

Multiway Array Decomposition Analysis of EEGs in Alzheimer's Disease

Charles-Francois V. Latchoumane^a, Francois-Benois Vialatte^b, Jordi Solé-Casals^c, Monique Maurice^b, Sunil R. Wimalaratna^d, Nigel Hudson^e, Jaeseung Jeong^{a,*} and Andrzej Cichocki^b

^aDepartment of Bio and Brain Engineering, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea 305-701

^bLab for Advanced Brain Signal Processing, Brain Science Institute, RIKEN, 2-1, Hirosawa, Wako Saitama, Japan, 351-0198

^cUniversity of Vic, Barcelona, Spain

^dJohn Radcliffe Infirmary, Oxford, UK

^eDepartment of Clinical Neurophysiology, Derriford Hospital, Plymouth, PL6 8DH, UK

Abstract— Methods for the extraction of features from physiological datasets are growing needs as clinical investigations of Alzheimer's disease (AD) in large and heterogeneous population increase. General tools allowing diagnostic regardless of recording sites, such as different hospitals, are essential and if combined to inexpensive non-invasive methods could critically improve mass screening of subjects with AD. In this study, we applied three state of the art multiway array decomposition (MAD) methods to extract features from electroencephalograms (EEGs) of AD patients obtained from multiple sites. In comparison to MAD, spectral-spatial average filter (SSFs) of control and AD subjects were used as well as a common blind source separation method, algorithm for multiple unknown signal extraction (AMUSE). We trained a feed-forward multilayer perceptron (MLP) to validate and optimize AD classification from two independent databases. Using a third EEG dataset, we demonstrated that features extracted from MAD outperformed features obtained from SSFs AMUSE in terms of root mean squared error (RMSE) and reaching up to 100% of accuracy in test condition. We propose that MAD maybe a useful tool to extract features for AD diagnosis offering great generalization across multi-site databases and opening doors to the discovery of new characterization of the disease.

Index Terms— Alzheimer's disease, EEG, diagnosis, multiway array decomposition, PARAFAC, NTD

1. INTRODUCTION

Alzheimer's disease (AD) is the most prevalent form of progressive degenerative dementia, affecting 5-10% of the population above the age of 65 (Ferri et al., 2006). The disease's clinical symptoms are characterized by progressive amnesia, followed by a gradual but persistent decline in all cognitive domains that results in global dementia. The neuropathology of AD includes widespread neuronal cell loss, neurofibrillary tangles (such as aggregation of tau proteins), and senile plaques (including extracellular aggregates of amyloid β -peptides) starting in the entorhinal cortex and limbic areas during the early stages of the disease, followed by spreading to other parts of the cortex (Braak and Braak, 1991). Recently, several studies have suggested that pharmacological treatment of mild cognitive impairment and early AD can slow the development of the disease (Hashimoto et al., 2005; Pietrzik and Behl, 2005). Thus, accurate early diagnosis of AD is an important issue in both psychiatry and neurology.

Numerous clinical methods have been used extensively to diagnose AD, including neuroimaging techniques (Alexander, 2002; Deweer et al., 1995; Waldemar et al., 2005; Wang et al., 2006), physiological markers (Andreasen et al., 2001), and genetic analyses (Tanzi and Bertram, 2001). Among these, the electroencephalogram (EEG) allows low-cost mass screening and may be combined with cognitive tests like the mini-mental score evaluation (MMSE) for AD diagnosis and symptom quantification (van der Hiele et al., 2007). EEG-based classification studies of AD using spectral or nonlinear methods have shown about 70-95% diagnostic accuracy for various degrees of disease severity, such as mild cognitive impairment (MCI) and mild, moderate, and severe stages of AD. The current application of classification in disease diagnosis is still lacking consistency to reveal their true power for real clinical applications. Part of the issue and limitations in previous classifications resides in the usage of databases for validation and test as most of these studies used test EEG datasets that were identical to the training datasets (Huang et al., 2000; Musha et al., 2002; Pritchard et al., 1994) or subset of the same database (Besthorn et al., 1997; Buscema et al., 2007; Cichocki et al., 2005; Woon et al., 2007). Some groups have reported about 90% accurate classification of mild AD against healthy subjects

*correspondance author. Tel: +82-42-350-4319; fax: +82-42-350-4310.

E-mail: jsjeong@kaist.ac.kr (J. Jeong).

URL: raphe.kaist.ac.kr/people_01.htm

using resting state EEGs (Buscema et al., 2007; Lehmann et al., 2007; Rossini et al., 2008), yet these studies utilized methods that were tuned and targeted to particular datasets, often had too large sets of features and classifiers either too complex (i.e. nonlinear), introducing a bias toward the data's specific distribution (Haykin, 1994). Few studies have reported AD classification using EEGs recorded at multiple sites that possess different and independent training, validation, and test datasets (Musha et al., 2002; Rossini et al., 2008; Woon et al., 2007). An AD classification approach using independent EEG databases may provide a more clinically applicable and accurate tool for AD diagnosis than biased model selection and database over fitting.

The hallmarks of EEG abnormalities in AD patients are slowed rhythms and a decrease in coherence among different brain regions (see (Dauwels et al., 2010; Jeong, 2004) for extensive review). An increase in theta and delta activity and a decrease in alpha and beta activity are often observed (Coben et al., 1985; Coben et al., 1983; Giaquinto and Nolfe, 1986; Pijnenburg et al., 2004) and reduced coherence of the alpha and beta bands is frequently found (Adler et al., 2003; Dunkin et al., 1994; Locatelli et al., 1998). Moreover, these abnormalities are correlated with disease severity (Kowalski et al., 2001; Schreiter-Gasser et al., 1994). EEGs in AD patients, however, usually show high inter-subject variability, particularly when patient groups exhibiting different severities and comorbidities are considered (Nobili et al., 1999; Pucci et al., 1999). Due to this significant variability, it is difficult to determine general features among large AD patient populations from different hospitals.

We sought to apply multiway array decomposition (MAD) analysis to the EEGs of patients with mild AD and of age- and sex-matched healthy subjects to extract signature features for disease diagnosis. MAD is an unsupervised method for investigating the information content of multidimensional and multivariate physiological data like EEGs, which for example, track time, frequency, and electrode domains simultaneously. Meanwhile, methods such as principal component analysis (PCA) (Cichocki et al., 2005) and independent component analysis (ICA) (Melissant et al., 2005) only capture structures within two-way multivariate matrices (e.g., temporal and spatial features, spatial and spectral features). MAD is particularly suitable for analyzing multichannel EEG data, such as third-order tensors (*Subjects* \times *Frequency* \times *Electrodes*) exhibiting multiway linkages and interactions between subjects and characteristics in EEGs recorded at different hospitals, which cannot be accomplished using conventional one- or two-way analysis. MAD methods have been applied to EEGs to examine drug effects (Estienne et al., 2001), to detect multiple cognitive states (Lee et al., 2007; Miwakeichi et al., 2004), to determine artifacts (Acar et al., 2007; Miwakeichi et al., 2004), and to identify seizure foci in epileptic patients (Acar et al., 2007; Acar et al., 2006) in the time-frequency-electrode space.

In this study, we used two MAD methods including parallel factor analysis (PARAFAC) and non-negative tensor decomposition (NTD) to decompose higher-order arrays and reveal the common signatures of AD. Since we mainly used space and frequency as discriminative features, we compared our analysis to a simple group average based on spectral-spatial filters (SSFs) as well as more common two-way decomposition method: algorithm for multiple unknown signals extraction (AMUSE). We used the weights of each extracted component obtained after application of MAD, SSFs or AMUSE as features, and we compared AD and healthy subjects with a feed-forward multilayer perceptron (MLP). We trained, validated, and optimized the feed-forward MLP for classification of AD using two independent EEG databases, only retaining most accurate and least complex models, i.e. least number of hidden neurons. We also tested the performance of the MLP using a third independent EEG database to determine the general applicability of MAD to clinical situations. Finally, we demonstrated that this approach could offer a plausible physiological interpretation of the spectral-spatial properties of AD patients and healthy subjects.

2. METHODS

2.1. Three EEG Databases of AD and Healthy Subjects

EEGs were recorded in three different countries of the European Union (EU) for patients at rest and with their eyes closed. The number of electrodes and the sampling frequencies of the three EEG databases were all different. These three databases for AD patients and healthy subjects were produced using BIOPATTERN (University of Plymouth, Plymouth, UK), an EU project aimed at developing a generalized, clinical EEG model for AD diagnosis and symptom quantification by analyzing subject bio-profiles and clinical measures recorded from hospitals in different EU countries (Goh et al., 2007; Hamadicharef et al., 2008). The local institutional ethics committees approved this research and informed consent was obtained from all subjects and caregivers prior to recording and experimentation. These EEG datasets have been used previously (Goh et al., 2006) and detailed information regarding the three EEG databases is as follows:

2.1.1. Database A

This EEG database was composed of 24 healthy subjects (aged 69.4 ± 11.5 years; 10 males) and 17 patients diagnosed with a mild form of Alzheimer's disease (aged 77.6 ± 10.0 years; 9 males) of British nationality and was compiled at the Derriford Hospital, Plymouth, Devon, UK. It was shown that no significant effect stemmed from either down sampling and discrepancy of age between the two groups (Henderson et al., 2006). The patient group underwent full neuroimaging and cognitive testing, but this test information was not always available due to disparities between districts. EEGs were recorded during 4 minutes at a sampling frequency of 256 Hz (later down sampled to 128Hz) using 19 electrodes disposed according to the Maudsley System, which is similar to the 10-20 international system.

