NeuroImage 197 (2019) 470–481

Contents lists available at ScienceDirect

NeuroImage

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

Decrypting the electrophysiological individuality of the human brain: Identification of individuals based on resting-state EEG activity

Seyed Abolfazl Valizadeh [a,b,c,d,**], Robert Riener [b,e], Stefan Elmer [a,1], Lutz Jäncke [a,c,f,*,1]

[a] Division Neuropsychology, Institute of Psychology, University of Zurich, Switzerland
[b] Sensory-Motor Systems Lab, Department of Health Sciences and Technology, ETH Zurich, Zurich, Switzerland
[c] University Research Priority Program (URPP) "Dynamic of Healthy Aging", University of Zurich, Switzerland
[d] Department of Internal Medicine, University Hospital Zurich, Switzerland
[e] Spinal Cord Injury Center, University Hospital Balgrist, University of Zurich, Zurich, Switzerland
[f] Department of Special Education, King Abdulaziz University, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Keywords:
Biomarker
Identification
EEG
Resting-state
Functional connectivity
Effective connectivity
Machine learning

ABSTRACT

Biometric identification (BI) of individuals is a fast-growing field of research that is producing increasingly sophisticated applications in several spheres of everyday life. Previous magnetic resonance imaging (MRI) studies have demonstrated that based on the high inter-individual variability of brain structure and function, it is possible to identify individuals with high accuracy. Otherwise, there is the common belief that electroencephalographic (EEG) data recorded at the surface of the scalp are too noisy for identification purposes with a comparably high hit rate. In the present work, we compared BI quality (F1-scores, accuracy, sensitivity, and specificity) between different types of functional (instantaneous, lagged, and total coherence, phase synchronization, correlation, and mutual information) and effective (Granger causality, phase synchronization, and coherence) connectivity measures. Results revealed that across functional connectivity metrics, identification accuracy was in the range of 0.98–1, whereas sensitivity and F1-scores were between 0.00 and 1 and specificity was between 0.99 and 1. BI was higher for the connectivity metrics that are contaminated by volume conduction (instantaneous connectivity) compared to those that are unaffected by this variable (lagged connectivity). Support vector machine and neural network algorithms yielded the highest BI, followed by random forest and weighted *k*-nearest neighborhood, whereas linear discriminant analysis was less accurate. These results provide cross-validated counterevidence to the belief that EEG data are too noisy for identification purposes and demonstrate that functional and effective connectivity metrics are particularly suited for BI applications with comparable accuracy to MRI. Our results have important implications for fast, low-cost, and mobile BI applications.

1. Introduction

The identification of biomarkers that can be used to identify dysfunction has a long tradition in the fields of medicine (Majewski and Bernards, 2011; Ng et al., 2016) and neuroscience (Barry et al., 2010; Davidson, 2003; John et al., 1988; Kam et al., 2013; Kumari et al., 2004; Putman, 2011; Stam et al., 2005; Thatcher et al., 2005; Venables et al., 2009; Woltering et al., 2012). However, although this procedure enables the assignment of individuals to fixed categories (i.e., health or disease) based on biomarkers that are shared within a specific population, it does not take into account situational and personal variability (Gordon et al., 2015). Otherwise, biometric identification (BI) is producing increasingly sophisticated applications in several spheres of everyday life by focusing on individual variability rather than inter-individual similarity (Jain et al., 2000).

In recent years, a multidisciplinary field of research rooted in neuroscience and computer science has begun to evaluate the suitability and reliability of the human brain for BI purposes. This approach is anchored on the notion that the human brain demonstrates a high degree of inter-individual variability in terms of histology (Gage and Muotri,

* Corresponding author. Institute of Psychology, Division Neuropsychology, University of Zurich, Binzmühlestrasse 14/25, 8050, Zurich, Switzerland.
** Corresponding author. Division Neuropsychology, Institute of Psychology, University of Zurich, Switzerland.
E-mail addresses: valizadeh@hest.ethz.ch (S.A. Valizadeh), robert.riener@hest.ethz.ch (R. Riener), s.elmer@psychologie.uzh.ch (S. Elmer), lutz.jaencke@uzh.ch (L. Jäncke).
[1] Shared last authorship.

https://doi.org/10.1016/j.neuroimage.2019.04.005
Received 30 January 2019; Accepted 1 April 2019
Available online 9 April 2019

2012), anatomy (Scheperjans et al., 2008; Uylings et al., 2005), and function (Kanai and Rees, 2011; Pelofi et al., 2017) that can be estimated by means of neuroimaging techniques such as magnetic resonance imaging (MRI) and diffusion tensor imaging (DTI). Until now, several machine learning approaches have been used to identify individuals, including linear discriminant analysis (Yeh et al., 2016), matching (Finn et al., 2015), K-nearest neighbor (Wang et al., 2010), Support Vector Machine (Wang et al., 2010), and voting discriminate analysis (Wachinger et al., 2014). Such machine learning applications have been shown to be particularly fruitful in identifying individuals based on a variety of MRI-related brain parameters, such as for example, local gyrification index and brain shape (Wachinger et al., 2014), white matter architecture (Yeh et al., 2016), volume and cortical thickness measures (Valizadeh et al., 2018), as well as functional connectivity (Finn et al., 2015).

Electroencephalography (EEG) constitutes an alternative technique that can be used for BI purposes. In fact, the inter-individual variability of brain functions can be quantified using different electrophysiological parameters, including event-related potentials (ERPs), power spectra, time-frequency distributions, microstates, as well as different types of functional and effective connectivity measures (Phung et al., 2014; Poulos et al., 1999; Su et al., 2010). However, the scalp EEG signal is noisy, characterized by small signal amplitudes, and only enables to capture the synchronous activity of several thousand or millions of activated nerve cells. In addition, due to the simultaneous convergence of multiple electromagnetic fields at the surface of the scalp and their reciprocal interactions, it is still a matter of debate whether the EEG signal demonstrates sufficient sensitivity and specificity for determining an individual brain signature with high reliability (Michel, 2009). Previous EEG studies (DelPozo-Banos et al., 2015; Gui et al., 2015; Huang et al., 2012; Khalifa et al., 2012; Kostilek and St'astny, 2012; Lan et al., 2015; Marcel and Millan, 2007; Mu and Hu, 2011; Shedeed, 2011; Su et al., 2010; Wang et al., 2012; Zhao et al., 2010) reported controversial results in terms of BI accuracy (i.e., in the range of 72–100%). Furthermore, previous studies strongly differed in terms of the number of participants (i.e., in the range of N = 4–203), did not include information about sensitivity and specificity of BI, and did not take advantage of parametric and non-parametric predictors in order to estimate the variability of identification accuracy as a function of number of participants. Currently, it is also unknown how strongly BI accuracy is dependent upon the experimental design applied, and the classifiers used.

Functional connectivity is an influential concept in cognitive neuroscience (Fingelkurts et al., 2005), and can be described in terms of statistical dependencies (i.e., functional connectivity) or causal interactions (i.e., effective connectivity) between distinct electrodes or brain regions (Bastos and Schoffelen, 2016). In the present work, we used EEG and tested the usefulness of different connectivity measures for BI during resting-state conditions. In particular, we compared BI quality (i.e., F1-scores, accuracy, sensitivity, and specificity) between different types of functional (i.e., instantaneous, lagged, and total coherence, phase synchronization, correlation, and mutual information) and effective (i.e., Granger causality, phase synchronization, and coherence) connectivity in a large (N = 119) and heterogeneous sample of participants. Thereby, we used five of the most widespread machine learning methods (i.e., linear discriminant analysis, weighted *k*-nearest neighbor, support vector machine, neural network, and random forest). Furthermore, in order to assess the reliability and stability of BI, we applied several cross-validation procedures. These procedures consisted of (a) varying sample size, (b) adding Gaussian noise to the connectivity matrices, (c) using different EEG epoch durations for learning (i.e., 6, 12, 18 s), (d) comparing independent-component-analysis (ICA)-corrected EEG data with uncorrected ones, and (e) applying different tasks in between two resting-state periods in order to stimulate different brain functions. In addition, (f) we tested the sensitivity and specificity of BI by repeatedly measuring a single individual across three weeks and slightly modifying the spatial position of the EEG cap, varying the circadian measurement time, and applying different cognitive tasks in between two resting-state periods. Finally, (g) we also included five additional participants who were measured twice at time point 0 and 6 months later. This additional analysis focused on result stability over time and test-retest reliability.

2. Methods

2.1. Participants

In the present EEG study, we included the resting-state data from different projects. The entire dataset included 144 healthy participants of different age cohorts (i.e., young and older adults) with a differential degree of expertise (i.e., musicians, non-musicians, and synesthetes). From the original pool of 144 participants, 25 subjects have been excluded because the duration of the artifact-free resting-state periods were shorter than 2 min. For 95 participants, two resting-state periods were available (i.e., pre- and post-task resting-state, eyes closed), whereas for 24 participants only one resting-state period was collected. The entire sample of 119 participants included 46 men and 73 women in the age range of 18–69 years. In between two resting-state periods, a part of the participants performed a word learning task (Dittinger et al., 2016), a phonetic discrimination task (Dittinger et al., 2016; Elmer et al., 2017), or a pitch discrimination task (Jancke et al., 2012). All participants were consistent right-handers according to the Annett-Handedness-Questionnaire (Annett, 1970), healthy and had no history of neurological or psychiatric disorders, migraine, diabetes or tinnitus. Some of the data implemented in this study originate from previous projects of our groups (Elmer et al., 2017; Jancke et al., 2012; Kuhnis et al., 2013).

