Modeling and Detection of Epileptic Seizures using Multi-modal Data Construction and Analysis

Evrim Acar, *Computer Science Department, Rensselaer Polytechnic Institute, Troy, NY USA*
Canan Aykut-Bingol, *Department of Neurology, Yeditepe University, Istanbul, Turkey*
Haluk Bingol, *Computer Engineering Department, Bogazici University, Istanbul, Turkey*
Rasmus Bro, *Faculty of Life Sciences, Copenhagen University, Copenhagen, Denmark*
Anthony L. Ritaccio, *Department of Neurology and Neurosurgery, Albany Medical College, NY, USA*
Bülent Yener, *Computer Science Department, Rensselaer Polytechnic Institute, Troy, NY USA*

***Abstract*— The identification of epileptic seizures significantly relies on monitoring and visual analysis of large amounts of multi-channel electroencephalographic (EEG) signals. With a goal of automating this time-consuming and subjective task, we develop a patient-specific seizure recognition model for multi-channel scalp EEG signals.**

We differentiate between seizure and non-seizure periods by representing multi-channel EEG signals using a set of features from both time and frequency domains. Our contributions are threefold: First, we rearrange multi-channel EEG recordings as a third-order tensor called an Epilepsy Feature Tensor with modes: time epochs, features and channels. Second, we model the Epilepsy Feature Tensor using a multi-linear discriminant analysis based on Multi-linear Partial Least Squares, which is the generalization of Partial Least Squares regression to tensors. This two-step approach facilitates the analysis of EEG data from multiple channels represented by several features from different domains. Third, our multi-modal approach enables us to understand the differences between seizures of different patients by finding a subset of features capturing the seizure characteristics of each patient.

We evaluate the performance of our model considering both sensitivity and specificity. Our results based on the analysis of 29 seizures from 8 patients demonstrate that multiway analysis of an Epilepsy Feature Tensor can detect patient-specific seizures with g-means (geometric mean of sensitivity and specificity) ranging between 77%-97%. Furthermore, we compare our model with a two-way model and demonstrate that our multi-modal approach can improve a two-way analysis approach in terms of detecting and understanding epileptic seizures.

I. Introduction

Monitoring and analysis of EEG signals is one of the diagnosis tools used in identifying epileptic seizure onsets, localizing seizure origins and determining the adequate type of treatments like medications or surgeries. Large amounts of multi-channel EEG signals are visually analyzed by neurologists with a goal of understanding when and where the seizures start and how they propagate within the brain. However, visual analysis of EEG signals has some drawbacks. It is a time-consuming and subjective task. Furthermore, it is error-prone due to fatigue, etc. Therefore, automation of the detection of the underlying brain dynamics in EEG signals is significant in order to obtain fast and objective EEG analysis.

A common approach in seizure recognition/detection and also in prediction is to extract information; in other words, features that can characterize seizure morphologies, from EEG recordings [1]–[5]. The procedure for feature extraction from multi-channel EEG data is often as follows: First, an EEG signal from a channel is divided into I time epochs (overlapping or non-overlapping) and then J features are extracted from each epoch. Consequently, a signal from a single channel can be represented as a matrix of size $I \times J$. A great deal of effort from different disciplines has been invested in exploring the features in order to define the signature of a seizure. These features include statistical complexity measures (e.g., fractal dimension, approximate entropy, lyapunov exponents, etc.) as well as other features from time (e.g., higher-order statistics of the signal in time domain, Hjorth parameters, etc.) and frequency domains (e.g., spectral skewness, spectral entropy, etc.). A list of features used in characterization of epileptic seizure dynamics can be found in recent studies [3]–[5].

In the literature, studies use either multiple features from a single channel or a single feature from multiple channels since data construction and data analysis techniques are often restricted to two dimensions. For instance, in [3], seizure dynamics are analyzed solely on a specific recording, which represents the characteristics of a seizure well. Then the performance of various features from different domains on that particular signal is analyzed simultaneously. On the other hand, [5] analyzes multi-channel EEG data but assesses the performance of each feature one at a time. Furthermore, different studies extract different features and employ different algorithms to distinguish between seizure and non-seizure periods (e.g., [6] and references therein), which makes it difficult to compare the performance of features. An approach capable of simultaneously analyzing features would enable the performance comparison of the features on the same data using the same classifier. Simultaneous analysis of features is also important because it may consider linear or non-linear combinations of features. While a single feature may not be very effective in discriminating between epileptic periods, combinations of several features may well be [7]. Taking into consideration the challenges addressed in the literature, in this study we introduce a multi-modal data construction and analysis approach, which rearranges signals from K channels as a third-order tensor of size $I \times J \times K$ as shown in Figure 1. We then model this tensor using multi-linear discriminant analysis by facilitating simultaneous analysis of EEG data from multiple channels based on several features from different domains.