2.1.2. Database B

This EEG database consisted of five age-matched healthy subjects (aged 76.6 ± 5.6 years; 3 males; MMSE 29.3 ± 0.7) and five AD patients (aged 78.8 ± 2.4 years; 2 males; MMSE 22.3 ± 3.1) of Italian nationality. The subjects were diagnosed with dementia of the Alzheimer's type (early-stage, mild form) according to the national institute of neurological and communicative disorders and stroke and the Alzheimer's disease and related disorders association (NINCDS-ADRDA) (McKhann, 1984) and diagnostic and statistical manual of mental disorder, 4th edition (DSM-IV) criteria. They underwent general medical, neuroimaging (magnetic resonance imaging (MRI)), and psychiatric assessments. Cognitive evaluations were conducted including the MMSE (Folstein et al., 1975), the clinical dementia rating scale (CDRS) (Hughes et al., 1982), and the geriatric depression scale (GDS) (Yesavage et al., 1983). EEG signals were recorded during 1 minute at a sampling frequency of 128 Hz using 21 electrodes disposed according to the 10-20 international system at the University of Malta, Msida MSD06, Malta.

2.1.3. Database C

This database was composed of three healthy subjects (aged 73.5 ± 2.2 years; 3 males; range available only, MMSE 28-30) and eight age-matched AD patients (aged 75.0 ± 3.4 years; 6 males; range available only, MMSE 20-25) of Romanian nationality and was obtained from the Ecological University of Bucharest, Bucharest, Romania. Patients were diagnosed with a mild form of AD using psychometric tests (MMSE, CDR), neuroimaging (computed tomography (CT)), and clinical examinations. EEG time series were recorded during 10 to 20 minutes at a sampling frequency of 512 Hz using 22 electrodes disposed according to the international federation of clinical neurophysiology (IFCN) standards for digital recording of clinical EEG (Nuwer et al., 1998).

2.2. EEG preprocessing

For each database, EEGs were band-pass filtered (digital second-order Butterworth filter) between 0.5 and 30 Hz. We also obtained a version of the filtered data that was preprocessed for artifact removal using ICA, a blind source separation (BSS) technique (Cichocki et al., 2007; Tichavsky et al., 2008) (here called 'the ICA-processed data'), as compared with the whole duration data that were only band-pass filtered (here called 'the filtered data'). The ICA-processed data would represent a data that is further free from artifact that consistently propagated spatially, and would distort the signal and contribute to a greater inter-subject variability. For the ICA-processed data, three technicians visually inspected the EEGs and chose the least corrupted continuous 20-second recordings for our analysis. Each trial was then decomposed using a recently developed version of the ICA algorithm, an improved weight-adjusted second-order blind identification (IWASOBI). We discarded the sources corresponding to artifacts such as electrooculograms, electromyograms, and electrocardiograms using three criteria: abnormal wave forms, isolated sources on the scalp, and sources of abnormally high amplitude. An additional concurrent comparison of each technician's inspection was performed. The detailed preprocessing of the EEGs was presented in a previous study (Vialatte et al., 2009). This preprocessing phase has for main objective to show how effective a technique can be under simple filtering and ICA-filtering conditions, also highlighting that the later condition might be essential to extract most general features out of a database.

2.3. Data organization

Since EEGs were recorded from the subjects in the resting condition, the focus of this study was oriented towards region-specific and spectral differences between AD patients and healthy subjects rather than the brain temporal (Moretti et al., 2004; Signorino et al., 1995). The dataset was organized into the form Subjects \times Frequency Band Power \times Region. The frequencies of EEGs were divided into the five bands δ (1-4 Hz), θ (4-8 Hz), α_1 (8-10 Hz), α_2 (10-12 Hz), and β (12-25 Hz), obtained from the fast Fourier transform (Welch method, Hanning window, 500-millisecond epochs overlapping by 50%). The regions represented clusters of specific electrodes (frontal, left temporal, central, right temporal, and posterior regions, supplementary Fig. 1). This representation reduced the number of features to extract and maintained consistency between subjects across databases (database A and B have 19 channels, while database C has 20 channels). A similar organization of tensors was previously proposed to extract common characteristics among subjects based on regional or disease interaction information (Cole and Ray, 1985).

2.4. MAD methods

The Multiway Array Decomposition is a method that is able to extract common interaction between dimensions (e.g. space, time, frequency, groups, etc) and summarized them within separate components. In our study, MAD is applied with the assumption that a limited set of interactions exists and are extractable; those interactions are also assumed to contain discriminative information for group classification. Note that feature extraction using MAD is unsupervised and constraints were applied to guarantee unique decompositions.

MAD was performed in this study according to the following steps. (1), an encoding variable matrix S_A was estimated from the largest EEG database (database A) after application of MAD (Fig. 1a). The encoding variable matrix contains the summary of all extracted common interactions between every dimension of interest, i.e. frequency and space in this study (see section 2.4.2 for the

extraction of encoding variable matrix), and if applied to a subject, could differentiate that subject based on estimated weights. (2), we extracted the weights or features of every subjects used for classification by applying the least-square (LS) projection between the pseudo-inverse of the encoding variable matrix, S_A^{-1} , and the matricized tensor towards the ‘subject’ mode, $[X_A]_{(subject)}$, $[X_B]_{(subject)}$, and $[X_C]_{(subject)}$ of the databases A, B, and C, respectively (Fig. 1b, c and d, left section). A tensor that is unfolded along a defined dimension to obtain a matrix is referred as matricized tensor (see Fig. 2c). For each subject, we obtained a number of features equal to the number of component extracted through MAD, typically lower than three. (3), we used the features extracted from database A to train and validate the MLP models and selected the best model, which exhibited the highest accuracy and lowest complexity (Fig. 1b). Using the least complex classifier was essential to guarantee the generalizability of the classification results; typically, linear classifiers were preferred. (4), we utilized the features extracted from database B to regularize the best MLP model (Fig. 1c). (5) Finally, we tested the regularized version of the model using the features extracted from database C (Fig. 1d).

A similar approach was used for the Spatial-Spectral Filters method (SSFs). In step (1), we estimated the SSFs as group average from database A. We then extracted features as in step 2 to 5 using a matrix distance measure on each database following the identical scheme used for MAD (see section 2.5 for details). For the blind source separation method AMUSE, we limited the study up to the step 3 (Fig. 1b), due to unsatisfactory classification results compared to the simplest method, SSFs.

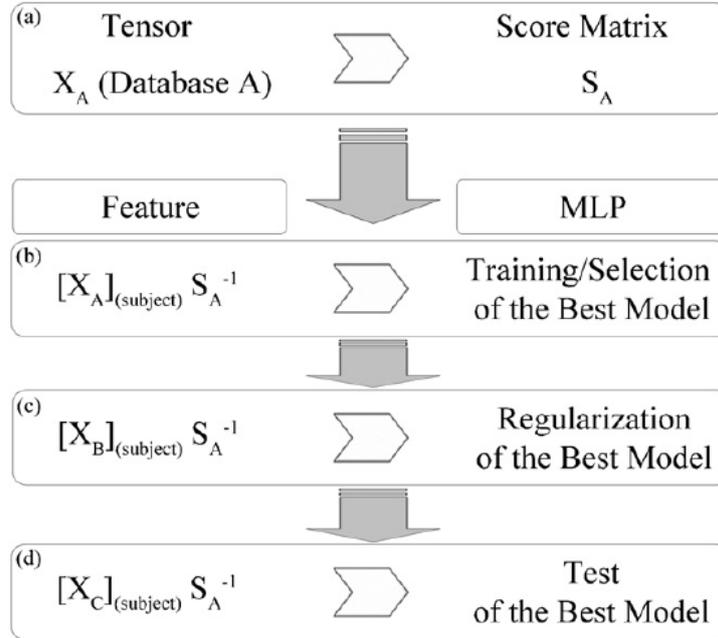


Fig. 1. Diagram for the classification of AD using multiway array decomposition and SSFs; (a) the encoding variable matrix S_A is originally estimated from database A using MAD. We extract features by applying the LS-projection between the inverse encoding variable matrix and the ‘subject’ mode-matricized tensor of a database; (b) using the features extracted from database A, we train MLPs and selected the best model; (c) using the features extracted from database B, we regularize the best MLP; and (d) the regularized MLP is finally test using the features extracted from the database C.

2.4.1. Comparison between MAD algorithms

Multiway-Array Decomposition (MAD) is a method that tends to separate a multi-dimensional data into either a sum of components of the same dimensionality, or a core tensor and matrices related to each dimension (Fig. 2a and 2b). MAD components can be understood as independent components that contains a representative relation between a set of dimensions; these components are considered as the basis to decompose the data into few templates of interactions just as independent component analysis (ICA) would decompose a two-way data template independent components. Two MAD methods were used in this study: PARAFAC (Harshman, 1970) and NTD, which was equivalent to Tucker3 decomposition (Tucker, 1966) with a non-negativity constraint (Kim et al., 2008).

The main idea underlying MAD is that the extraction of a multidimensional structure from EEG data could highlight important interactions that are otherwise invisible at lower dimensions. Among them, the PARAFAC model estimates all components simultaneously based on their multilinear interaction (i.e. trilinearity in this study), and results in an easily interpretable, unique fit

to the data due to the simplicity of the mathematical model (i.e. linear interaction between dimensions). This model is often easier to interpret than two-way analysis (Bro, 1997) or complex models, such as Tucker3 and NTD. The NTD model is more general and flexible than the PARAFAC model, allowing all components of each dimension (also called mode) to interact through the mean of a core tensor, whereas PARAFAC provides a component-wise tri-linear interaction. The NTD model is applicable to data with non-negative values and improves the convergence and robustness of the model. We tested the NTD model calculated from the algorithm based on alpha divergence, a generalized cost function for the optimization of the factorization problem and sparseness constraint on the core tensor to guarantee uniqueness (Cichocki et al., 2009a; Cichocki et al., 2009b; Mørup et al., 2008).