2.2. EEG measurements

During recording of the resting-state EEG, the participants were required to close their eyes, not to move, and to not think about specific situations of their life while they sat on a comfortable chair. The EEG was recorded with a 32-channel system, a sampling rate of 1000 Hz and a high pass filter of 0.1 Hz using an EEG-amplifier (Brainproducts, Munich, Germany). The 32 silver/silver-chloride electrodes were located according to the 10/10 system at frontal, central, parietal, occipital, and temporal scalp sites (Fp1, Fp2, F7, F3, Fz, F4, F8, FT7, FC3, FCz, FC4, FT8, T7, C3, Cz, C4, T8, TP9, TP7, CP3, CPz, CP4, TP8, Tp10, P7, P3, Pz, P4, P8, O1, Oz, O2). The reference electrode was placed on the tip of the nose, and electrode impedance was kept below 10 kΩ using an electrically conductive gel.

2.3. EEG data analyses

The EEG data were preprocessed with the Brain Vision Analyzer software package (version 2.0.1, Brain Products GmbH, D-82205 Gilching). In a first step, raw EEG data were filtered with a band-pass filter of 1–40 Hz, including a notch-filter of 50 Hz. Eye movements and muscle artifacts were corrected by applying an independent-component analysis (ICA) implemented in the Brain Vision Analyzer toolbox. Remaining muscle artifacts were identified and eliminated using automatic raw data inspection (max. gradient 50 μV/ms, max. allowed absolute signal difference 200 μV, max amplitude ±100 μV, lowest allowed activity 0.5 μV). Afterwards, the EEG data were segmented into 2-s epochs and connectivity metrics were calculated for each of these epochs.

2.4. Connectivity measures

In the present study, we computed twelve different types of connectivity metrics that have previously been frequently used in the field of neuroscience, namely (1) coherence correlation, (2) phase correlation

(Jeong et al., 2001; Na et al., 2002), (3) instantaneous coherence, (4) lagged coherence, (5) total coherence (sum of instantaneous and lagged coherence), (6) instantaneous phase connectivity, (7) lagged phase connectivity, and (8) total phase connectivity (sum of instantaneous and lagged phase connectivity) (Pascual-Marqui, 2007). In addition, we computed the following functional connectivity measures which are commonly used in the fields of computer science and electrical engineering, namely (9) coherence mutual information and (10) phase mutual information (Jeong et al., 2001; Na et al., 2002). We also included two effective connectivity measures (Granger causality), namely (11) coherence Granger causality and (12) phase Granger causality (LeRoy, 2004). In the following paragraphs, we will shortly describe the connectivity metrics we used.

(1) Cross-correlation (*Cor*) (from which coherence correlation and phase correlation are derived) is a measure of similarity between two time series as a function of the displacement of one signal relative to the other. If we assume that X and Y are multivariate time series, then:

$$corr(X.Y) = \frac{cov(X.Y)}{\sigma_x \sigma_y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_x \sigma_y} \tag{1}$$

$$\mu_x = E[X] = \sum_{i=1}^{n} x_i p_i \tag{2}$$

$$\sigma_x = \sqrt{E\left[(X-\mu_x)^2\right]} = \sqrt{E[X^2] - E[x]^2} \tag{3}$$

The signal has two parts: (1) the phase (Phs) and (2) the amplitude (Coh). Accordingly, there are two types of connectivity measures that are calculated using this formula (coherence correlation: *CohCor* and phase correlation: *PhsCor*). The real part of this formula is the *CohCor,* while the imagery part is *PhsCor*.

(2) Total linear dependency is defined as:

$$F_{X.Y}(\omega) = ln \frac{\begin{vmatrix} S_{YY\omega} & 0 \\ 0 & S_{XX\omega} \end{vmatrix}}{\begin{vmatrix} S_{YY\omega} & S_{YX\omega} \\ S_{XY\omega} & S_{XX\omega} \end{vmatrix}} \tag{4}$$

S_{xy} is the power spectral density of the signal which is defined as:

$$S_{XY\omega} = \frac{1}{N_R} \sum_{j=1}^{N_R} X_{j\omega} Y^*_{j\omega} \tag{5}$$

By using the coherence information, we get *CohTot*. The coherence information is obtained by the following equation:

$$X_{j\omega} = \sum_{t=0}^{N_T - 1} X_{jt} e^{-2\pi i \omega t / N_t} \tag{6}$$

In addition, by using phase we get *PhsTot*. Phase information is obtained by the following equation:

$$\check{X}_{j\omega} = \left(X^*_{j\omega} X_{j\omega}\right)^{1/2} X_{j\omega} \tag{7}$$

(3) Instantaneous linear dependency based on coherence (*CohIns)* is defined as:

$$F_{X..Y}(\omega) = ln \frac{\left| Re\begin{pmatrix} S_{YY\omega} & 0 \\ 0 & S_{XX\omega} \end{pmatrix} \right|}{\left| Re\begin{pmatrix} S_{YY\omega} & S_{YX\omega} \\ S_{XY\omega} & S_{XX\omega} \end{pmatrix} \right|} \tag{8}$$

In addition, by using the phase information, we have *PhsIns*.

$$G_{\check{X}.\check{Y}}(\omega) = ln \frac{\begin{vmatrix} S_{\check{Y}\check{Y}\omega} & 0 \\ 0 & S_{\check{X}\check{X}\omega} \end{vmatrix}}{\begin{vmatrix} S_{\check{Y}\check{Y}\omega} & S_{\check{Y}\check{X}\omega} \\ S_{\check{X}\check{Y}\omega} & S_{\check{X}\check{X}\omega} \end{vmatrix}} \tag{9}$$

(4) Lagged linear coherence is defined as the subtraction between total and instantaneous dependence. The lagged coherence (*CohLag*) is defined in the following manner:

$$F_{X \leftrightarrows Y}(\omega) = F_{X,Y}(\omega) - F_{X\ .Y}(\omega) \tag{10}$$

The lagged phase (*PhsLag*) is defined as:

$$G_{\check{X} \leftrightarrows \check{Y}}(\omega) = G_{\check{X}.\check{Y}}(\omega) - G_{\check{X}.\check{Y}}(\omega) \tag{11}$$

(5) The mutual information (*MuI*) of two random variables is a measure of the mutual dependence between the two variables. More specifically, it quantifies the "amount of information" obtained about one random variable through the other random variable. Formally, the mutual information of two discrete random variables X and Y can be defined as I (X,Y)

$$I(X.Y) = \sum_{y \in Y} \sum_{x \in X} p(x.y) log\left(\frac{p(x.y)}{p(x)p(y)}\right) \tag{12}$$

Thereby, p(x,y) is the joint probability distribution function of X and Y, and p(x) and p(y) are the marginal probability distribution functions of X and Y, respectively. In our case, it is difficult to find the joint probability directly. Therefore, we used entropy instead of probability:

$$H(X) = -\sum_{i=1}^{n} P(x_i) log_2 P(x_i) \tag{13}$$

$$I(X.Y) = H(X) + H(Y) - H(X.Y) = H(X) - H(X|Y) \tag{14}$$

H represents the entropy of the signal and *P* is a probability mass function. By using the coherence part, we get *CohMuI* and by using the phase part we get *PhsMuI*.

(6) Granger causality is based on the multivariate regression of two time series, where one time series is predicted by the information of the other time series. In other words, causality is realized by a linear model between time series 1 (X) and time series 2 (Y) (LeRoy, 2004).

$$X(t) = \sum_{j=1}^{p} A_{11} X(t-j) + \sum_{j=1}^{p} A_{12} Y(t-j) + e_1 \tag{15}$$

$$Y(t) = \sum_{j=1}^{p} A_{21} X(t-j) + \sum_{j=1}^{p} A_{22} Y(t-j) + e_2 \tag{16}$$

By the assumption of lag size (p) equal to signal size, and if the exclusion condition is satisfied and the number of the internal and external variable was the same, than the A matrix is nonsingular, and the solution for this equation is the following:

$$A = \left[X \times \acute{X}\right] \times X \times Y \tag{17}$$

In this formula, $\acute{X}$ is a transform version of X. In other words, the regression between input and output represents the causality (*CohGrC and PhsGrC)*.

2.5. Machine learning algorithms

For the identification of single participants, we used five different

techniques, namely (1) linear discriminant analysis (LDA), (2) weighted *k*-nearest neighborhood (WKNN), (3) support vector machine (SVM), (4) neural network (NN), and (5) random forest (RF). LDA is one of the most straightforward identification techniques currently available. LDA classifiers strive at finding the best linear combination of predictors to optimize the separation between multiple classes. Often the primary goal of a LDA is to project a feature space onto a smaller subspace while maintaining the class-discriminatory information. The 1-nearest neighbor rule identifies the class of unknown data points based on its closest neighbor data point (i.e., the class of the latter is already known). Cover and Hart (1967) described the *k*-nearest neighbor (kNN) method, where the nearest neighbor is calculated by the value of *k* that specifies how many nearest neighbors have to be considered to define a class of a sample data point. Here, we used the weighted *k*NN method (with the square inverse distance) that uses a weight for the distance function. Wettschereck and colleagues (Wettschereck et al., 1997) revealed that kNN was not sensitive to the exact choice of *k* when *k* was large enough. They found that for small values of k, the *k*NN algorithm was more robust than the single nearest neighbor algorithm (1NN) for the majority of large datasets tested. The support vector machine (SVM) technique uses high (or infinite) hyperplanes (i.e., dimensions) for classification. The best hyperplane is the one that can result in the largest distance to the training data classes. SVM is a parametric technique, which transfers the non-linearities into linearities (Cristianini and Shawe-Taylor, 2000). As a fourth approach, we used a neural network (NN). In particular, we applied 11 different learning methods for NN. The best result was achieved for the second back propagation, as suggested by Baitti (Battiti, 1992). This algorithm has the advantage that it requires less storage and computation time in comparison to other learning methods. NN is a parametric classification algorithm that is able to solve linear and nonlinear problems (i.e., the XOR problem). Finally, we used the Random Forest (RF) technique (Breiman, 2001) that relies on the bagging (bootstrap aggregating) and decision tree techniques. Breiman showed that the accuracy of RF depends on the strength of the individual tree classifiers and the measure of the dependence between them. This algorithm is called RF because it generates a relatively large number of random trees. The features for classification using random forest are selected randomly in each tree. In addition, the training and test time points are randomly selected. Although RF uses randomness to bagging, Breiman showed that the RF is stable and convergent. More detailed information about random forest can be found in a previous work of Strobl and colleagues (Strobl et al., 2009). RF divides the non-linear space into several smaller linear sub-spaces. All programming codes were written in Matlab by one of the authors (V.S.A.) and run on an iMac machine (iMac, processor: 4 GHz i7-L2 Cache: 256 KB - L3 Cache 8 MB, memory: 32 GB 1600 MHz DDR3, OS X Sierra).