In this study, we are particularly interested in distinguishing a seizure (ictal) period from a pre-seizure (pre-ictal) and a post-seizure (post-ictal) period. Moreover, we want to be able to characterize seizures of patients using a subset of features and understand the differences between seizures of different patients.

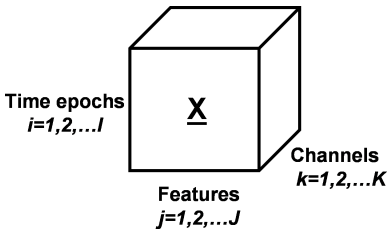

Fig. 1. Epilepsy Feature Tensor. $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ represents the multi-channel EEG data transformed into the feature space by computing certain measures characterizing seizure dynamics. Each entry of $\underline{\mathbf{X}}$, x_{ijk}, corresponds to the value of j^{th} feature of i^{th} time epoch at k^{th} channel.

Our ultimate goal is to mark the seizure period but not to predict an upcoming seizure or to detect the seizure onset with minimum delay. This study, therefore, differs from the related work on seizure detection and prediction, e.g., [1], [5], [8]. They either focus on the identification of features distinguishing between inter-ictal and pre-ictal periods or aim to detect an epileptic seizure with possible minimum delay using features from a particular domain. Nevertheless, multiway data construction and analysis approach introduced in this paper can be easily extended to seizure prediction and detection.

Multi-linear models have been previously employed in several applications in neuroscience. In [9], EEG data and data collected through experiments with different doses of a drug are arranged as a six-way array with modes: EEG, patients, doses, conditions, time and channels. The analysis of the six-way dataset demonstrates that significant information is successfully extracted from a complex drug dataset by a multi-linear model rather than two-way models such as Principal Component Analysis (PCA). Multiway models have become more popular in neuroscience with the idea of decomposing EEG data into space, time and frequency components [10]. The three-way array constructed from multi-channel EEG data in [10] with modes $time\ samples \times frequency \times channels$ is analyzed using a multi-linear component model called Parallel Factor Analysis (PARAFAC) [11]. Components extracted by a PARAFAC model are observed to demonstrate the temporal, spectral and spatial signatures of EEG. PARAFAC models with nonnegativity constraints are later used in another study on event-related potentials (ERP) to find the underlying structure of brain dynamics [12]. These studies have also motivated the application of multiway models for understanding the structure of epileptic seizures [13]–[15]. Similar to the three-way array constructed in [10], multi-channel ictal EEG data are arranged as a third-order tensor with modes $time\ samples \times frequency \times channels$ using the power of wavelet coefficients in [13] and [14] and using pure wavelet coefficients in [15]. Components extracted by PARAFAC are later used to explore the signatures of a seizure in the frequency and time domains as well as localize the seizure origin. Based on the extracted signatures, artifacts have also been identified and later removed by multi-linear subspace analysis in [14]. In addition to the applications of multi-linear component models, multi-linear regression models have also been previously employed in neuroscience, e.g., in [16] for extracting the connection between EEG and fMRI (functional magnetic resonance imaging) recordings.

A. Our Contributions

We address the problem of identifying an epileptic seizure automatically from multi-channel scalp EEG signals. We introduce a novel approach, which combines the seizure recognition power of several features from different domains and classifies epochs of signals from multiple channels as seizure or non-seizure periods. This paper is an extension of our preliminary study on seizure recognition using Epilepsy Feature Tensors [17] and our contributions are as follows:

- We rearrange multi-channel scalp EEG recordings as a third order tensor, Epilepsy Feature Tensor, with modes: $time\ epochs \times features \times channels$. We extract features from both the time and frequency domains and represent a signal using a set of feature vectors. We have omitted some of the features used in [17] and added new features like mean absolute slope and spatial information. We do not make any assumptions about the seizure origin and analyze the signals from all channels simultaneously.
- We model Epilepsy Feature Tensors using multi-linear discriminant analysis based on Multi-linear Partial Least Squares (N-PLS) and Linear Discriminant Analysis (LDA). We develop a patient-specific seizure recognition model and compare the performance of this multi-modal approach with that of a two-way approach based on Support Vector Machines (SVM).
- We extend a feature selection method used in two-way regression analysis to three-way regression models. Feature selection enables us to determine a subset of features for each patient, which can improve our understanding of the differences between seizures of different patients.