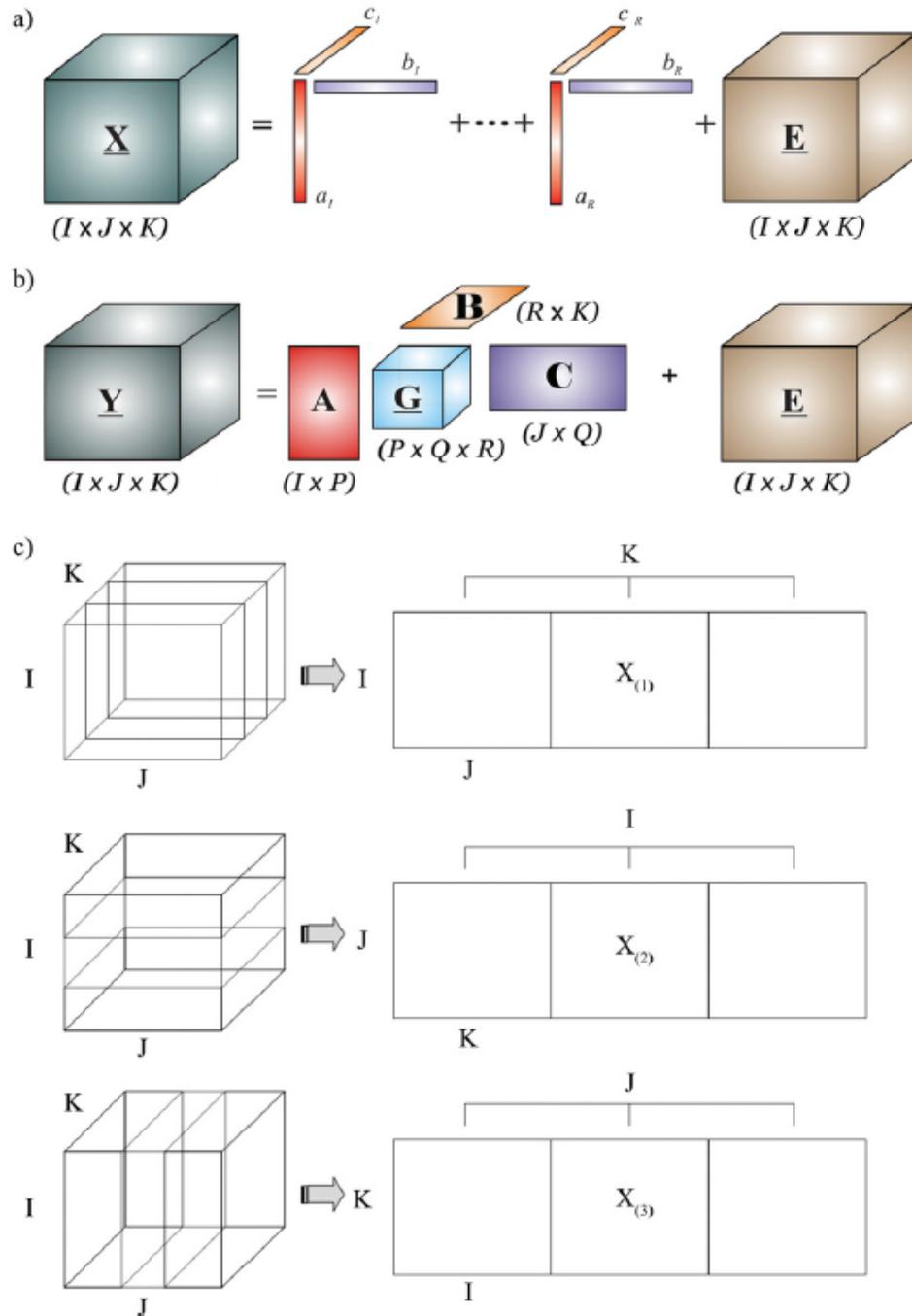


Fig. 2. (a) PARAFAC modeling of a 3-way tensor; there are R components defined as the outer product of a rank-1 tensor a , b , and c , and E is a residual tensor; (b) the (P, Q, R) -component NTD decomposition of a three-way tensor. The NTD decomposition results in one matrix per mode (here three modes) of size $I \times P$ (i.e. A), $K \times R$ (i.e. B) and $J \times Q$ (i.e. C), respectively, and a core matrix of size $P \times Q \times R$ (i.e. G). Here, R is the number of components of the third mode, which is distinct from the R for the PARAFAC; (c) the matricization of a 3-way tensor (flattening of tensor). The matricization along its n th mode correspond to the unfolding of the other modes along this mode; the resulting matrix has the dimension $I \times JK$, $J \times IK$ and $K \times IJ$ for the first, second and third mode matricization, respectively.

2.4.2. Extraction of encoding variable matrix using MAD

PARAFAC is often referred to as a multilinear version of bi-linear factor models. From a given three-way tensor, $\underline{X} \in \mathfrak{R}^{I \times J \times K}$, this model is capable of extracting the R-decompositions of rank-1 tensors, where R designates the number of components, which is identical for each dimension referred as mode in for the remainder of this paper (Fig. 2a):

$$x_{ijk} = \sum_{r=1}^R (a_r \circ b_r \circ c_r)_{ijk} + e_{ijk}, \quad (1)$$

where \mathbf{a}_r is the r^{th} column of $A \in \mathfrak{R}^{I \times R}$, \mathbf{b}_r is the r^{th} column vector of $B \in \mathfrak{R}^{J \times R}$, and \mathbf{c}_r is the r^{th} column vector of $C \in \mathfrak{R}^{K \times R}$, which are the component matrices (also referred to as loadings or factors) of the first, second, and third modes, respectively. $\underline{E} \in \mathfrak{R}^{I \times J \times K}$ is the residual error tensor. The operator \circ designates the vector outer product, defined as follows: given three column vectors \mathbf{x} , \mathbf{y} , and \mathbf{z} of respective sizes I, J, and K, $\underline{W} = \mathbf{x} \circ \mathbf{y} \circ \mathbf{z}$ is a tensor of size I \times J \times K if and only if $w_{ijk} = x_i y_j z_k$. In this analysis, the component matrices A, B, and C refer to the mode ‘subject,’ ‘frequency band,’ and ‘region,’ respectively.

Given the three-way tensor $\underline{X} \in \mathfrak{R}^{I \times J \times K}$, its (P, Q, R)-component NTD model, where P, Q, and R designate the number of components for each of the three modes, is given by:

$$x_{ijk} = \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_{ip} b_{jq} c_{kr} + e_{ijk}, \quad (2)$$

where $A \in \mathfrak{R}^{I \times P}$, $B \in \mathfrak{R}^{J \times Q}$, and $C \in \mathfrak{R}^{K \times R}$ are the component matrices of the first, second, and third modes, respectively. $\underline{G} \in \mathfrak{R}^{P \times Q \times R}$ is the core tensor of the model and $\underline{E} \in \mathfrak{R}^{I \times J \times K}$ is the residual error tensor. An illustration of the NTD model is shown in Fig. 2b.

The mode-n matricization of tensor X (i.e., tensor unfolding in the direction of the n^{th} dimension) (Fig. 2c) can be formulated after PARAFAC decomposition, which was expressed as the column-wise Kronecker product (Khatri and Rao, 1968) of PARAFAC modes, as shown in Eq. 3:

$$X_{(n)} \approx M_{(n)} \left[M_{(n-1)} \otimes \dots \otimes M_{(1)} \otimes M_{(N)} \otimes \dots \otimes M_{(n+1)} \right]^T$$

$$X_{(n)} \approx M_{(n)} S_M^{(n)}, \quad (3)$$

where $M(n)$ is the component matrix of mode n and $S_M^{(n)}$ is the corresponding encoding variable matrix. In our analysis, $M_{(1)} = A$ and $S_A \in \mathfrak{R}^{R \times JK}$, $M_{(2)} = B$ and $S_B \in \mathfrak{R}^{R \times IK}$, or $M_{(3)} = C$ and $S_C \in \mathfrak{R}^{R \times IJ}$ for the modes $n = 1$ (subject), 2 (frequency), and 3 (region), respectively. The operator \otimes denotes the column-wise Kronecker product of two matrices. It is particularly interesting to examine Eq. 3 for $n = 1$ (subject mode) (Fig. 1a), where S_A is the encoding variable matrix whose folded version contains matrices of Frequency Bands Power \times Region (the outer product of ‘mode frequency’ and ‘mode region’) for each component and A (‘mode subject’) is the weighted contribution of each subject to each component’s matrix (Frequency Bands Power \times Region).

From Eq. 2, it is possible to derive the equivalent of Eq. 3 for the NTD models as follows:

$$x_{ijk} \approx \sum_{p=1}^P \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} a_{ip} b_{jq} c_{kr}$$

$$\approx \sum_{p=1}^P a_{ip} \sum_{q=1}^Q \sum_{r=1}^R g_{pqr} b_{jq} c_{kr}$$

$$\equiv \sum_{p=1}^P a_{ip} S_{Pjk}^A$$

$$, \quad (4)$$

where $A \in \mathfrak{R}^{I \times P}$, $S_A \in \mathfrak{R}^{R \times JK}$, and P is the number of components in the first mode (the equivalent of R for the PARAFAC model). The matricization for the case of $n = 1$ (‘mode subject’ on the left side and ‘mode component’ on the right side) of Eq. 4 leads to Eq. 3. For the NTD model, we utilized the percentage of explained variance to estimate the number of components in database A.