2.6. Machine learning procedures

The resting-state EEG was randomly segmented into an epoch of 1 min (i.e., the minimal resting-state duration across participants was of two minutes) that served as the training data set. The rest of the resting-state EEG was used for testing (test data set). The resting-state EEG period selected for training was further segmented into epochs of two seconds, resulting in 30 successive epochs for each participant. For each of these epochs, we calculated the 12 connectivity measures between all possible pairs of electrodes (32 * 31/2 = 496 connectivities), resulting in 496 connectivity measures per epoch. In total, this resulted in 30 * 496 = 14880 connectivity metrics per participant and connectivity measure, with the exception of the Granger causality measures for which 32 * 32 = 1024 connectivity measures were obtained. Since we calculated the Granger causality matrices for each 2-s epoch, we obtained 30 * 1024 = 30720 connectivity matrices.

The resting-state EEG data set used for testing comprised epochs that lasted from 1 to 4 min. In a similar way as the training data set, the test data set was segmented into 2 s epochs, resulting in 30–120 epochs. For each epoch and participant, we calculated the 12 connectivity measures between all electrodes exactly in the same manner as described for the training data set. Since we were motivated to extract the temporal dynamics of connectivity as precisely as possible, we chose 2 s epochs for our analyses. Two seconds are deemed as being stable enough to calculate connectivity measures from EEG data (Pascualmarqui et al., 1995).

The machine learning procedures were applied to the training data sets separately for each connectivity measure. The results of this training procedure were then applied to the test data sets. Since the number of epochs of the test data sets were partly different from those of the training data sets, we adjusted the results obtained for the test data set accordingly. For the comparison between the identification results obtained by the different classifier techniques, we used the Cochran's Q test, which is an extension of the McNemar test that is used when the response variable is dichotomous and there are multiple times points (repeated measurements). In the case of a significant Cochran's Q test, subsequent Bonferroni-adjusted McNemar tests were performed (SPSS version 22 for Mac OS X). In total, we performed 96 Cochran's Q tests. For comparing the performance of the different identification methods, we computed 5 McNemar tests (one for each dataset), resulting in a total of 101 statistical comparisons. In order to correct for multiple comparisons, we applied the Bonferroni procedure (Holm, 1979) with an adjusted threshold of $p = 0.0004$ (i.e., $p = 0.05/101 = 0.0004$).

2.7. Description of the cross-validation procedures

In order to assess the reliability and stability of the results, we applied several cross-validation procedures. First, we randomly selected different participants and built different data pools differing in sample size (i.e., in the range of 10–110, with increments of 10 participants), resulting in 12 different sample sizes. For each sample, we iterated the identification procedure by running the classifiers for the training data set and applied the results to the test data set. Second, we added Gaussian noise (i.e., for each epoch, ranging from 5% to 40%, in incremental steps of 5%) to the connectivity matrices of the test data obtained for the different classifiers. By adding noise to the connectivity matrices, we simulated noisy EEG signals that are expected to result in a change of the connectivity measures. Third, we used different durations for the training data sets ranging from 6 s to 1 min (i.e., in incremental steps of 6 s). This procedure was applied to examine whether the identification results remained stable or changed as a function of different durations of the training data sets. For this validation, we only used the LDA technique because it was the technique revealing the least good results and requiring less computation time. Fourth, we compared BI results between ICA-corrected EEG data and non-ICA-corrected EEG periods. For the latter, EEG samples were simply removed when they contained artifacts (i.e., automatic raw data inspection, see methods section). Furthermore, we randomly chose 12 participants from our datasets containing two resting-state time points. We have chosen this smaller sample in order to reduce computation time. Fifth, as a further validation, we conducted repeated measurements with a single participant in order to evaluate whether the identification results remained stable across repeated time points. With this purpose in mind, we changed the size of the EEG cap and the circadian recording time (i.e., early in the morning, afternoon, or evening), inserted different cognitive tasks in between successive resting-state measurements (i.e., playing video games, attentive listening to music, solving mathematical problems), and performed EEG measurements under the influence of caffeine or without caffeine. Sixth, as a final validation we also tested BI of 5 participants (the data were taken from a new pool of EEG data from our lab) who were measured twice with 6 months in-between the two measurement times.

2.8. Definition of identification assessment

In order to evaluate the identification results obtained for the different connectivity measures and classifiers, we used accuracy,

sensitivity, specificity, and F1-scores. These metrics are typically used in the machine learning community and are reported elsewhere (Chicco, 2017). Accuracy (ACC) is defined as the total number of correctly classified participants divided by the total number of subjects. Sensitivity (also called true positive ratio, TPR) is the ratio between the number of true positives (TP) over the sum of the true positives and false negatives (FN). Specificity (also termed true negative rate, TNR) is the ratio of true negatives (TN) over the sum of false positive (FP) and true negatives (TN). F1-score is the harmonic mean of precision and sensitivity, where precision is defined as the ratio of TP over the sum of TP and FN. This F1-score combines sensitivity and precision and gives a more comprehensive picture of the classification results. In this work, we arbitrarily defined F1-scores > 0.98 as perfect, F1-scores > 0.9 and < 0.98 as high, F1-scores > 0.7 and < 0.9 as moderate, and F1-scores < 0.7 as not satisfactory. Finally, it is important to mention that there are two strategies for solving the multi-class classification problem (i.e., TN, TP, FP, and FN). The first one uses all combinations of the classes (e.g., for 3 classes the following combinations are compared: class 1 vs. class 2, class 1 vs. class 3 and class 2 vs. class 3). Accordingly, for n classes n!/2 comparison should be done. The second strategy compares one class vs.

Table 1
Summary of the identification results. Results are shown separately for the different methods (LDA, WKNN, SVM, NN, RF) and feature extractors. The right column reports those comparisons revealing no significant differences between the different classifiers. ** P-value < 0.001 for different classifiers and each feature extractor; * P-value < 0.001 for different feature extractors and each classifier.

		LDA*[a]	WKKN*[b]	SVM*[c]	NN*[d]	RF*[e]	Significant differences according to the McNemar Tests
CohIns**	ACC	1	1	1	1	1	SVM > WKN + NN + RF + LDA
	SENS	0.77	0.97	0.98	0.94	0.85	WKNN=NN > RF + LDA
	SPEC	1	1	1	1	1	RF > LDA
	F1	0.76	0.96	0.98	0.93	0.86	
CohLag**	ACC	0.99	0.99	0.98	0.99	1	LDA = RF > NN + WKKN + SVM
	SENS	0.65	0.49	0.01	0.54	0.43	NN > WKNN + SVM
	SPEC	1	1	0.99	1	1	
	F1	0.66	0.46	0	0.53	0.42	
CohTot**	ACC	1	1	1	1	1	SVM > LDA + WKNN + NN + RF
	SENS	0.77	0.97	0.98	0.95	0.83	WKNN > LDA + NN + RF
	SPEC	1	1	1	1	1	NN > LDA + RF
	F1	0.76	0.96	0.98	0.95	0.85	RF > LDA
CohCor**	ACC	1	1	1	1	1	NN > LDA + WKNN + SVM + RF
	SENS	0.73	0.87	0.95	0.98	0.88	SVM = RF > LDA + WKNN
	SPEC	1	1	1	1	1	
	F1	0.71	0.86	0.95	0.97	0.9	
CohMuI**	ACC	1	1	1	1	1	SVM=WKNN > LDA + NN + RF
	SENS	0.81	0.98	0.98	0.98	0.86	NN > LDA + RF
	SPEC	1	1	1	1	1	RF > LDA
	F1	0.81	0.98	0.98	0.97	0.87	
CohGrC**	ACC	1	1	1	1	1	WKNN=SVM=NN = RF > LDA
	SENS	0.98	1	1	1	1	
	SPEC	1	1	1	1	1	
	F1	0.98	1	1	1	1	
PhsIns**	ACC	1	1	1	1	1	SVM=NN=RF > LDA + WKNN
	SENS	0.72	0.77	0.92	0.94	0.82	NN > LDA + WKNN + RF
	SPEC	1	1	1	1	1	RF > LDA + WKNN
	F1	0.71	0.77	0.92	0.93	0.84	
PhsLag**	ACC	0.99	0.98	0.98	0.98	0.99	LDA > WKNN,SVM=NN,NN,RF
	SENS	0.24	0.05	0.01	0.01	0.06	WKNN > SVM + NN + RF
	SPEC	0.99	0.99	0.99	0.99	1	RF > SVM + NN
	F1	0.23	0.04	0	0	0.04	
PhsTot**	ACC	1	1	1	1	1	SVM=NN=RF > LDA + WKNN
	SENS	0.72	0.77	0.92	0.92	0.81	
	SPEC	1	1	1	1	1	
	F1	0.71	0.76	0.92	0.92	0.83	
PhsCor**	ACC	1	1	1	1	1	SVM=NN > LDA + WKNN + RF
	SENS	0.77	0.87	0.95	0.94	0.86	RF > LDA + WKNN
	SPEC	1	1	1	1	1	WKNN > LDA
	F1	0.76	0.86	0.96	0.93	0.88	
PhsMuI**	ACC	1	1	1	1	1	WKNN=SVM > LDA + NN + RF
	SENS	0.82	0.98	0.98	0.98	0.86	NN > LDA + RF
	SPEC	1	1	1	1	1	RF > LDA
	F1	0.82	0.98	0.98	0.97	0.88	
PhsGrC**	ACC	1	1	1	1	1	LDA > WKNN + SVM + NN + RF
	SENS	0.92	0.88	0.89	0.87	0.8	SVM > WKNN + NN + RF
	SPEC	1	1	1	1	1	WKNN = RF
	F1	0.92	0.88	0.9	0.86	0.81	

[a] Non-significant McNemar test results for LDA: *PhsMuI-CohMuI, CohIns-CohTot, CohTot-PhsCor, CohIns-PhsCor, CohCor-PhsIns, PhsIns-PhsTot, CohCor-PhsTot*. All others are significant.

