rehabilitation [Internet]. 1993; 8(3). Available from: https://journals.lww.com/headtraumarehab/Fulltext/1993/09000/Definition_of_mild_traumatic_brain_injury.10.aspx.

33. Taulu S, Simola J. Spatiotemporal signal space separation method for rejecting nearby interference in MEG measurements. Physics in Medicine and Biology. 2006; 51(7):1759–68. https://doi.org/10.1088/0031-9155/51/7/008 PMID: 16552102

34. Hyvärinen A, Oja E. Independent component analysis: Algorithms and applications. Neural Networks. 2000; 13(4–5):411–30. https://doi.org/10.1016/s0893-6080(00)00026-5 PMID: 10946390

35. Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, et al. MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience. 2013; 7:267. https://doi.org/10.3389/fnins.2013.00267 PMID: 24431986

36. Gramfort A, Luessi M, Larson E, Engemann D, Strohmeier D, Brodbeck C, et al. MNE software for processing MEG and EEG data. NeuroImage. 2014; 23(1):1–7. https://doi.org/10.1016/j.neuroimage.2013.10.027 PMID: 24161808

37. Dale AM, Fischl B, Sereno MI. Cortical surface-based analysis: I. Segmentation and surface reconstruction. NeuroImage. 1999; 9(2):179–94.

38. Fischl B, Sereno MI, Dale AM. Cortical surface-based analysis: II. Inflation, flattening, and a surface-based coordinate system. NeuroImage. 1999; 9(2):195–207.

39. Ségonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, et al. A hybrid approach to the skull stripping problem in MRI. NeuroImage. 2004; 22(3):1060–75. https://doi.org/10.1016/j.neuroimage.2004.03.032 PMID: 15219578

40. Oostenveld R, Fries P, Maris E, Schoffelen JM. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Computational Intelligence and Neuroscience [Internet]. 2011; 2011. Available from: https://www.fieldtriptoolbox.org/. https://doi.org/10.1155/2011/156869 PMID: 21253357

41. Khan S, Hashmi JA, Mamashli F, Michmizos K, Kitzbichler MG, Bharadwaj H, et al. Maturation trajectories of cortical resting-state networks depend on the mediating frequency band. NeuroImage. 2018; 174:57–68. https://doi.org/10.1016/j.neuroimage.2018.02.018 PMID: 29462724

42. Nayak J, Naik B, Behera HS. A comprehensive survey on support vector machine in data mining tasks: applications & challenges. International Journal of Database Theory and Application. 2015; 8(1):169–86.

43. Breiman L. Random forests. Machine Learning. 2001; 45(1):5–32.

44. 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. Journal of Neurotrauma. 2009; 26(8):1213–26. https://doi.org/10.1089/neu.2008.0672 PMID: 19385722

45. Dunkley BT, Da Costa L, Bethune A, Jetly R, Pang EW, Taylor MJ, et al. Low-frequency connectivity is associated with mild traumatic brain injury. Neuroimage Clin. 2015 Mar 3; 7:611–21. https://doi.org/10.1016/j.nicl.2015.02.020 PMID: 25844315

46. Huang MX, Nichols S, Robb A, Angeles A, Drake A, Holland M, et al. An automatic MEG low-frequency source imaging approach for detecting injuries in mild and moderate TBI patients with blast and non-

blast causes. NeuroImage. 2012; 61(4):1067–82. https://doi.org/10.1016/j.neuroimage.2012.04.029 PMID: 22542638

47. Robb Swan A, Nichols S, Drake A, Angeles AM, Diwakar M, Song T, et al. Magnetoencephalography slow-wave detection in patients with mild traumatic brain injury and ongoing symptoms correlated with long-term neuropsychological outcome. Journal of Neurotrauma. 2015; 32(19):1510–21. https://doi.org/10.1089/neu.2014.3654 PMID: 25808909

48. Cao C, Tutwiler RL, Slobounov S. Automatic classification of athletes with residual functional deficits following concussion by means of EEG signal using support vector machine. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2008; 16(4):327–35. https://doi.org/10.1109/TNSRE.2008.918422 PMID: 18701381

49. 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: Clinical. 2021 Jan 1; 31:102697. https://doi.org/10.1016/j.nicl.2021.102697 PMID: 34010785

50. 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 Dec 1; 124(6):1948–58. https://doi.org/10.1152/jn.00333.2020 PMID: 33052746

51. 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. Cerebral Cortex. 2020; 30(1):283–95. https://doi.org/10.1093/cercor/bhz087 PMID: 31041986

52. Popescu M, Hughes JD, Popescu EA, Riedy G, DeGraba TJ. Reduced prefrontal MEG alpha-band power in mild traumatic brain injury with associated posttraumatic stress disorder symptoms. Clinical Neurophysiology. 2016; 127(9):3075–85. https://doi.org/10.1016/j.clinph.2016.06.004 PMID: 27472543

53. Shin SS, Bales JW, Edward Dixon C, Hwang M. Structural imaging of mild traumatic brain injury may not be enough: overview of functional and metabolic imaging of mild traumatic brain injury. Brain Imaging and Behavior. 2017; 11(2):591–610. https://doi.org/10.1007/s11682-017-9684-0 PMID: 28194558

54. Yuh EL, Hawryluk GWJ, Manley GT. Imaging concussion: A review. Neurosurgery. 2014; 75(4):S50–63. https://doi.org/10.1227/NEU.0000000000000491 PMID: 25232884

55. Watson DS, Krutzinna J, Bruce IN, Griffiths CEM, McInnes IB, Barnes MR, et al. Clinical applications of machine learning algorithms: Beyond the black box. BMJ (Online). 2019; 364(March):1–5. https://doi.org/10.1136/bmj.l886 PMID: 30862612

56. Williams CKI, Rasmussen CE. Gaussian processes for regression. Advances in Neural Information Processing Systems. 1996; 8:514–20.

57. Bethlehem R a. I, Seidlitz J, White SR, Vogel JW, Anderson KM, Adamson C, et al. Brain charts for the human lifespan. Nature. 2022 Apr; 604(7906):525–33. https://doi.org/10.1038/s41586-022-04554-y PMID: 35388223

58. Nuwer MR, Hovda DA, Schrader LM, Vespa PM. Routine and quantitative EEG in mild traumatic brain injury. Clinical Neurophysiology. 2005; 116(9):2001–25. https://doi.org/10.1016/j.clinph.2005.05.008 PMID: 16029958

59. Treacher AH, Garg P, Davenport E, Godwin R, Proskovec A, Bezerra LG, et al. MEGnet: Automatic ICA-based artifact removal for MEG using spatiotemporal convolutional neural networks. NeuroImage. 2021 Nov 1; 241:118402. https://doi.org/10.1016/j.neuroimage.2021.118402 PMID: 34274419

60. Duan F, Phothisonothai M, Kikuchi M, Yoshimura Y, Minabe Y, Watanabe K, et al. Boosting specificity of MEG artifact removal by weighted support vector machine. In: 2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2013. p. 6039–42.

61. Hasasneh A, Kampel N, Sripad P, Shah NJ, Dammers J. Deep Learning Approach for Automatic Classification of Ocular and Cardiac Artifacts in MEG Data. Journal of Engineering. 2018 May 2; 2018: e1350692.