In any MAD model, it is possible to extract the weighted contribution of the i^{th} subject (that is, the i^{th} row of the component

matrix of the ‘mode subject’) relating to the mode of frequency and region (Eq. 3 and 4). We used database A to obtain the encoding variable matrix S_A (Eq. 3 for PARAFAC, Eq. 4 for NTD) (Fig. 3a). It was then possible to calculate the approximate values of the weighted contributions of subjects of any database using the LS projection $[X_{test}]_{(subject)} S_A^{-1} \approx \tilde{A}$ (Lee et al., 2007). Here, S_A^{-1} denotes the pseudo-inverse of S_A and $[X_{test}]_{(subject)}$ is the matricization of the tensor X_{test} along the ‘mode subject’ (Fig. 3b).

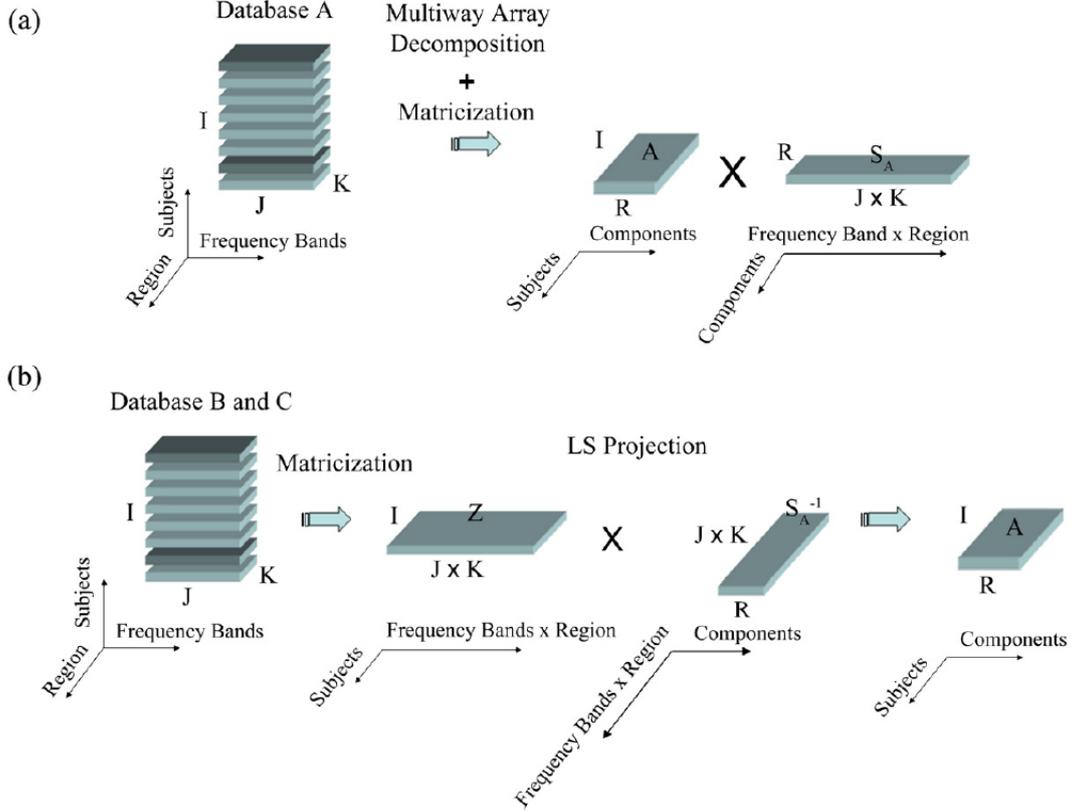


Fig. 3. Feature extraction using multiway array decomposition: (a) from an original database (database A), the MAD allows one to extract the encoding variable matrix S_A ; (b) the LS projection of the encoding variable matrix S_A on the unfolded (i.e. first mode matricization) version of any database (three-way tensor Subjects \times Frequency Band Power \times Region) returns the feature matrix A that contains R (PARAFAC) features.

2.4.3. Estimation of number of components

The optimal fitting of the model depends on the number of components R (PARAFAC) or (P, Q, R) (NTD). In this study, we opted for validation of the model (i.e. choice of number of components) based on the core consistency (Bro and Kiers, 2003). The core consistency diagnosis roughly tests a MAD complexity by assessing how much an additional component would be affected by random variation. For the PARAFAC model estimation using database A, we found that $R = 2$ was a reasonable choice for the number of components. For the NTD models, the values (P, Q, R) = (2, 1, 2) (the number of components for the ‘mode subject,’ ‘frequency,’ and ‘region’) produced the maximum of the explained variances of the model’s core (Pravdova et al., 2002) as well as the maximal percentage of fit.

2.5. Spectral-spatial filters

Spectral-spatial filters (SSFs) are matrices that are mean representations of each group from their spectral and regional repartition of powers. We obtained two SSFs by averaging the values of the matrices (Frequency Band Power \times Region) over the subjects in the control group (F_{CTR}) and the AD group (F_{AD}). The MAD method fits a model and decomposes it into two components with relation to (Frequency Band Power \times Region), F_1 and F_2 whereas SSFs are mere averaging of the original data knowing the group compositions. By analogy, this step corresponds to step (1) of the MAD method described in Fig. 1a. The approach preserves the regional repartition of power for each frequency band. Similar use of spectral filters has been applied to

classification, yet this study did not consider the regional grouping of electrodes (Wu et al., 2008). It is important to note that because each SSF contains information from a single group, the classification associated with SSFs is supervised, as compared with MAD, where the features are extracted in an unsupervised fashion.

To extract the features associated with the SSFs, the matrix scalar product, also known as the normalized Frobenius inner product, was applied as a distance metric D between subject's (Frequency Band Power \times Region) matrices, F_{subjects} , and the SSFs, as given in Eq. 5:

$$D(A, B) = \frac{A : B}{\|A\| \|B\|}, \quad (5)$$

where A and B are Frequency Band Power \times Region matrices; $A : B$ is the Frobenius inner product of the matrices A and B , defined as $A : B = \text{trace}(A^T B)$; and $\|\cdot\|$ is the Frobenius norm, defined as $\|A\| = \sqrt{\text{trace}(A^T A)}$, where T denotes the conjugate transpose of a matrix. The distance D is used to compare the SSFs F_{CTR} and F_{AD} with each subject's matrix (Frequency Band Power \times Region), F_{subjects} , thus yielding two features per subject. This step is analogous to the second step of the MAD methods, in which we extracted features from the database tensors by applying the normalized Frobenius inner product subject-wise to the Frequency Band Power \times Region matrix, instead of the LS projection.

As for the MAD methods, after the extraction of features, MLPs were trained using features from database A and the model with the highest accuracy and lowest complexity was selected. This model was regularized using the features from database B and then tested using the features from database C.

2.6. Classification: training, validation, regularization, and performance testing

Classification procedures include the estimation of a classifier's quality based on its probability of correctly identifying unknown data. According to Cover's theorem, the more complex a model, the higher the risk of over fitting, which biases a model toward its training set (Dreyfus, 2005; Myung, 2000). The estimate of true error rates loses its reliability in the case of small data sets and/or a large number of parameters (Buscema et al., 2007; Rossini et al., 2008).

In general, model comparison using classification results is irrelevant, since classifiers are often dependent on thresholds. The gold standard was to implement the validation set only for model comparisons using area under the curve (AUC) statistics or root mean squared errors (RMSEs) which measures offers a global perspective on the classification performance, independently of a threshold, and thus a more suitable to compare models' generalizability. We implemented a feed-forward multilayer perceptron (MLP) as a classifier using the Matlab® Neural Network toolbox™ 4.2 (Demuth and Beale, 1994). Since the number of features for each subject was equal to two for the MAD-based classification ($R = 2$ for the PARAFAC, $P = 2$ for the NTD) and for the SSF-based classification (the comparison of two SSFs for each subject), the number of neurons in the input layer was equal to two. The MLP was trained using a Levenberg-Marquardt backpropagation with sigmoid transfer function over 100 iterations. The linear neural networks (i.e., no hidden neurons) returned a post-training unique solution, while solutions from the nonlinear versions (i.e., up to three neurons in one hidden layer) were non-unique. A sigmoid transfer function was used in the MLP.

We modulated the classifier's complexity using the number of hidden units, and we applied a leave-one-out cross-validation RMSE to the validation set (database A). Over-fitting was prevented using an early stopping criterion on the validation set (database A) and the regulation set (database B). Once the best model was trained, performance estimation was conducted using an independent test set (database C).

3. RESULTS

Using the normalized Frobenius inner product (i.e., matrix scalar product; see Eq. 5), we computed the distance D as similarity between each component of the folded version of the encoding variable matrix S_A of the PARAFAC and NTD models (the equivalent of the SSF obtained from MAD; see Eq. 3 and 4) and the SSFs (F_{CTR} and F_{AD}). We found high D values between the F_{AD} and the first component of the PARAFAC ($D = 0.9346$) and the NTD ($D = 0.9867$) models, as summarized in Table 1. This result indicates that MAD characterized common interactions in AD group closely to the average spectral-spatial characteristics that group in an unsupervised manner. Meanwhile, the spectral-spatial characteristics of the healthy subjects closely matched the second component of the NTD ($D = 0.9456$), while the second component of the PARAFAC visually exhibited a strong non-region-specific influence from high frequencies (α_1 , α_2 , and β bands) yet did not exhibit the characteristic peak in α_1 -band power (that is, characteristic of a healthy subject in the resting state condition). This result implies that MAD could differentiate spectral-spatial characteristics in the data that were common to AD, indicating a slowing of frequencies. The NTD was able to extract additional information in its second component that was closely related to the SSF of the healthy subjects.