The organization of this paper is as follows: In Section 2, we include a brief introduction on higher-order datasets and multi-linear regression models. Features extracted from EEG signals and the characteristics of the EEG dataset are described concisely in Section 3 and 4, respectively. We introduce our seizure recognition model based on N-PLS and LDA in Section 5. The performance of the model on the sample EEG dataset is discussed in Section 6. We conclude, in Section 7, with future directions in seizure recognition.

II. Methodology

Regression models, e.g., multiple linear regression, PLS and Principal Component Regression, are commonly applied in prediction and classification problems in diverse disciplines. While these models are employed on datasets of order no higher than two (vectors or matrices), the independent variable in this study, i.e., multi-channel EEG data, is a third-order tensor (Figure 1). This section briefly introduces higher-order arrays and the regression model, i.e., Multi-linear Partial Least Squares, developed for higher-order data analysis.

A. Notation and Background

Multiway arrays, also referred to as tensors, are higher-order generalizations of vectors and matrices. Higher-order arrays are represented as $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$, where the order of $\underline{\mathbf{X}}$ is N ($N > 2$) while a vector and a matrix are arrays of order 1 and 2, respectively. In higher-order array terminology, each dimension of a multiway array is called a mode (way) and the number of variables in each mode is used to indicate the dimensionality of

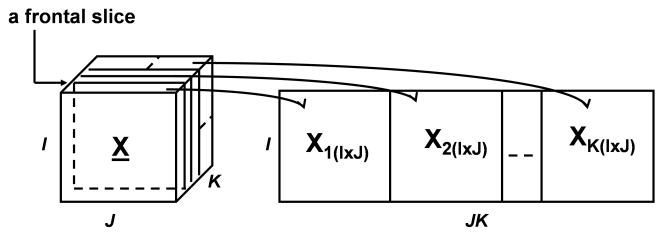

Fig. 2. Matricization of a three-way array in the first mode. A three-way array $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ is unfolded in the first mode and a matrix of size $I \times JK$, denoted by $\mathbf{X}_{(1)}$, is formed. Subscript in $\mathbf{X}_{(i)}$ indicates the mode of matricization.

a mode. For instance, $\underline{\mathbf{X}} \in \mathbb{R}^{I_1 \times I_2 \ldots \times I_N}$ is a multiway array with N modes (called N-way array or N^{th}-order tensor) with I_1, I_2, ...,I_N dimensions in the first, second, ... , N^{th} mode, respectively.

A multiway array can be rearranged as a two-way array by unfolding the slices in a certain mode, e.g., the first mode as shown in Figure 2. This operation is called matricization (or unfolding/flattening). Rearranging multiway arrays as two-way datasets enables the application of traditional component and regression models for two-way datasets on multiway arrays. However, analyzing multiway datasets with two-way methods may result in a more complex model harder to interpret and in some cases with low prediction accuracy if the data are noisy [18]. Therefore, we preserve the multi-modality of the dataset and employ a generalized version of a regression model, i.e., PLS, to higher-order arrays.

We denote higher-order arrays using underlined boldface letters, e.g., $\underline{\mathbf{X}}$, following the standard notation in the multiway literature [19]. Matrices and vectors are represented by boldface capital, e.g., $\mathbf{X}$, and boldface lowercase letters, e.g., $\mathbf{x}$, respectively. Scalars are denoted by lowercase or uppercase letters, e.g., x or X.

B. Multi-linear Partial Least Squares (N-PLS)

Multi-linear PLS is introduced as a generalization of PLS to multiway datasets [20]. This method can handle the situations where dependent and/or independent variables are multiway arrays. In this study, we confine our attention to the case where the independent variable, $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$, is a three-way array of type Epilepsy Feature Tensor and the dependent variable, $\mathbf{y} \in \mathbb{R}^I$, is a vector containing the labels of time epochs (seizure or non-seizure). Multi-linear PLS models the dataset $\underline{\mathbf{X}}$ by extracting a component, $\mathbf{t} \in \mathbb{R}^I$, from the first mode such that $cov(\mathbf{t}, \mathbf{y})$ is maximized. A pre-defined number of components, N, is extracted iteratively and the matrix $\mathbf{T} \in \mathbb{R}^{I \times N}$, whose columns are the extracted components ($\mathbf{t}$'s), is constructed. In addition to $\mathbf{T}$, component matrices, $\mathbf{W}^J$ and $\mathbf{W}^K$, corresponding to the second and third modes, respectively are also formed. The steps of the algorithm are briefly summarized in Algorithm 1 and discussed in detail in [21].