[b] Non-significant McNemar test results for WKNN: *CohMuI-PhsMuI, CohMuI-CohTot, CohTot-CohIns, CohCor-PhsCor, PhsIns-PhsTot*. All others are significant.

[c] Non-significant McNemar test results for SVM: *CohIns-PhsMuI, CohIns-CohTot, CohTot-PhsMuI, CohIns-CohMuI, CohLag-PhsLag, PhsIns-PhsTot, CohTot-CohMuI, CohCor-PhsCor, CohMuI-PhsMuI*. All others are significant.

[d] Non-significant McNemar test results for NN: *CohCor-CohMuI, CohIns-CohTot, CohIns-CohMuI, CohTot-PhsCor, CohIns-PhsCor, CohIns-PhsMuI, PhsCor-PhsMuI, PhsIns-PhsTot*. All others are significant.

[e] Non-significant McNemar test results for RF: *CohIns-CohTot, CohCor-CohMuI, CohTot-PhsTot, CohMuI-PhsCor, CohIns-PhsMuI, CohIns-CohMuI, CohIns-PhsCor, PhsIns-PhsTot, CohMuI-PhsMuI, PhsCor-PhsMuI, CohTot-PhsIns*. All others are significant.

other classes. In this context, for n classes only n comparisons are necessary (e.g., for 3 classes the following combinations are compared: class 1 vs. class 2 and 3, class 2 vs. class 1 and 3, and class 3 vs. class 1 and 2). In our work we used the second strategy.

3. Results

3.1. Identification of single participants

The results of the identification procedures computed for the different classifiers are visible in Table 1. This table reports accuracy, sensitivity, specificity, and F1-scores. Accuracy values were in the range between 0.98 and 1 (mean = 0.99, SD = 0.001), sensitivity between 0.00 and 1 (mean = 0.66, SD = 0.087), specificity between 0.99 and 1 (mean = 0.99, SD = 0.002), and F1-scores between 0.00 and 1 (mean = 0.65, SD = 0.087). Furthermore, we compared the identification results between the classifiers separately for each connectivity measure using McNemar tests. The results are reported in the rightmost column of Table 1. In Fig. 1, the identification results (expressed as F1-scores) are shown separately for the different classifiers and connectivity measures. As one can see from this figure, most of the F1-scores were high (i.e., >0.9, according to our definition mentioned in the methods section), indicating a very good identification result. However, the results for lagged coherence (*CohLag*) and lagged phase (*PhsLag*) were not satisfactory. In the same Figure it becomes also visible that LDA classification varied from not satisfactory (*PhsLag*) to perfect (*CohGrC*). The detailed results of the between-classifier comparisons are reported in Table 1. This table shows that most classifiers differed, even when using corrected and thus conservative cut-off values. Fig. 1 provides an overview of the results. In summary, SVM and NN provided the best identification results, while LDA was the least good classifier. WKNN and RF were a bit better than LDA but partly less good than SVM and NN. However, the differences between SVM and NN vs. RF and WKNN were not that strong, although partly significant depending on the connectivity method used.

3.2. Identification of single participants using different sample sizes

Figs. 2 and 3 indicate the sample size effect for the different identification techniques and the different datasets. As a reminder, we have randomly eliminated a subset of subjects from our dataset to examine a putative influence of number of participants on identification results. As one can see from these Figures, even for different sample sizes the identification results were good for NN, WKNN, and SVM. Otherwise, the identification results became worse when sample size increased, especially for RF and LDA. This happened especially for the *CohLag* and *PhsLag* datasets (Fig. 2). For *CohGrC,* the results were always perfect (Fig. 2).

3.3. Identification of single participants using different artifact removal procedures

From our dataset, we selected 12 participants with two resting-state data sets in order to test for ICA-based influences. We only evaluated the data of 12 participants because the analyses for the entire sample of 119 subjects would take several months due to heavy computational demands. For these 12 participants, we performed two different types of artifact removals procedures, one with and one without ICA-based artifact correction. We conducted these additional analyses in order to exclude that ICA-based artifact removal might have influenced the coherence measures. In order to rule this out, we compared the coherences and identification results for the artifact corrected and uncorrected EEG data. Table 2 summarizes the identification results for both data sets. As one can see from this Table, the identifications are very similar for both datasets. The lowest F1-score was found for *PhsMuI* and the WKNN classifier (F1 = 0.75). Most of the F1-scores were around 0.9. There were some significant differences between the two datasets (Table 3).

3.4. Identification of single participants using different noise levels

Adding Gaussian noise to the connectivity matrices resulted in a substantial decrement of subject identification accuracy for most of the classifiers. Even by adding only 10% of noise to the entire matrix, we observed a substantial drop in identification accuracy (Fig. 4). The Granger causality measures turned out to be relatively stable across the different noise levels. *CohIns, CohTot, and CohMuI* were less good than the Granger causality measures but still good.

3.5. Identification of single subjects as a function of epoch lengths

By increasing the epoch length from 3 to 30 epochs of 2-s length (9 steps), we obtained 9 different training data sets. By applying the classification results obtained from these different training sets to the test data sets, we verified whether epochs' length might have an influence on the classification results. As visible in Fig. 5, for epoch length ≥ 30 s, the classification results were moderately good and did not change (at least by using LDA).

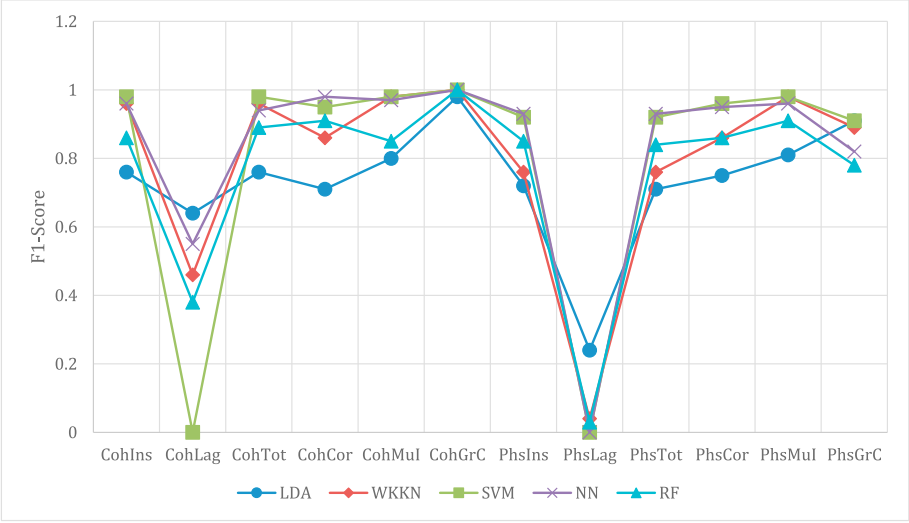

Fig. 1. Summary of the identification results. Results are shown separately for the different classifiers and connectivity measures and are expressed as F1-scores.

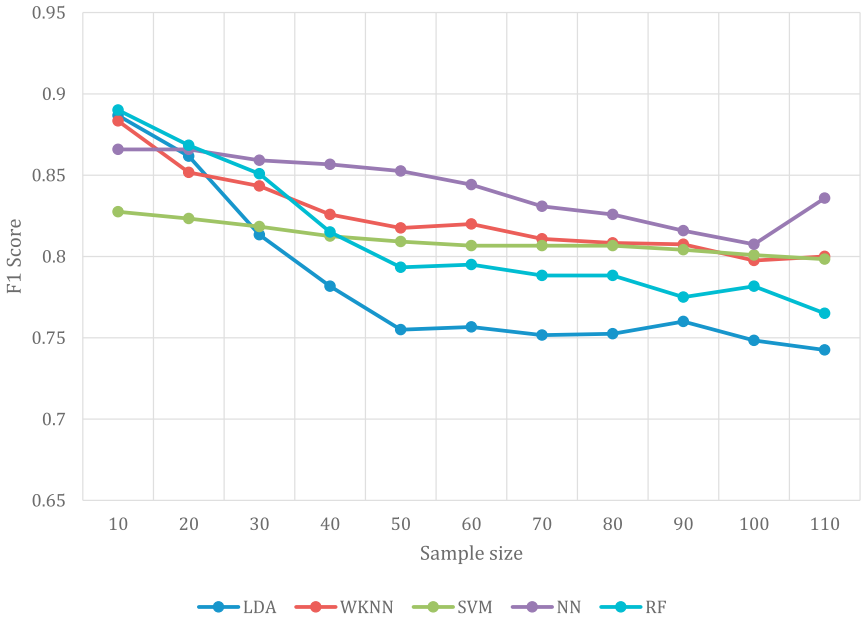

Fig. 2. Mean identification results. F1-scores are shown across all coherence measures for different sample sizes and the different classifiers.