Table 1. Distance D between the SSF (average filters; supervised) of ADs (F_{AD}) and controls (F_{CTR}) and each spectral-spatial filters extracted from the multiway decomposition methods (unsupervised); The number of component for the PARAFAC and NTD was equal to two, hence two filters (#1 and #2) were obtained for each model. The metric used is the normalized Frobenius inner product; a distance ≥ 0.90 and ≥ 0.95 indicate similar (*) and very similar (**) matrices, respectively.

	F_{AD}	PARAFAC		NTD	
SSF	-	#1	#2	#1	#2
F_{CTR}	0.785	0.525	0.349	0.864	0.946*
F_{AD}	1.000	0.935*	-0.295	0.987**	0.576

We then estimated the classification performance of the ANN model with zero (linear) to three neurons (one hidden layer) using features extracted from the filtered version of database A. We compared the five features of frequency band powers (mean of each frequency band power over all electrodes, with no regional distinction) extracted using both a BSS method (i.e., algorithm for multiple unknown signals extraction, or AMUSE (Cichocki et al., 2007); see also Supplementary Information 1) and the supervised method based on the SSFs proposed in this paper. Since the quality of extracted features is of interest, we decided to focus on the results describing the use of linear classification (MLP with no hidden layer). We found that AMUSE (70.7%, RMSE = 0.51) significantly increased the classification performance of the average power on filtered data taken alone (61%, RMSE = 0.52) (mean frequency band power over all electrodes) (Table 2). Organization into five frequency bands and five regions of filtered data (average *Frequency Band Power* \times *Region filters*, SSF, over the control and AD groups) further improved classification accuracy (85.4%, RMSE = 0.35, linear classifier). This result indicates that the use of additional dimension (the ‘mode region’) improves the classification performance. We observed that AMUSE did not show satisfactory classification results either using filtered or ICA-processed data (data not shown). Following, we only included SSFs and MAD method for the classification on ICA-processed data.

Table 2. Comparison of the classification accuracies after LOO cross-validation for the filtered data (frequency band only, no regional grouping), AMUSE method and SSF on filtered data; for nonlinear neural network model (i.e. # hidden neuron > 0), it provides the mean and standard deviation of accuracy, and the maximum accuracy is shown in bracket. Unique solutions are found for linear model of neural network (i.e. # hidden neuron = 0), hence no mean accuracy is provided. The classification accuracies obtained by linear classification is marked in bold. The corresponding RMSE can be found in the supplementary table 1.

	# of Hidden Neurons	0	1	2	3
Filtered Data	Training	71.8%	91.9 \pm 0.5 (92.8%)	89.7 \pm 1.5 (93.0%)	97.3 \pm 1.2 (100)
	Validation	61.0%	61 \pm 3.9 (68.3%)	63.4 \pm 7.2 (80.5%)	70.7 \pm 5.4 (80.5%)
AMUSE on Filtered Data	Training	80.5%	85.5 \pm 0.1 (85.6%)	88.9 \pm 1.7 (92.0%)	96.3 \pm 1.0 (97.9%)
	Validation	70.7%	75.6 \pm 1.9 (78.0%)	61.0 \pm 5.0 (70.7%)	58.5 \pm 5.4 (68.3%)
SSF on Filtered Data	Training	90.3%	89.3 \pm 2.2 (93.7%)	92.2 \pm 2.2 (96.5%)	93.9 \pm 1.8 (97.4%)
	Validation	85.4%	71.9 \pm 6.0 (87.8%)	71.4 \pm 6.1 (85.4%)	71.2 \pm 5.7 (85.4%)

We trained and validated an MLP using the features based on the ICA-processed database A using SSFs, PARAFAC and NTD. We trained and validated an MLP using the features based on the ICA-processed database A using SSFs, PARAFAC and NTD. In addition, we tested the classification accuracy of conventional a two-dimensional decomposition method, singular value decomposition (SVD), on the subject-mode matricized tensor of database A (supplementary information 1): the tensor Subjects \times Frequency Band Power \times Region becomes a matrix of size Subjects \times (Frequency Band Power \times Region). This decomposition method represents a simpler alternative to the extraction of unique features for classification from tensors database which could or not take advantage of additional mode (e.g. region) information. The classification based on SSFs and ICA-processed data yielded a classification accuracy of 87.8% (RMSE = 0.30) (Table 3). This indicates that ICA preprocessing improves the classification performance, as compared with that of filtered data (85.4%, RMSE = 0.35 (Table 3)). The classification based on unfolding and SVD showed accuracy of 87.1% (RMSE = 0.38). This indicates that the conventional two-dimensional decomposition combined to matricization seems to not take advantage of the additional modes of the tensor data (i.e. regional grouping), also showing the superiority of multiway structure for feature extraction. Among the MAD models, NTD exhibited the highest classification accuracy (97.6%, RMSE = 0.20, linear classifier), while PARAFAC had lower classification performance (92.7%, RMSE = 0.30,

linear classifier). This result indicates that MAD methods are able to generalize better than SSFs.

Table 3. Comparison of the classification accuracies after LOO cross-validation for SSF on ICA processed data, SVD on ICA processed and matricized data, and PARAFAC ($R = 2$) and NTD ($P, Q, R = (2, 1, 2)$) on ICA processed data; for the nonlinear neural network model (i.e. # hidden neuron > 0), it is provided the mean and standard deviation of accuracy, and the maximum accuracy is shown in bracket. Unique solutions are found for linear model of neural network (i.e. # hidden neuron = 0), hence no mean accuracy is provided. The classification accuracies obtained by linear classification is marked in bold. The corresponding RMSE can be found in the supplementary table 2.

		# of Hidden Neurons	0	1	2	3
SSF on ICA processed Data	Training		97.4%	90.3±2.6 (95.7%)	92.1±2.2 (97.4%)	94.4±1.8 (98.2%)
	Validation		87.8%	83.9±3.7 (92.7%)	83.4±3.7 (92.7%)	84.0±3.8 (92.7%)
SVD on ICA processed matricized Data	Training		97.6%	96.7±0.8 (97.5%)	94.6±1.0 (93.7%)	97.8±1.0 (97.6%)
	Validation		87.1%	89.5±3.6 (95.1%)	87.8±3.4 (92.7%)	88.3±3.0 (92.7%)
PARAFAC	Training		97.6%	86.2±2.6 (91.8%)	93.0±2.1 (93.7%)	95.8±1.6 (92.7%)
	Validation		92.7%	82.6±5.1 (92.3%)	87.3±3.5 (92.7%)	89.4±2.7 (95.1%)
NTD	Training		97.6%	93.0±2.0 (96.6%)	94.4±1.8 (97.6%)	95.0±1.3 (97.6%)
	Validation		97.6%	90.9±3.0 (95.1%)	92.4±2.8 (97.6%)	93.9±2.6 (97.6%)

The regularization dataset, database B, was used to tune the classifiers using the early stopping method. The regularization based on an independent database ensured reduction of over fitting and thus helped to obtain a generalized, better classifier. The results for the linear model are summarized in Table 4. We found that SSF (RMSE = 0.31), PARAFAC (RMSE = 0.30) and NTD (RMSE = 0.31) had a classification accuracy of 90.0%, indicating only one misclassified subject (Supplementary Information 5).

We verified the trained and tuned models using database C as the test set. The models were left unmodified at this point. All models exhibited classification accuracies of 100% (Table 4). Notably, PARAFAC showed the best generalization properties among the three MAD models, with an RMSE of less than 0.11, a sensitivity of 94.1%, and a specificity of 91.7% for database A; a sensitivity of 80.0% and a specificity of 100% for database B; and a sensitivity and specificity of 100% for database C. From a classification perspective, the MAD- and SSF-based methods performed similarly. However, the RMSE result indicated that unsupervised MAD-based methods are more reliable than the supervised SSF-based method in terms of generalizability (i.e., lower RMSE values). We also verified that PARAFAC was able to perform better than SSFs when applied to the filtered data using the model trained on ICA-processed data, indicating that features were less influenced by noise after data processing and could recover discriminative information even in the presence of noise.

Table 4. Recapitulative classification using linear classifiers (MLP with no hidden neurons) results in training, validation, regulation/optimization and test conditions using SSF, PARAFAC and NTD. The two best models are marked in bold for each of the four steps of classification.

		Database A		Database B	Database C
		Training	Validation	Regulation	Test
SSF	- Accuracy (%)	97.4%	87.8%	90.0%	100%
	- RMSE	0.16	0.30	0.31	0.19
PARAFAC	- Accuracy (%)	97.6%	92.7%	90.0%	100%
	- RMSE	0.15	0.30	0.30	0.11
NTD	- Accuracy (%)	97.6%	97.6%	90.0%	100%
	- RMSE	0.16	0.20	0.31	0.13

4. DISCUSSION

In this study, we used unsupervised MAD models to extract features for classification of mild AD and have demonstrated that the methods are generalizable using three multi-centric, independent databases. The MAD methods using five frequency bands and five region groupings have performed better on average than the supervised SSF-based method: classification accuracy over 97.4% (RMSE < 0.16) for training, 92.7% (RMSE < 0.30) for validation, 90.0% (RMSE < 0.31) for regulation, and 100% (RMSE < 0.20) for testing, respectively. This approach has outperformed previous methods in AD classification and diagnosis (Buscema et al., 2007; Lehmann et al., 2007; Rossini et al., 2008) while reducing 25 features (5 frequency bands \times 5 regions) to two linearly

separable features ($R = 2$ for PARAFAC, $P = 2$ for NTD). The extracted feature also showed a meaningful association with the average spectral-spatial characteristics of healthy and AD subjects (Table 1).