One advantage of N-PLS over two-way regression analysis is that when we use N-PLS, we obtain component matrices corresponding to each mode of a third-order Epilepsy Feature Tensor: $\mathbf{T}$, $\mathbf{W}^J$ and $\mathbf{W}^K$ corresponding to the time epochs, features and channels modes. We will see in Section 5 how extracting components separately from each mode makes feature selection possible. If we used PLS on an unfolded dataset of type: *time epochs* by $features - channels$, then we would have only the component matrices $\mathbf{T}$ and $\mathbf{W}$, where $\mathbf{W}$ would correspond to both features and channels modes, making the interpretation and feature selection much harder.

Algorithm 1 Multi-linear PLS($\underline{\mathbf{X}}$, $\mathbf{y}$, N)

$\mathbf{y}_0 = \mathbf{y}$, $\mathbf{X}_0 = \mathbf{X}_{(1)}$
for $i = 1$ to N **do**
  $\mathbf{z} = \mathbf{y}_{i-1}^T \mathbf{X}_{i-1}$
  Reshape $\mathbf{z}$ as a matrix $\mathbf{Z} \in \mathbb{R}^{J \times K}$ such that $\mathbf{Z}(m,n) = \mathbf{z}(m + J * (n-1))$
  {Compute singular value decomposition of matrix $\mathbf{Z}$}
  $\mathbf{Z} = \mathbf{U}\mathbf{S}\mathbf{V}^T$
  $\mathbf{w}^J = \mathbf{U}(:,1)$, $\mathbf{w}^K = \mathbf{V}(:,1)$
  $\mathbf{W}^J(:,i) = \mathbf{w}^J$, $\mathbf{W}^K(:,i) = \mathbf{w}^K$
  $\mathbf{T}(:,i) = \mathbf{X}_{i-1}(\mathbf{w}^K \otimes \mathbf{w}^J)$
  $\mathbf{X}_i = \mathbf{X}_{i-1} - \mathbf{T}(:,i)(\mathbf{w}^K \otimes \mathbf{w}^J)'$
  $\mathbf{b}_i = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{y}_{i-1} = \mathbf{T}^+\mathbf{y}_{i-1}$
  {Regression and Deflation}
  $\mathbf{y}_i = \mathbf{y}_{i-1} - \mathbf{T}\mathbf{b}_i = (\mathbf{I} - \mathbf{T}\mathbf{T}^+)\mathbf{y}_{i-1}$
end for

*

$\mathbf{X}_{(1)}$ stands for the tensor $\underline{\mathbf{X}}$ matricized in the first mode. $\mathbf{X}_i$ indicates matricized data in the first mode updated/deflated by the computation of i components. $\mathbf{A}(i,j)$ represents the entry of matrix $\mathbf{A}$ at the i^{th} row and j^{th} column while $\mathbf{A}(:,j)$ represents the j^{th} column of matrix A. $\mathbf{W}^J$ and $\mathbf{W}^K$ correspond to the component matrices in the second and third modes, respectively. $\mathbf{T}^+$ stands for pseudo-inverse defined as $\mathbf{T}^+ = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$. $\otimes$ indicates the Kronecker product [22].

III. Features

An EEG recording from a single channel is a sequence of time samples. One approach for analyzing a time series is to divide the time series into time epochs and inspect whether there are certain underlying dynamics in a particular epoch. This could be achieved by extracting measures that characterize those dynamics. Then each epoch can be represented using a set of measures called *features*. Let $s(j)$ denote the time sample at time j and $\mathbf{s} = \{s(1), s(2), ...s(N)\}$ be the time sequence for a particular epoch of length N. We represent each feature as $f_i(s)$, which denotes the i^{th} feature computed on the time epoch $\mathbf{s}$. In this section, we give brief definitions of the features used in this paper.

A. Time domain

1) Hjorth parameters: Hjorth parameters including activity, mobility and complexity are computed as defined in [3] as follows:

$$\begin{aligned} Activity: &\quad f_1(s) = \sigma_s^2 \\ Mobility: &\quad f_2(s) = \sigma_{s'}/\sigma_s \\ Complexity: &\quad f_3(s) = (\sigma_{s''}/\sigma_{s'})/(\sigma_{s'}/\sigma_s) \end{aligned}$$

where σ_s stands for the standard deviation of a time sequence $\mathbf{s}$; $\mathbf{s}'$ and $\mathbf{s}''$ denote the first and second difference of a time series $\mathbf{s}$, respectively. d^{th} difference of a time series can be denoted as follows $(1 - \boldsymbol{B})^d s(t)$, where $\boldsymbol{B}$ is the backshift operator. The backshift operator applied to a time sample can be represented as $\boldsymbol{B}^j s(t) = s(t-j)$ [23].