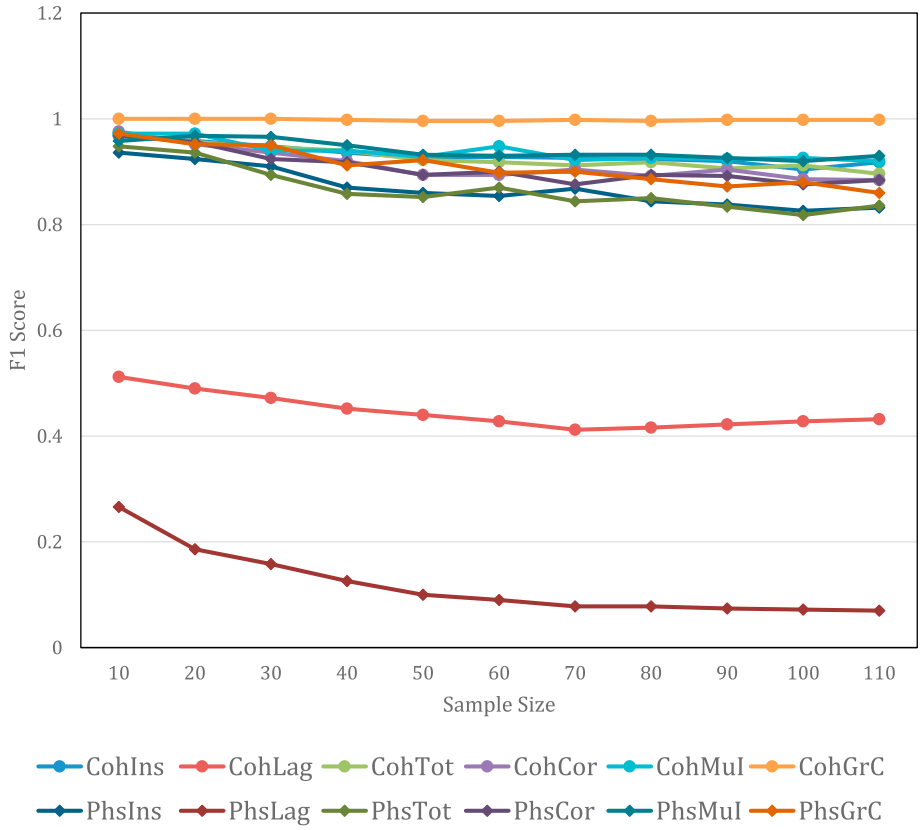

Fig. 3. Mean identification results. F1-scores across all classifiers in different sample sizes.

3.6. *Stability of identification over time*

In a next analysis, we tested whether EEG resting-state data of a single participant remained stable over a time period of 3 months. The identification rate of this participant in the entire sample ranged from moderate to perfect using the EEG data obtained at three different time points. Table 4 shows the identification results of this particular subject.

Finally, since the evaluation of a single participant does not enable generalizability of test-retest stability, we additionally evaluated BI in five participants (new sample from our databank) who were measured twice with a time lag of 6 months in-between the two measurement points. The results of these additional analyses are reported in Table 5.

Table 2
Summary of identification results. Results are shown separately for different methods and feature extractors as well as for ICA-based artifact corrected and uncorrected EEG data.

		ICA					NO ICA				
		LDA	WKKN	SVM	NN	RF	LDA	WKKN	SVM	NN	RF
CohIns	Accuracy	0.98	1.00	1.00	1.00	1.00	0.99	0.96	0.98	0.99	0.98
	sensitivity	0.91	0.99	1.00	1.00	0.97	0.91	0.76	0.91	0.93	0.87
	specificity	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.99	0.99	0.99
	F-score	0.91	0.99	1.00	1.00	0.97	0.91	0.76	0.91	0.93	0.87
CohLag	Accuracy	0.97	0.96	0.80	0.92	0.96	0.99	0.96	0.99	0.99	0.97
	sensitivity	0.86	0.78	0.14	0.53	0.80	0.91	0.79	0.91	0.93	0.85
	specificity	0.99	0.98	0.86	0.96	0.98	0.99	0.98	0.99	0.99	0.99
	F-score	0.86	0.78	0.03	0.53	0.80	0.91	0.79	0.91	0.93	0.85
CohTot	Accuracy	0.98	1.00	1.00	1.00	0.99	0.99	0.96	0.98	0.99	0.98
	sensitivity	0.89	0.98	0.99	1.00	0.97	0.91	0.77	0.91	0.93	0.87
	specificity	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.99	0.99	0.99
	F-score	0.89	0.98	0.99	1.00	0.97	0.91	0.78	0.91	0.93	0.87
CohCor	Accuracy	0.97	0.98	1.00	1.00	0.99	0.98	0.96	0.98	0.99	0.98
	sensitivity	0.87	0.89	0.99	1.00	0.93	0.90	0.78	0.90	0.94	0.86
	specificity	0.99	0.99	1.00	1.00	0.99	0.99	0.98	0.99	0.99	0.99
	F-score	0.87	0.89	0.99	1.00	0.93	0.90	0.78	0.90	0.93	0.86
CohMuI	Accuracy	0.92	0.87	0.80	0.85	0.94	0.99	0.96	0.98	0.99	0.98
	sensitivity	0.60	0.32	0.14	0.08	0.38	0.92	0.78	0.91	0.93	0.88
	specificity	0.96	0.93	0.86	0.93	1.00	0.99	0.98	0.99	0.99	0.99
	F-score	0.59	0.30	0.03	0.07	0.35	0.92	0.78	0.91	0.93	0.87
CohGrC	Accuracy	0.97	0.98	1.00	1.00	0.99	0.99	0.97	0.98	0.99	0.98
	sensitivity	0.86	0.89	0.99	0.99	0.95	0.92	0.79	0.91	0.94	0.78
	specificity	0.99	0.99	1.00	1.00	1.00	0.99	0.98	0.99	0.99	0.99
	F-score	0.86	0.89	0.99	0.99	0.95	0.92	0.79	0.91	0.94	0.78
PhsIns	Accuracy	0.98	0.99	1.00	1.00	0.99	0.99	0.96	0.98	0.99	0.98
	sensitivity	0.89	0.97	0.99	1.00	0.96	0.92	0.78	0.91	0.94	0.77
	specificity	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.99	1.00	1.00
	F-score	0.89	0.97	0.99	1.00	0.96	0.93	0.78	0.91	0.94	0.76
PhsLag	Accuracy	0.98	0.99	1.00	1.00	0.99	0.99	0.96	0.99	0.99	0.98
	sensitivity	0.88	0.94	0.99	0.99	0.97	0.91	0.78	0.91	0.94	0.87
	specificity	0.99	0.99	1.00	1.00	1.00	0.99	0.98	0.99	0.99	0.99
	F-score	0.89	0.94	0.99	0.99	0.97	0.92	0.78	0.91	0.94	0.87
PhsTot	Accuracy	0.99	1.00	1.00	1.00	0.99	0.99	0.97	0.98	0.99	0.98
	sensitivity	0.93	0.99	1.00	1.00	0.96	0.92	0.80	0.91	0.94	0.79
	specificity	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.99	0.99	1.00
	F-score	0.93	0.99	1.00	1.00	0.96	0.92	0.80	0.91	0.94	0.79
PhsCor	Accuracy	0.98	1.00	1.00	1.00	0.99	0.99	0.97	0.98	0.99	0.98
	sensitivity	0.89	0.99	1.00	0.99	0.96	0.92	0.80	0.91	0.94	0.86
	specificity	0.99	1.00	1.00	1.00	1.00	0.99	0.98	0.99	0.99	0.99
	F-score	0.89	0.99	1.00	0.99	0.96	0.92	0.79	0.91	0.94	0.85
PhsMuI	Accuracy	1.00	1.00	1.00	1.00	1.00	0.98	0.96	0.99	0.99	0.97
	sensitivity	1.00	1.00	1.00	1.00	1.00	0.90	0.75	0.92	0.92	0.84
	specificity	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.99	0.99	0.99
	F-score	1.00	1.00	1.00	1.00	1.00	0.91	0.75	0.92	0.92	0.84
PhsGrC	Accuracy	0.99	0.99	1.00	1.00	0.99	0.99	0.96	0.98	0.99	0.98
	sensitivity	0.96	0.96	0.97	0.99	0.93	0.92	0.79	0.91	0.93	0.86
	specificity	1.00	1.00	1.00	1.00	1.00	0.99	0.98	0.99	0.99	0.99
	F-score	0.96	0.96	0.97	0.99	0.83	0.92	0.80	0.91	0.93	0.86

The best results were obtained by using the *CohGrc* and *PhsGrc* measures. F1-scores were highest using RF and WKKN, followed by NN and SVM. By taking into account *CohGrc*, the lowest F1-scores were reached using LDA.

4. Discussion

In the present study, we tested the suitability of resting-state EEG data for BI by comparing different classifiers and connectivity metrics. With this purpose in mind, we focused on the temporal dynamics of coherence measures instead of evaluating temporally restricted spectrographic profiles at single electrodes or clusters of electrodes. The main reason for focusing on dynamic coherence measures was related to previous work showing that functional connectivity metrics at the sensor level can be used as a powerful approach for BI purposes with high identification accuracy (Schetinin et al., 2018). In our opinion, this approach is promising because functional and effective connectivity measures convey important information about the neural basis of cognition, emotion, and behavior (Fingelkurts et al., 2005). Furthermore, functional connectivity has been shown to reflect the underlying neural activation pattern with high individuality (Syed et al., 2017). Because rhythmic neuronal oscillations can be quantified using multiple metrics and each of them has its own advantages and disadvantages (Bastos and Schoffelen, 2016), we compared BI accuracy between different types of functional and effective connectivity measures. In this context, it is noteworthy to mention that even though we used well-known machine learning algorithms that have previously already been tested by other groups (Palaniappan and Mandic, 2007), our work exhibits several aspects of novelty. In particular, we (1) provide first evidence for the stability of BI in short EEG epochs, (2) compared BI between a wide range of commonly used functional connectivity metrics and machine learning algorithms, (3) focused on functional connectivity by taking into account the time domain, and (4) used different time points that were temporally segregated by a task for training and classification. In addition, (5) we assessed the reliability and stability of BI using several cross-validation procedures.

Results demonstrated high accuracy (i.e., F1-scores, specificity, and sensitivity) of resting-state metrics for BI and provided counterevidence

Table 3
Summary of statistical tests targeting at comparing identification results for ICA-based corrected and uncorrected datasets. The results are shown separately for different classifiers and coherence measures (*** < 0.001).