Using the NTD method, we obtained the highest classification accuracy, 97.6%, under training and validation conditions, as compared with PARAFAC (training, 92.7%; validation, 92.7%). In contrast, the supervised approach using SSF obtained a classification performance of 97.4% in the training condition and 87.8% in the validation condition. Both the supervised (SSF) and unsupervised (MAD) methods used in this study performed similarly, as compared with previous studies of resting state EEGs in mild AD and control patients under training and validation conditions (Buscema et al., 2007; Lehmann et al., 2007; Rossini et al., 2008). As far as regulation conditions are concerned, all methods exhibited 90.0% classification accuracy, where a single AD outlier was consistently misclassified (Supplementary Information 5). Using a small number of simple features, the MAD methods extracted characteristics that linearly differentiate AD patients from controls with accuracies around the current limit of the field (97.7%) in an unsupervised manner.

More significantly, application of this method without any further modification to the third database resulted in 100% classification accuracy with an RMSE as low as 0.11 for PARAFAC and 0.13 for NTD. This result is of clinical importance since the data were obtained from different hospitals and since different EEG recording facilities had varying sampling rates, channel positioning, and possibly data quality. The method presented here demonstrates strength in identifying meaningful features as well as in its general application to physiological datasets including different parameters. To our best of knowledge, none of the previous reports (Buscema et al., 2007; Lehmann et al., 2007; Rossini et al., 2008) used a test database that was completely independent from the training and validation datasets used to construct the classifier for mild AD. This method of diagnosing AD based on EEGs recorded at different hospitals may thus be clinically applicable for mass screening.

We assessed the similarity between unsupervised SSFs obtained using MAD and supervised SSFs calculated from each clinical group (Table 1). The MAD methods decomposed database A into two components for the ‘subject mode’ ($R = 2$, PARAFAC; $P = 2$, NTD), which was highly similar to the SSFs (Table 1). Therefore, it may be possible to extract the common characteristics of each group from their spectral-spatial profile. Additionally, MAD may be capable of efficiently differentiating dominant structures in the database, such as the spectral-spatial characteristics of healthy and AD subjects. MAD could then distinguish subjects based on modal interactions in an unsupervised manner, based on the assumption that the features in the modes (Fourier Power over different regions and bands in this study) contain the information of discriminative power (Cole and Ray, 1985).

We observed that, for healthy controls, the repartition of power was equivalent between all regions, which showed a characteristic peak in the α_1 band of the EEG. The AD patients exhibited the typical ‘slowing of frequencies,’ with globally decreased powers in the α_1 band and high powers in the θ band, as compared with control subjects. We suggest that by using novel types of features, including the interaction of several modes (e.g., frequency, region, and time), MAD could help to characterize and differentiate groups in a mode-wise fashion, thus allowing new hypothesis formulation on group profiles.

Comparing the classification performance of the SSF and MAD methods between filtered-only and ICA-processed data (Tables 2 and 3), ICA-processed data performed better for both the SSF and MAD methods. Thus, we suggest that the presence of noise may reduce classification performance and data preprocessing, and that noise reduction using ICA or other techniques should be included as a crucial step for mass screening methods. We additionally showed using SVD on subject-mode matricized data (Table 3), that MAD are able to take advantage of the multiway structure to improve feature extraction, which was not an effect due to filtering or addition of mode only. With the increasing number of mode, the matricized tensor of data might not reveal a complete interaction between every variable presented, resulting in less meaningful decomposition (Cichocki, 2011). Two-dimensional decompositions coupled with matricization are, however, a powerful and simple approach to unique decomposition of tensor. The choice of constraints (e.g. statistical independence, orthogonality, sparseness, non negativity and so on) shall play a critical role in the meaningfulness and efficiency of this decomposition approach, and such demonstration shall be reserved for future studies.

There are several limitations to the current study, including the size of the databases used for regularization and testing. To produce a better estimate of the true error rates in the testing condition, more than several tens of test subjects in each clinical group are required. To compensate for such limitations, importance should be placed on extensive outlier analyses and anti-over fitting techniques that can be applied before the actual training and regularization of a classifier. We are investigating the classification performance of this method using large populations of AD and healthy subjects from different hospitals. In addition, although MAD methods are unsupervised, the actual organization of data (i.e., no random shuffle between subjects of each group) possibly helps to extract the correct spectral-spatial characteristics for classification, indicating a possible influence of the EEG data structure for the extraction of the original features. This has not been tested yet but will be examined in future investigations.

ACKNOWLEDGMENT

This work was supported by a Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MOST) (No. R01-2007-000-21094-0 and No. M10644000013 -06N4400-01310). The first author would like to thank the Minister of Information and Technology of South Korea, Institute for Information and Technology Advancement (IITA) for his

financial support. Jordi Solé-Casals acknowledges support from the Fundació Alicia Koplowitz and from the University of Vic (grant R0904). The authors would like to thank Drs. Cindy Goh and Emmanuel Ifeachor (University of Plymouth, UK), Drs. Cristin Bigan and Mircea Besleaga (Ecological University of Bucharest, Romania) and Dr. Paolo Capotosto (Istituto di Tecnologie Avanzate Biomediche Università "G. D'Annunzio", Chieti, Italy) for their precious contribution in the subjects management, data recording and data management.

REFERENCES

- Acar E, Aykut-Bingol C, Bingol H, Bro R, Yener B. Multiway analysis of epilepsy tensors. *Bioinformatics*, 2007; 23: i10-8.
- Acar E, Bing CA, Bing H, Yener B. Computational analysis of epileptic focus localization. *Proceedings of the 24th IASTED international conference on Biomedical engineering: Innsbruck, Austria, 2006: 317-22.*
- Adler G, Brassen S, Jajcevic A. EEG coherence in Alzheimer's dementia. *J. Neural Transm.*, 2003; 110: 1051-8.
- Alexander GE. Longitudinal PET Evaluation of Cerebral Metabolic Decline in Dementia: A Potential Outcome Measure in Alzheimer's Disease Treatment Studies. *A- J. Psychiatry*, 2002; 159: 738-45.
- Andreasen N, Minthon L, Davidsson P, Vanmechelen E, Vanderstichele H, Winblad B, Blennow K. Evaluation of CSF-tau and CSF-Aβ42 as Diagnostic Markers for Alzheimer Disease in Clinical Practice. *Arch. Neurol.*, 2001; 58: 373-9.
- Besthorn C, Zerfass R, Geiger-Kabisch C, Sattel H, Daniel S, Schreiter-Gasser U, Forstl H. Discrimination of Alzheimer's disease and normal aging by EEG data. *Electroencephalogr. Clin. Neurophysiol.*, 1997; 103: 241-8.
- Braak H, Braak E. Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol. (Berl)*. 1991; 82: 239-59.
- Bro R. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 1997; 38: 149-71.
- Bro R, Kiers HAL. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics*, 2003; 17: 274-86.
- Buscema M, Rossini P, Babiloni C, Grossi E. The IFAST model, a novel parallel nonlinear EEG analysis technique, distinguishes mild cognitive impairment and Alzheimer's disease patients with high degree of accuracy. *Artif. Intell. Med.*, 2007; 40: 127-41.
- Cichocki A. Tensor Decompositions: new concepts in brain data analysis? *Journal of Cont. Meas. and Sys. Int. (SICE)*, 2011; 7: 1-10.
- Cichocki A, Amari S, Siwek K, Tanaka T. ICALAB Toolboxes. URL: <http://www.bsp.brain.riken.jp/ICALAB>, 2007.
- Cichocki A, Shishkin SL, Musha T, Leonowicz Z, Asada T, Kurachi T. EEG filtering based on blind source separation (BSS) for early detection of Alzheimer's disease. *Clin. Neurophysiol.*, 2005; 116: 729-37.
- Cichocki A, Zdunek R, Phan A-H, Amari S. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis*. John Wiley, 2009a: 552.
- Cichocki A, Zdunek R, Phan AH, Amari S. *Nonnegative Matrix and Tensor Factorizations: Applications to Exploratory Multi-way Data Analysis and Blind Source Separation*. Wiley, 2009b: 120-9.
- Coben LA, Danziger W, Storandt M. A longitudinal EEG study of mild senile dementia of Alzheimer type: changes at 1 year and at 2.5 years. *Electroencephalogr. Clin. Neurophysiol.*, 1985; 61: 101-12.
- Coben LA, Danziger WL, Berg L. Frequency analysis of the resting awake EEG in mild senile dementia of Alzheimer type. *Electroencephalography and Clinical Neurophysiology*, 1983; 55: 372-80.
- Cole HW, Ray WJ. EEG correlates of emotional tasks related to attentional demands. *Int. J. Psychophysiol.*, 1985; 3: 33-41.
- Dauwels J, Vialatte F, Cichocki A. Diagnosis of Alzheimers Disease from EEG Signals: Where Are We Standing? *Current Alzheimer Research*, 2010; 7: 487-505.
- Demuth HB, Beale MH. *MATLAB: Neural Network Toolbox*. MathWorks, Inc., 1994.
- Deweere B, Lehericy S, Pillon B, Baulac M, Chiras J, Marsault C, Agid Y, Dubois B. Memory disorders in probable Alzheimer's disease: the role of hippocampal atrophy as shown with MRI. *Br. Med. J.*, 1995; 58: 590.
- Dreyfus G. *Neural networks: methodology and applications*. Springer, 2005.
- Dunkin JJ, Leuchter AF, Newton TF, Cook IA. Reduced EEG coherence in dementia: State or trait marker? *1. Biol. Psychiatry*, 1994; 35: 870-9.
- Estienne F, Matthijs N, Massart DL, Ricoux P, Leibovici D. Multi-way modelling of high-dimensionality electroencephalographic data. *Chemometrics and Intelligent Laboratory Systems*, 2001; 58: 59-72.
- Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Hall K, Hasegawa K, Hendrie H, Huang Y. Global prevalence of dementia: a Delphi consensus study. *The Lancet*, 2006; 366: 2112-7.
- Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Psychiatric Research*, 1975; 12: 189-98.
- Giaquinto S, Nolfé G. The EEG in the normal elderly: A contribution to the interpretation of aging and dementia. *Electroencephalogr. Clin. Neurophysiol.*, 1986; 63: 540-6.
- Goh C, Ifeachor E, Henderson G, Latchoumane C, Jeong J, Bigan C, Besleaga M, Hudson N, Capotosto P, Wimalaratna S. P04. 3 Characterisation of EEG at different stages of Alzheimer's disease (AD). *Clin. Neurophysiol.*, 2006; 117: 138-9.
- Goh C, Ifeachor EC, Cassar T, Latchoumane CF, Bigan C, Henderson G, Camilleri K, Fabri S, Jeong J, Hudson N. Comparison of Methods for Early detection of Alzheimer's disease. 3rd International Conference on Computational Intelligence in Medicine and Healthcare (CIMED'07): Plymouth, UK, 2007.
- Hamadicharef B, Guan C, Ifeachor EC, Hudson N, Wimalaratna S. Performance Evaluation and Fusion of Methods for Early Detection of Alzheimer Disease. 2008.
- Harshman RA. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 1970; 16: 84.
- Hashimoto M, Kazui H, Matsumoto K, Nakano Y, Yasuda M, Mori E. Does donepezil treatment slow the progression of hippocampal atrophy in patients with Alzheimer's disease? *A- J. Psychiatry*, 2005; 162: 676.
- Haykin S. *Neural Network*. Macmillian Publishing: Upper Saddle River, New Jersey, 1994.
- Henderson G, Ifeachor E, Hudson N, Goh C, Outram N, Wimalaratna S, Del Percio C, Vecchio F. Development and assessment of methods for detecting dementia using the human electroencephalogram. *IEEE Transaction on Biomedical Engineering* 2006; 53: 1557-68.
- Huang C, Wahlund LO, Dierks T, Julin P, Winblad B, Jelic V. Discrimination of Alzheimer's disease and mild cognitive impairment by equivalent EEG sources: a cross-sectional and longitudinal study. *Clin. Neurophysiol.*, 2000; 111: 1961-7.
- Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. *The British Journal of Psychiatry*, 1982; 140: 566.
- Jeong J. EEG dynamics in patients with Alzheimer's disease. *Clin. Neurophysiol.*, 2004; 115: 1490-505.
- Khatri CG, Rao CR. Solutions to some functional equations and their applications to characterization of probability distributions. *Sankhy : The Indian Journal of Statistics, Series A*, 1968: 167-80.
- Kim YD, Cichocki A, Choi S. Nonnegative Tucker decomposition with alpha-divergence. *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP-2008)*. IEEE: Las Vegas, Nevada, 2008.

- Kowalski JW, Gawel M, Pfeffer A, Barcikowska M. The diagnostic value of EEG in Alzheimer disease: correlation with the severity of mental impairment. *J. Clin. Neurophysiol.*, 2001; 18: 570.
- Lee H, Kim YD, Cichocki A, Choi S. Nonnegative tensor factorization for continuous EEG classification. *Int. J. Neural Syst.*, 2007; 17: 305.
- Lehmann C, Koenig T, Jelic V, Prichep L, John RE, Wahlund LO, Dodge Y, Dierks T. Application and comparison of classification algorithms for recognition of Alzheimer's disease in electrical brain activity (EEG). *J. Neurosci. Methods*, 2007; 161: 342-50.
- Locatelli T, Cursi M, Liberati D, Franceschi M, Comi G. EEG coherence in Alzheimer's disease. *Electroencephalogr. Clin. Neurophysiol.*, 1998; 106: 229-37.
- McKham G. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *AAN Enterprises*, 1984: 939-44.
- Melissant C, Ypma A, Frietman EEE, Stam CJ. A method for detection of Alzheimer's disease using ICA-enhanced EEG measurements. *Artif. Intell. Med.*, 2005; 33: 209-22.
- Miwakeichi F, Mart nez-Montes E, Valde-Sosa PA, Nishiyama N, Mizuhara H, Yamaguchi Y. Decomposing EEG data into space-time-frequency components using parallel factor analysis. *Neuroimage*, 2004; 22: 1035-45.
- Moretti DV, Babiloni C, Binetti G, Cassetta E, Dal Forno G, Ferreric F, Ferri R, Lanuzza B, Miniussi C, Nobili F. Individual analysis of EEG frequency and band power in mild Alzheimer's disease. *Clin. Neurophysiol.*, 2004; 115: 299-308.
- Mørup M, Hansen LK, Arnfred SM. Algorithms for sparse nonnegative Tucker decompositions. *Neural Comput.*, 2008; 20: 2112-31.
- Musha T, Asada T, Yamashita F, Kinoshita T, Chen Z, Matsuda H, Uno M, Shankle WR. A new EEG method for estimating cortical neuronal impairment that is sensitive to early stage Alzheimer's disease. *Clin. Neurophysiol.*, 2002; 113: 1052-8.
- Myung IJ. The importance of complexity in model selection. *J. Math. Psychol.*, 2000; 44: 190-204.
- Nobili F, Copello F, Vitali P, Prastaro T, Carozzo S, Perego G, Rodriguez G. Timing of Disease Progression by Quantitative EEG in Alzheimer's Patients. *J. Clin. Neurophysiol.*, 1999; 16: 566.
- Nuwer MR, Comi G, Emerson R, Fuglsang-Frederiksen A, Guerit JM, Hinrichs H, Ikeda A, Jose C, Luccas F, Rappelsburger P. IFCN standards for digital recording of clinical EEG. *Electroencephalogr. Clin. Neurophysiol.*, 1998; 106: 259-61.
- Pietrzik C, Behl C. Concepts for the treatment of Alzheimer's disease: molecular mechanisms and clinical application. *Int. J. Exp. Pathol.*, 2005; 86: 173.
- Pijnenburg YAL, Vd Made Y, Van Cappellen van Walsum AM, Knol DL, Scheltens P, Stam CJ. EEG synchronization likelihood in mild cognitive impairment and Alzheimer's disease during a working memory task. *Clin. Neurophysiol.*, 2004; 115: 1332-9.
- Pravdova V, Boucon C, de Jong S, Walczak B, Massart DL. Three-way principal component analysis applied to food analysis: an example. *Anal. Chim. Acta*, 2002; 462: 133-48.
- Pritchard WS, Duke DW, Coburn KL, Moore NC, Tucker KA, Jann MW, Hostetler RM. EEG-based, neural-net predictive classification of Alzheimer's disease versus control subjects is augmented by non-linear EEG measures. *Electroencephalogr. Clin. Neurophysiol.*, 1994; 91: 118-30.
- Pucci E, Belardinelli N, Cacchio G, Signorino M, Angeleri F. EEG power spectrum differences in early and late onset forms of Alzheimer's disease. *Clin. Neurophysiol.*, 1999; 110: 621-31.
- Rossini PM, Buscema M, Capriotti M, Grossi E, Rodriguez G, Del Percio C, Babiloni C. Is it possible to automatically distinguish resting EEG data of normal elderly vs. mild cognitive impairment subjects with high degree of accuracy? *Clin. Neurophysiol.*, 2008.
- Schreiter-Gasser U, Gasser T, Ziegler P. Quantitative EEG analysis in early onset Alzheimer's disease: correlations with severity, clinical characteristics, visual EEG and CCT. *Electroencephalogr. Clin. Neurophysiol.*, 1994; 90: 267.
- Signorino M, Pucci E, Belardinelli N, Nolfe G, Angeleri F. EEG spectral analysis in vascular and Alzheimer dementia. *Electroencephalogr. Clin. Neurophysiol.*, 1995; 94: 313-25.
- Tanzi RE, Bertram L. New Frontiers in Alzheimer's Disease Genetics. *Neuron*, 2001; 32: 181-4.
- Tichavsky P, Yeredor A, Nielsen J. A fast approximate joint diagonalization algorithm using a criterion with a block diagonal weight matrix. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP): Las Vegas, USA*, 2008: 3321-4.
- Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika*, 1966; 31: 279-311.
- van der Hiele K, Vein AA, Reijntjes R, Westendorp RGJ, Bollen E, van Buchem MA, van Dijk JG, Middelkoop HAM. EEG correlates in the spectrum of cognitive decline. *Clin. Neurophysiol.*, 2007; 118: 1931-9.
- Vialatte FB, Sole-Casals J, Maurice M, Latchoumane C, Hudson N, Wimalaratna S, Jeong J, Cichocki A. Improving the Quality of EEG Data in Patients with Alzheimer's Disease Using ICA. *Springer*, 2009: 986.
- Waldemar G, Høgh P, Paulson OB. Functional brain imaging with single-photon emission computed tomography in the diagnosis of Alzheimer's disease. *Int. Psychogeriatr.*, 2005; 9: 223-7.
- Wang L, Zang Y, He Y, Liang M, Zhang X, Tian L, Wu T, Jiang T, Li K. Changes in hippocampal connectivity in the early stages of Alzheimer's disease: evidence from resting state fMRI. *Neuroimage*, 2006; 31: 496-504.
- Woon WL, Cichocki A, Vialatte F, Musha T. Techniques for early detection of Alzheimer's disease using spontaneous EEG recordings. *Physiol. Meas.*, 2007; 28: 335-47.
- Wu W, Gao X, Hong B, Gao S. Classifying Single-Trial EEG During Motor Imagery by Iterative Spatio-Spectral Patterns Learning (ISSPL). *IEEE Transaction on Biomedical Engineering*, 2008; 55: 1733-43.
- Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO. Development and validation of a geriatric depression screening scale: a preliminary report. *Psychiatric Research*, 1983; 17: 37-49.