	LDA	WKNN	SVM	NN	RF
CohIns	0.831	***	***	***	***
CohLag	***	1.000	***	***	***
CohTot	1.000	***	***	***	***
CohCor	1.000	***	***	***	0.004
CohMuI	***	1.000	***	***	***
CohGrC	***	***	***	***	***
PhaseIns	1.000	***	0.002	***	***
PhaseLag	1.000	***	***	***	***
PhaseTot	1.000	***	***	***	***
PhaseCor	0.054	***	***	***	***
PhaseMU	***	***	***	***	***
PhaseCa	0.001	***	***	***	***

to the common belief that EEG data are too noisy for identification purposes with a high likelihood. Generally, BI was high (F1-scores between 0.71 and 1), with the exception of lagged coherence (*CohLag*) and lagged phase synchronization (*PhsLag*) metrics. Such high BI rates confirm that dynamic connectivity measures provide important individual-specific information that cannot be inferred from averaged spectral profiles (Syed et al., 2017). Furthermore, all classifiers yielded a high BI. However, LDA was less accurate (i.e., F1-scores in the range of 0.24–0.98) than the other algorithms, and this lower accuracy was particularly the case for *CohLag* and *PhsLag* connectivity metrics. Otherwise, SVM and NN resulted in the highest accuracy for these two specific connectivity measures, followed by RF and WKNN. Taken together, our results clearly demonstrate that the human brain is highly individual not only in terms of anatomy (Scheperjans et al., 2008; Uylings et al., 2005) but also function (Kanai and Rees, 2011; Pelofi et al., 2017). Accordingly, EEG coherence metrics can be used in a fruitful manner for BI with a high degree of specificity.

BI was paticularly high for those connectivity measures that are known to be affected by volume conduction (i.e., *CohIns, CohTot, CohCor, CohMul, CohGrC, PhsIns, PhsTot, PhsCor, PhsMul*, and *PhsGrC*) compared to those that are unaffected (i.e., *CohLag* and *PhsLag*) (Bastos and Schoffelen, 2016). Volume conduction refers to the transmission of electric fields through biological tissue from a primary current source within the brain (i.e., intracortical dipole) towards scalp electrodes (Nunez et al., 1997). Since the EEG signal measured at the surface of the scalp is not only dependent on intracortical dipole strength but also on the electrical transmission properties of brain tissues and compartments, connectivity metrics that are influenced by volume conduction indirectly reflect individual-specific biological and anatomical brain properties. Consequently, inter-individual differences in brain morphology have a stronger influence on the volume conduction-dependent connectivity metrics compared to those that are not influenced by volume propagation (Bastos and Schoffelen, 2016). Such functional-anatomical relationships can be deduced, for example, from professional musicians who are characterized by training-related structural changes in several brain regions (Bermudez et al., 2008; Elmer et al., 2012, 2013; Habibi et al., 2017; Münte et al., 2002; Schlaug, 2015), and at the same time demonstrate increased whole-brain instantaneous coherence (*CohIns*) during resting-state compared to non-musicians (Klein et al., 2016). By contrast, the connectivity measures that are not contaminated by volume conduction (i.e., *CohLag* and *PhsLag*) were less suited for BI purposes, possibly because they convey a lower degree of information about the underlying brain properties. These connectivity metrics are often referred to as "true physiological connectivity" measures (Pascual-Marqui et al., 2011) because they represent the intrinsic neurophysiological activity of the human brain. However, these connectivity profiles are also characterized by a high degree of fast neurophysiological fluctuations that lead to increased intra-individual variance and reduce the likeliness of BI. Such highly variable coherence measures are even more

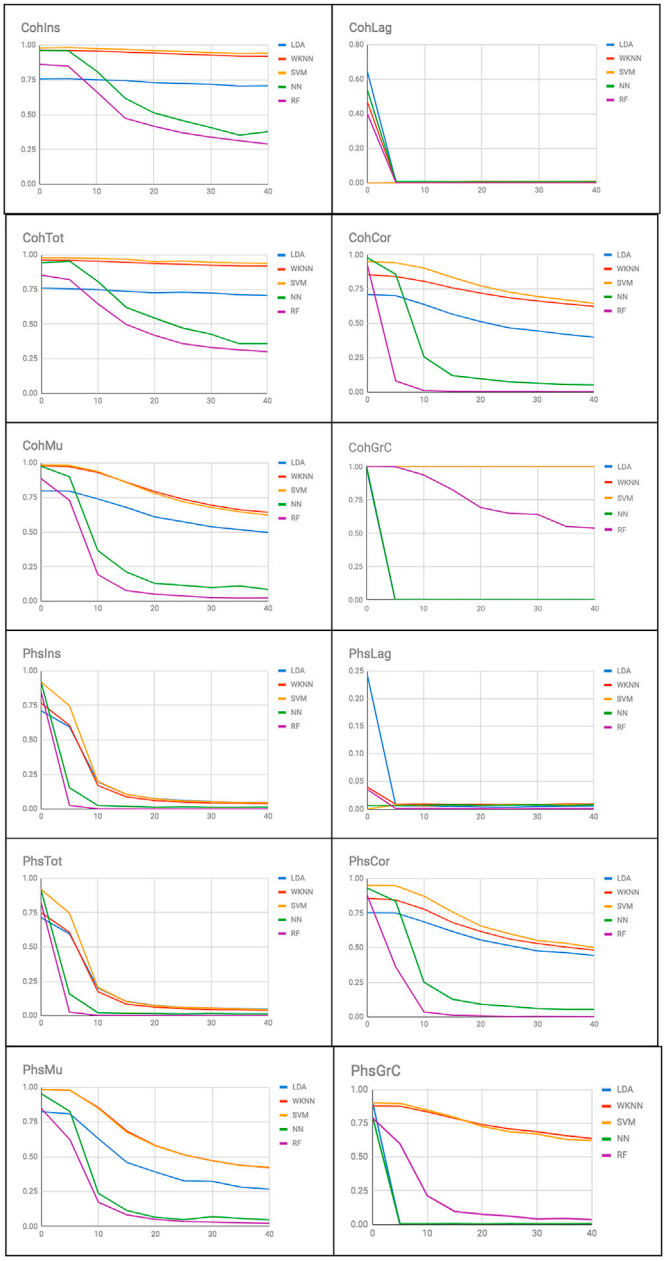

Fig. 4. Identification results for different noise levels and all connectivity types. The X-axis represents the percentage of noise and the Y-axis depicts F1-Scores.

accentuated during resting-state conditions, where cognitive and emotional states fluctuate more strongly than during a structured task condition. The different BI accuracy we revealed between connectivity metrics as a function of volume conduction may pave the way for new future applications. For example, the more "stable" (i.e., useful for subject identification) coherence measures could be used as biomarkers for psychiatric disorders, whereas the more "variable" (i.e., useless for subject identification) ones are possibly more eligible candidates for assessing neural dynamics underlying transient cognitive and emotional events. The next step for future research will be to focus on intracortical coherences that will provide more information about "true" physiological information. Such an intracortical approach could also help to identify which brain areas are particularly involved in generating stable and individual coherences.

In the present work, we also used different cross-validation methods consisting of (a) varying sample size, (b) adding Gaussian noise, (c) using different epoch durations for learning, (d) comparing ICA-corrected with

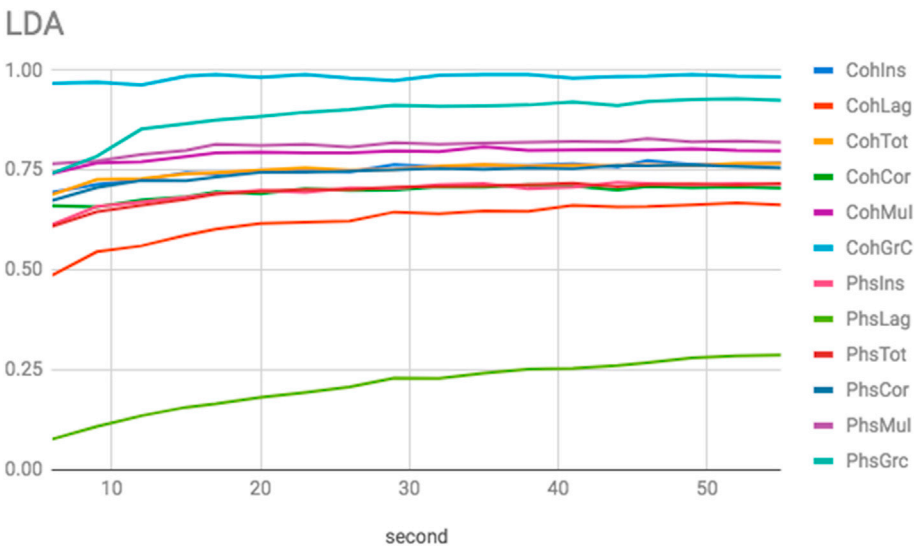

Fig. 5. Identification results in the learning data sets as a function of epochs duration and connectivity type (LDA).

Table 4
Identification of a single participant over a time period of three months. The results are shown separately for the different classifiers and connectivity metrics.

	LDA	WKKN	SVM	NN	RF
CohIns	0.89	1.00	1.00	0.92	0.14
CohLag	0.80	0.84	0.00	0.90	0.40
CohTot	0.87	0.94	0.98	0.86	0.89
CohCor	0.71	0.85	0.95	0.96	0.92
CohMuI	1.00	1.00	1.00	1.00	0.30
CohGrC	0.99	1.00	1.00	1.00	1.00
PhsIns	1.00	0.76	0.92	0.78	0.89
PhsLag	1.00	0.17	0.50	0.35	0.29
PhsTot	1.00	0.77	0.93	0.86	0.88
PhsCor	0.93	0.86	0.95	0.85	0.92
PhsMuI	1.00	0.97	0.98	0.89	0.89
PhsGrC	0.94	0.62	0.94	0.78	0.28

Table 5
Identification of 5 participants over a time period of six months. The results are shown separately for the different classifiers and connectivity metrics.