SUPPLEMENTARY INFORMATION 1

Algorithm for Multiple Unknown Signals Extraction (AMUSE)

AMUSE is a BSS technique that relies on second-order statistics for spatiotemporal decorrelation. In its original usage, components are extracted from an EEG signal (i.e., combining several observables and/or electrodes) and the five most significant components are selected based on the ‘best linear predictability’ criterion. The assumption underlying its use is that components are spatiotemporally decorrelated and less complex, as indicated by their linear predictability, compared to a mixture of EEG components. The five selected components are ordered from the higher to the lower value of the singular value of their time-delayed covariance matrix.

Similar to PCA, the components extracted by AMUSE are uniquely defined and consistently ranked (unlike ICA). The algorithm is applied to the multichannel EEG time series $x(t)$ and can be described in two steps: (1) *prewhitening of the data*: application of standard and robust prewhitening (sphering) as the linear transformation $x_1(t) = Qx(t)$ with

$Q = R_x^{-1/2} = (V\Lambda V^T)^{-1/2} = V(\Lambda)^{-1/2} V^T$, where $R_x = E[x(t)x^T(t)]$ is a standard covariance matrix; and (2) *singular*

value decomposition (SVD): SVD can be obtained for a time-delayed covariance matrix using

$R_{x_1x_1} = E[x_1(t)x_1^T(t-1)] = U\Sigma V^T$, where Σ is a diagonal matrix with decreasing singular values (i.e., ordering of source),

and U and V are orthogonal matrices of left and right singular vectors. The final unmixing (separation of sources using observables) matrix W can then be estimated as the following factorization: $w = U^T Q$. The PCA used for AMUSE implementation was part of the ICALAB signal processing package [1].

Singular Value Decomposition (SVD)

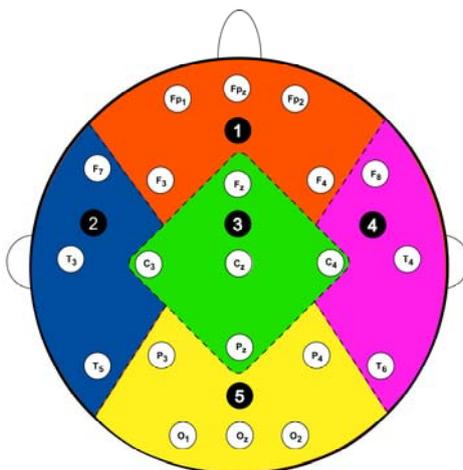
In this paper, SVD algorithm was applied to a the subject-mode unfolded tensor of the database A, A (subject). Using this approach, we can confirm that the introduction of additional variable (i.e. matrix form) or mode (i.e. tensor form) can be exploited differently. The SVD decomposition was performed using MATLAB® (MathWorks, Inc.). The number of component was set to three (i.e. corresponding to 90% of the explained variance). The reduced matrix (*Subject x Component*) was used for further classification in training condition and leave-one-out cross-validation condition (Table 3 and supplementary Table 3).

[1] A. Cichocki, S. Amari, K. Siwek, and T. Tanaka, "ICALAB Toolboxes," in *URL: <http://www.bsp.brain.riken.jp/ICALAB>*, 2007.

SUPPLEMENTARY INFORMATION 2

Topological Grouping of Electrodes

Supplementary Fig. 1. Topological grouping of electrodes in five regions. Numbers 1, 2, 3, 4 and 5 denote the frontal, left temporal, central, right temporal and posterior regions, respectively.



SUPPLEMENTARY INFORMATION 3

Supplementary Table 1 and 2 (RMSE values) associated with Table 2 and 3 (Classification accuracy), respectively

Supplementary Table 1. Comparison of the classification RMSE after LOO cross-validation for the filtered data (frequency band only, no regional grouping), AMUSE algorithm, and SSF on filtered data; for nonlinear neural network model (# hidden neuron > 0), it provides the mean and standard deviation of RMSE, and the lowest RMSE is shown in bracket. The RMSE obtained by linear classification is marked in bold.

		# of Hidden Neurons	0	1	2	3
Filtered Data	Training	0.3 8	0.38±0.0 2 (0.33)	0.22±0.1 1 (0)	0	
	Validation	0.5 2	0.54±0.0 3 (0.46)	0.56±0.0 5 (0.41)	0.53±0.0 5 (0.43)	
AMUSE on Filtered Data	Training	0.5 1	0.34±10 ⁻⁴ (0.34)	0.26±0.0 7 (0.16)	0.14±0.1 1 (0)	
	Validation	0.5 1	0.46±0.0 2 (0.44)	0.58±0.0 3 (0.50)	0.64±0.0 4 (0.55)	
SSF on Filtered Data	Training	0.2 1	0.25±0.1 3 (0.21)	0.22±0.1 3 (0.14)	0.19±0.1 5 (0)	
	Validation	0.3 5	0.36±0.0 4 (0.27)	0.46±0.0 3 (0.26)	0.48±0.0 4 (0.22)	

Supplementary Table 2. Comparison of the classification RMSE after LOO cross-validation for the SSF on ICA processed data, SVD on ICA processed and matricized data, and PARAFAC (R=2) and NTD (P, Q, R) = (2, 1, 2) on ICA processed data; for nonlinear neural network model (i.e. # hidden neuron > 0), it provides the mean and standard deviation of accuracy or RMSE, and the min RMSE is shown in bracket. Unique solutions are found for linear model of neural network (i.e. # hidden neuron = 0), hence no mean is provided. The RMSE obtained by linear classification is marker in bold.

		# of Hidden Neurons	0	1	2	3
SSF on ICA processed Data	Training	0.1 6	0.19±0.1 2 (0.15)	0.19±0.1 4 (0.13)	0.16±0.1 7 (0)	
	Validation	0.3 0	0.37±0.0 3 (0.28)	0.36±0.0 4 (0.23)	0.37±0.0 4 (0.27)	
SVD on ICA processed matricized Data	Training	0.1 5	0.16±0.0 2 (0.15)	0.16±0.2 3 (0.08)	0.03±0.0 6 (0)	

	Validation	0.35	0.31±0.05 (0.31)	0.34±0.04 (0.30)	0.34±0.04 (0.27)
PARAFAC	Training	0.16	0.26±0.16 (0.15)	0.15±0.16 (0)	0.14±0.14 (0)
	Validation	0.30	0.35±0.04 (0.25)	0.33±0.04 (0.24)	0.31±0.04 (0.21)
NTD	Training	0.16	0.22±0.15 (0.15)	0.21±0.15 (0.15)	0.19±0.13 (0.15)
	Validation	0.20	0.28±0.04 (0.20)	0.26±0.05 (0.16)	0.24±0.05 (0.10)

Classification using PARAFAC and SSFs on the filtered Data

We tested the classification performance of the selected models (i.e., after training, validation, and regularization on ICA-processed data) on filtered data (i.e., EEG time series that were filtered only). As shown in Supplementary Table 3, up to the regularization database (database B), all models exhibit similar performance for ICA-processed data except SSF, which exhibits poorer performance on the training set (database A; accuracy > 85.4%). The difference between the accuracies of PARAFAC (> 90.9%) and the SSF (> 72.7%) is even more evident on the test set (database C). The information extracted from the ICA-processed data (weighted contribution, modal interaction) seems to lose some accuracy when applied to filtered data. This demonstrates the importance of data being artifact-free such that any discriminative characteristics remain unaffected. Moreover, the method based on PARAFAC provides better generalization of the subject distinction from an unsupervised perspective as compared with that of supervised SSF.

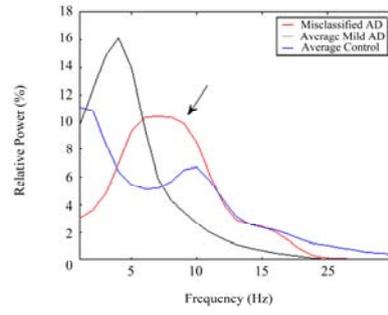
Supplementary Table 3. Classification performance on the filtered data using the linear classifier trained, validated, and regularized on ICA processed data using the PARAFAC method.

		Database A	Database B	Database C
PARAFAC	- Accuracy (%)	90.2%	90.0%	90.9%
	- RMSE	0.31	0.31	0.41
SSF	- Accuracy (%)	85.4%	90.0%	72.7%
	- RMSE	0.36	0.31	0.49

SUPPLEMENTARY 5

Misclassified Subjects in the Regulation Process (Database B)

We analyzed the misclassified subject present in database B. The subject presented in Supplementary Fig. 1 was diagnosed with mild AD (female; MMSE = 21; two years of higher education, mean group = 8.3). However, the global spectrum of this subject exhibited outlier properties as compared with the mean frequency repartition of control and AD subjects. This subject was systematically misclassified by all methods used in this paper.



Supplementary Fig. 2. Misclassified subject's Fourier spectrum in database B (regularization database) compared with the average Fourier spectrum of control and AD groups.