	LDA	WKKN	SVM	NN	RF
CohIns	0.19	0.20	0.21	0.15	0.21
CohLag	0.19	0.22	0.33	0.17	0.17
CohTot	0.19	0.22	0.22	0.21	0.21
CohCor	0.20	0.19	0.20	0.15	0.24
CohMuI	0.21	0.18	0.20	0.20	0.22
CohGrC	0.92	1.00	0.95	0.99	1.00
PhsIns	0.21	0.21	0.23	0.21	0.25
PhsLag	0.29	0.23	0.33	0.19	0.28
PhsTot	0.23	0.18	0.25	0.21	0.22
PhsCor	0.19	0.18	0.22	0.21	0.26
PhsMuI	0.23	0.21	0.25	0.19	0.22
PhsGrC	0.95	0.93	0.95	0.99	0.90

uncorrected data, (e) applying different tasks in between two resting-state periods, (f) repeatedly measuring a single individual across three weeks, and (g) repeatedly measuring five additional participants in order to determine BI stability after a time period of six months. Results revealed that (a) the SVM classifier was robust across different sample sizes, whereas LDA and RF showed the highest identification drop with increased sample size. Furthermore, the different functional connectivity metrics seem to be relatively unaffected by sample size, even though *CohGrc* and *PhsGrC* demonstrated the highest stability. By adding Gaussian noise to the features (b), we generally observed a rapid decrease of F1-scores. However, interestingly, for some combinations of classifiers and feature extractors, F1-scores remained relatively stable even after adding 40% of noise. This stability was particularly the case for *CohIns, CohTot, and CohGrC* in association with WKNN and SVM. Consequently, these combinations of classifiers and feature extractors may be particularly suited for BI in several spheres of everyday life. Notably, the evaluation of the number of EEG epochs required for an accurate BI (c) revealed that after only 30 s F1-scores were highest and remained almost stable. This result is particularly interesting in that it suggests that EEG metrics are especially suited for fast BI purposes in a time period of only 30 s. In a further hierarchical step, we verified the influence of ICA procedures on BI (d). Thereby, we revealed that when ICA was not used, F1-scores decreased. However, this reduction was not higher than 10%. This result was probably related to the fact that when it comes to performing analyses without ICA, there are still more artifacts (i.e., noise) in the data compared to signals that have been cleaned with a combination of ICA and automatic raw data inspection. Our results also demonstrated that a specific task inserted in between two resting-state periods did not affect F1-scores (e), indicating that BI is more strongly dependent upon other variables (i.e., sample size, feature extractors, and classifiers). We also evaluated the stability of BI across a period of three weeks by repeatedly measuring a single individual engaged in different tasks (i.e., listening to music, solving a math problem, mobile gaming, etc.) in between two successive resting-state periods (f). In this context, we randomly selected one minute from a single session for testing (i.e., randomly selected) and attempted to identify this individual among all the others. This additional cross-validation procedure revealed a high stability of BI across repeated measurements. Furthermore, BI was unaffected by different cap sizes, varying tasks inserted in between successive resting-state periods, or time-point of measurement and caffeine consumption. Finally, (g) we were even able to demonstrate a high test-retest BI stability in five participants over a time period of six months. In this context, it is noteworthy to mention that in contrast to previous work (Crobe et al., 2016; Fraschini et al., 2015; La Rocca et al., 2014) we also used effective connectivity metrics (*CohGrc* and *PhsGrc*). Notably, effective connectivity metrics demonstrated the highest stability over time, even after adding 40% of Gaussian noise. This additional result opens novel perspectives for BI applications across longer time periods.

Finally, it is important to mention that in the present work we also tested short- and long-term stability of BI by (1) inserting a task between the resting-state periods used for training and classification, (2) repeatedly measuring a single individual in a time period of three weeks before

and after cognitive demanding activities, and (3) re-measuring 5 participants after a time period of 6 month. Based on previous work showing that cognitive engagement can lead to rapid dynamic functional network reconfigurations (Braun et al., 2015), we reasoned that possibly BI will be less accurate if training and classification data sets are separated by a task or collected at different time points. However, according to our results, we did not reveal a strong influence of measurement time point on BI. This additional result underlines the suitability of EEG for BI purposes, at least within a time period of six months.

Author contribution statement

S.A.V. conducted all computations, wrote the scripts for subject identification, participated in writing the manuscript, and prepared the figures. S.E., L.J., and S.A.V. wrote the paper and R.R and S.E. supervised the entire project.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgement

This project was funded by the SNF (Swiss National Science Foundation) through two grants to LJ (Sinergia-Grant #136249 and grant # 320030_163149).

References

Annett, M., 1970. A classification of hand preference by association analysis. Br. J. Psychol. 61, 303–&.
Barry, R.J., Clarke, A.R., Hajos, M., McCarthy, R., Selikowitz, M., Dupuy, F.E., 2010. Resting-state EEG gamma activity in children with attention-deficit/hyperactivity disorder. Clin. Neurophysiol. 121, 1871–1877.
Bastos, A.M., Schoffelen, J.M., 2016. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. Front. Syst. Neurosci. 9.
Battiti, R., 1992. First-and second-order methods for learning: between steepest descent and Newton's method. Neural Comput. 4, 141–166.
Bermudez, P., Lerch, J.P., Evans, A.C., Zatorre, R.J., 2008. Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. Cerebr. Cortex 19, 1583–1596.
Braun, U., Schäfer, A., Walter, H., Erk, S., Romanczuk-Seiferth, N., Haddad, L., Schweiger, J.I., Grimm, O., Heinz, A., Tost, H., 2015. Dynamic reconfiguration of frontal brain networks during executive cognition in humans. Proc. Natl. Acad. Sci. Unit. States Am. 112, 11678–11683.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Chicco, D., 2017. Ten quick tips for machine learning in computational biology. BioData Min. 10.
Cover, T.M., Hart, P.E., 1967. Nearest neighbor pattern classification. Inform. Theory, IEEE Trans. 13, 21–27.
Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge university press.
Crobe, A., Demuru, M., Didaci, L., Marcialis, G.L., Fraschini, M., 2016. Minimum spanning tree and k-core decomposition as measure of subject-specific EEG traits. Biomed. Phys. Eng. Exp. 2, 017001.
Davidson, R.J., 2003. Affective neuroscience and psychophysiology: toward a synthesis. Psychophysiology 40, 655–665.
DelPozo-Banos, M., Travieso, C.M., Weidemann, C.T., Alonso, J.B., 2015. EEG biometric identification: a thorough exploration of the time-frequency domain. J. Neural Eng. 12, 056019.
Dittinger, E., Barbaroux, M., D'Imperio, M., Jancke, L., Elmer, S., Besson, M., 2016. Professional music training and novel word learning: from faster semantic encoding to longer-lasting word representations. J. Cogn. Neurosci. 28, 1584–1602.
Elmer, S., Greber, M., Pushparaj, A., Kuhnis, J., Jancke, L., 2017. Faster native vowel discrimination learning in musicians is mediated by an optimization of mnemonic functions. Neuropsychologia 104, 64–75.
Elmer, S., Hanggi, J., Meyer, M., Jancke, L., 2013. Increased cortical surface area of the left planum temporale in musicians facilitates the categorization of phonetic and temporal speech sounds. Cortex 49, 2812–2821.
Elmer, S., Meyer, M., Jancke, L., 2012. Neurofunctional and behavioral correlates of phonetic and temporal categorization in musically trained and untrained subjects. Cerebr. Cortex 22, 650–658.
Fingelkurts, A.A., Fingelkurts, A.A., Kahkonen, S., 2005. Functional connectivity in the brain - is it an elusive concept? Neurosci. Biobehav. Rev. 28, 827–836.
Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., Constable, R.T., 2015. Functional Connectome Fingerprinting: Identifying Individuals Using Patterns of Brain Connectivity. Nat Neurosci advance online publication.
Fraschini, M., Hillebrand, A., Demuru, M., Didaci, L., Marcialis, G.L., 2015. An EEG-based biometric system using eigenvector centrality in resting state brain networks. IEEE Signal Process. Lett. 22, 666–670.
Gage, F.H., Muotri, A.R., 2012. What makes each brain unique. Sci. Am. 306, 26–31.
Gordon, E.M., Devaney, J.M., Bean, S., Vaidya, C.J., 2015. Resting-state striato-frontal functional connectivity is sensitive to DAT1 genotype and predicts executive function. Cerebr. Cortex 25, 336–345.
Gui, Q., Jin, Z., Blondet, M.V.R., Laszlo, S., Xu, W., 2015. Towards EEG biometrics: pattern matching approaches for user identification. In: Identity, Security and Behavior Analysis (ISBA), 2015 IEEE International Conference on. IEEE, pp. 1–6.
Habibi, A., Damasio, A., Ilari, B., Veiga, R., Joshi, A.A., Leahy, R.M., Haldar, J.P., Varadarajan, D., Bhushan, C., Damasio, H., 2017. Childhood music training induces change in micro and macroscopic brain structure: results from a longitudinal study. Cerebr. Cortex 1–12.
Holm, S., 1979. A simple sequentially rejective multiple test procedure. Scand. J. Stat. 6, 65–70.
Huang, X., Altahat, S., Tran, D., Sharma, D., 2012. Human identification with electroencephalogram (EEG) signal processing. In: Communications and Information Technologies (ISCIT), 2012 International Symposium on. IEEE, pp. 1021–1026.
Jain, A., Hong, L., Pankanti, S., 2000. Biometric identification. Commun. ACM 43, 90–98.
Jancke, L., Rogenmoser, L., Meyer, M., Elmer, S., 2012. Pre-attentive modulation of brain responses to tones in coloured-hearing synesthetes. BMC Neurosci. 13.
Jeong, J., Gore, J.C., Peterson, B.S., 2001. Mutual information analysis of the EEG in patients with Alzheimer's disease. Clin. Neurophysiol. 112, 827–835.
John, E.R., Prichep, L.S., Fridman, J., Easton, P., 1988. Neurometrics: computer-assisted differential diagnosis of brain dysfunctions. Science 239, 162–169.
Kam, J.W., Bolbecker, A.R., O'Donnell, B.F., Hetrick, W.P., Brenner, C.A., 2013. Resting state EEG power and coherence abnormalities in bipolar disorder and schizophrenia. J. Psychiatr. Res. 47, 1893–1901.
Kanai, R., Rees, G., 2011. The structural basis of inter-individual differences in human behaviour and cognition. Nat. Rev. Neurosci. 12, 231–242.
Khalifa, W., Salem, A., Roushdy, M., Revett, K., 2012. A survey of EEG based user authentication schemes. In: Informatics and Systems (INFOS), 2012 8th International Conference on. BIO-55-BIO-60.
Klein, C., Liem, F., Hanggi, J., Elmer, S., Jancke, L., 2016. The "silent" imprint of musical training. Hum. Brain Mapp. 37, 536–546.
Kostilek, M., St'astny, J., 2012. EEG biometric identification: repeatability and influence of movement-related EEG. In: Applied Electronics (AE), 2012 International Conference on, pp. 147–150.
Kuhnis, J., Elmer, S., Meyer, M., Jancke, L., 2013. The encoding of vowels and temporal speech cues in the auditory cortex of professional musicians: an EEG study. Neuropsychologia 51, 1608–1618.
Kumari, V., ffytche, D.H., Williams, S.C., Gray, J.A., 2004. Personality predicts brain responses to cognitive demands. J. Neurosci. 24, 10636–10641.
La Rocca, D., Campisi, P., Vegso, B., Cserti, P., Kozmann, G., Babiloni, F., Fallani, F.D.V., 2014. Human brain distinctiveness based on EEG spectral coherence connectivity. IEEE Trans. Biomed. Eng. 61, 2406–2412.
Lan, M., Minett, J.W., Blu, T., Wang, W.S., 2015. Resting State EEG-based biometrics for individual identification using convolutional neural networks. Conf. Proc. IEEE Eng. Med. Biol. Soc. 2015, 2848–2851.
LeRoy, S., 2004. Causality in Economics. London School of Economics, Centre for Philosophy of Natural and Social Sciences.
Majewski, I.J., Bernards, R., 2011. Taming the dragon: genomic biomarkers to individualize the treatment of cancer. Nat. Med. 17, 304–312.
Marcel, S., Millan, J.D.R., 2007. Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. IEEE Trans. Pattern Anal. Mach. Intell. 29, 743–748.
Michel, C.M., 2009. Electrical Neuroimaging. Cambridge University Press.
Mu, Z.D., Hu, J.F., 2011. Research on EEG Identification Computing Based on Photo Images. Advanced Materials Research. Trans Tech Publ, pp. 1366–1371.
Münte, T.F., Altenmüller, E., Jäncke, L., 2002. The musician's brain as a model of neuroplasticity. Nat. Rev. Neurosci. 3, 473.
Na, S.H., Jin, S.H., Kim, S.Y., Ham, B.J., 2002. EEG in schizophrenic patients: mutual information analysis. Clin. Neurophysiol. 113, 1954–1960.
Ng, S.W., Mitchell, A., Kennedy, J.A., Chen, W.C., McLeod, J., Ibrahimova, N., Arruda, A., Popescu, A., Gupta, V., Schimmer, A.D., Schuh, A.C., Yee, K.W., Bullinger, L., Herold, T., Gorlich, D., Buchner, T., Hiddemann, W., Berdel, W.E., Wormann, B., Cheok, M., Preudhomme, C., Dombret, H., Metzeler, K., Buske, C., Lowenberg, B., Valk, P.J., Zandstra, P.W., Minden, M.D., Dick, J.E., Wang, J.C., 2016. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature 540, 433–437.
Nunez, P.L., Srinivasan, R., Westdorp, A.F., Wijesinghe, R.S., Tucker, D.M., Silberstein, R.B., Cadusch, P.J., 1997. EEG coherency. I: statistics, reference electrode, volume conduction, Laplacians, cortical imaging, and interpretation at multiple scales. Electroencephalogr. Clin. Neurophysiol. 103, 499–515.
Pascual-Marqui, R.D., 2007. Instantaneous and Lagged Measurements of Linear and Nonlinear Dependence between Groups of Multivariate Time Series: Frequency Decomposition arXiv preprint arXiv:0711.1455.
Pascual-Marqui, R.D., Lehmann, D., Koukkou, M., Kochi, K., Anderer, P., Saletu, B., Tanaka, H., Hirata, K., John, E.R., Prichep, L., Biscay-Lirio, R., Kinoshita, T., 2011. Assessing interactions in the brain with exact low-resolution electromagnetic tomography. Phil. Trans. Math. Phys. Eng. Sci. 369, 3768–3784.

Pascualmarqui, R.D., Michel, C.M., Lehmann, D., 1995. Segmentation of brain electrical-activity into microstates - model estimation and validation. IEEE Trans. Biomed. Eng. 42, 658–665.
Pelofi, C., de Gardelle, V., Egre, P., Pressnitzer, D., 2017. Interindividual variability in auditory scene analysis revealed by confidence judgements. Philos. Trans. R. Soc. Lond. B Biol. Sci. 372.
Putman, P., 2011. Resting state EEG delta-beta coherence in relation to anxiety, behavioral inhibition, and selective attentional processing of threatening stimuli. Int. J. Psychophysiol. 80, 63–68.
Scheperjans, F., Eickhoff, S.B., Homke, L., Mohlberg, H., Hermann, K., Amunts, K., Zilles, K., 2008. Probabilistic maps, morphometry, and variability of cytoarchitectonic areas in the human superior parietal cortex. Cerebr. Cortex 18, 2141–2157.
Schetinin, V., Jakaite, L., Nyah, N., Novakovic, D., Krzanowski, W., 2018. Feature extraction with GMDH-type neural networks for EEG-based person identification. Int. J. Neural Syst. 28, 1750064.
Schlaug, G., 2015. Musicians and music making as a model for the study of brain plasticity. Music, Neurol. Neurosci.: Evolution, the Musical Brain, Medical Conditions, and Therapies 217, 37–55.
Shedeed, H.A., 2011. A new method for person identification in a biometric security system based on brain EEG signal processing. In: Information and Communication Technologies (WICT), 2011 World Congress on, pp. 1205–1210.
Stam, C.J., Montez, T., Jones, B.F., Rombouts, S.A., van der Made, Y., Pijnenburg, Y.A., Scheltens, P., 2005. Disturbed fluctuations of resting state EEG synchronization in Alzheimer's disease. Clin. Neurophysiol. 116, 708–715.
Strobl, C., Malley, J., Tutz, G., 2009. An introduction to recursive partitioning: rationale, application and characteristics of classification and regression trees, bagging and random forests. Psychol. Methods 14, 323–348.
Su, F., Xia, L., Cai, A., Wu, Y., Ma, J., 2010. EEG-based personal identification: from proof-of-concept to a practical system. In: Pattern Recognition (ICPR), 2010 20th International Conference on. IEEE, pp. 3728–3731.
Syed, M.F., Lindquist, M.A., Pillai, J.J., Agarwal, S., Gujar, S.K., Choe, A.S., Caffo, B., Sair, H.I., 2017. Dynamic functional connectivity states between the dorsal and ventral sensorimotor networks revealed by dynamic conditional correlation analysis of resting-state functional magnetic resonance imaging. Brain Connect. 7, 635–642.
Thatcher, R.W., North, D., Biver, C., 2005. EEG and intelligence: relations between EEG coherence, EEG phase delay and power. Clin. Neurophysiol. 116, 2129–2141.
Uylings, H.B., Rajkowska, G., Sanz-Arigita, E., Amunts, K., Zilles, K., 2005. Consequences of large interindividual variability for human brain atlases: converging macroscopical imaging and microscopical neuroanatomy. Anat. Embryol. 210, 423–431.
Valizadeh, S.A., Liem, F., Mérillat, S., Hänggi, J., Jäncke, L., 2018 Apr 4. Identification of individual subjects on the basis of their brain anatomical features. Sci Rep. 8 (1), 5611. https://doi.org/10.1038/s41598-018-23696-6.
Venables, N.C., Bernat, E.M., Sponheim, S.R., 2009. Genetic and disorder-specific aspects of resting state EEG abnormalities in schizophrenia. Schizophr. Bull. 35, 826–839.
Wachinger, C., Golland, P., Reuter, M., 2014. BrainPrint: Identifying Subjects by Their Brain. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014. Springer, pp. 41–48.
Wang, J., She, M., Nahavandi, S., Kouzani, A., 2010. A review of vision-based gait recognition methods for human identification. In: Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on. IEEE, pp. 320–327.
Wang, Y., Veluvolu, K.C., Cho, J.H., Defoort, M., 2012. Adaptive estimation of EEG for subject-specific reactive band identification and improved ERD detection. Neurosci. Lett. 528, 137–142.
Wettschereck, D., Aha, D.W., Mohri, T., 1997. A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. Artif. Intell. Rev. 11, 273–314.
Woltering, S., Jung, J., Liu, Z., Tannock, R., 2012. Resting state EEG oscillatory power differences in ADHD college students and their peers. Behav. Brain Funct. 8, 60.
Yeh, F.C., Vettel, J.M., Singh, A., Poczos, B., Grafton, S.T., Erickson, K.I., Tseng, W.I., Verstynen, T.D., 2016. Quantifying differences and similarities in whole-brain white matter architecture using local connectome fingerprints. PLoS Comput. Biol. 12, e1005203.
Zhao, Q., Peng, H., Hu, B., Liu, Q., Liu, L., Qi, Y., Li, L., 2010. Improving individual identification in security check with an EEG based biometric solution. Brain Informatics 6334, 145–155.