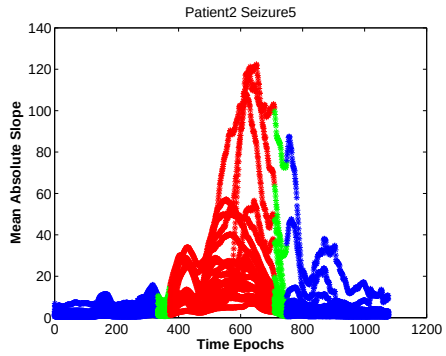

Fig. 3. Mean Absolute Slope of epochs from all channels for the fifth seizure of the second patient. Epochs marked with blue and red belong to non-seizure and seizure periods, respectively. Green epochs are the transition epochs from pre-seizure to seizure or seizure to post-seizure periods.

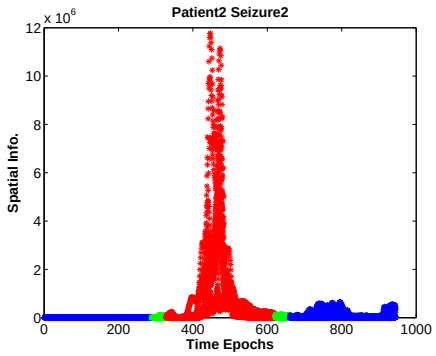

Fig. 4. Spatial Information of epochs from all channels for the second seizure of the second patient. Epochs marked with blue and red belong to non-seizure and seizure periods, respectively. We observe a clear increase in similarity between neighboring channels during a seizure period. Green epochs are the transition epochs from pre-seizure to seizure or seizure to post-seizure periods.

2) Mean Absolute Slope: Absolute slope is calculated using the consecutive differences between time samples in a time sequence: $AS(t) = |s(t+1) - s(t)|$ for each time sample $s(t)$ in a time epoch $\mathbf{s}$ [24]. In addition to its simplicity and efficiency, absolute slope can capture both high-amplitude slow and low-amplitude fast activities, that are often observed on seizure onsets. We extract the mean of absolute slopes computed for each time sample in a time epoch as the fourth feature, $f_4(s)$ (Figure 3). This feature should be a more reliable feature for intracranial EEG recordings, which are not contaminated with artifacts and less reliable, in our case, for scalp EEG recordings often contaminated with artifacts. However, we have observed that this feature contributes to seizure recognition in half of the patients in our dataset [Table III].

3) Spatial Information: During visual analysis, neurologists take into consideration not only the signal from a single channel but also the activity in other channels, especially in the neighboring channels and expect to observe synchronization. Therefore, in order to quantify the similarity between neighboring channels in each time epoch, we first define neighbors for each channel and then use the covariance between neighboring channels as a feature (Figure 4). Let $\mathbf{X}$ be a matrix of type: time samples by channels, for a particular time epoch $\mathbf{s}$. We define spatial information, the fifth feature extracted from an epoch s, for channel i as $f_5(s,i) = \sum_{j \in NEIGH_i} |\mathbf{C}_{ij}|$, where $NEIGH_i$ contains the neighbors of channel i and $\mathbf{C}$ is the covariance matrix corresponding to the channels in $\mathbf{X}$.

B. Frequency domain

1) Frequency Spectrum: We reduce the time series at least to a mean-stationary time series by taking the first difference of the signal before computing the amplitude spectrum. Given a time series $\mathbf{s}$ corresponding to a particular epoch, we use a Fast Fourier Transform (FFT) to obtain the Fourier coefficients, c_k, where $c_k = \frac{1}{N}\sum_{t=1}^{N} s(t) e^{-i\frac{2\pi k}{N}t}$. Based on the Fourier coefficients, we construct the amplitude spectrum using $|c_k|$. The amplitude spectrum is then used to extract the sixth feature, $f_6(s)$, which is the median frequency.

2) Spectral Entropy: The last feature is a measure used to quantify the uncertainty in the frequency domain. Five frequency bands in accordance with the traditional EEG frequency bands ([25] and references therein) are chosen: δ (0.5 - 3.5Hz), θ (3.5 - 7.5Hz), α (7.5 - 12.5Hz), β (12.5 - 30Hz), γ (> 30Hz). We apply continuous wavelet transform between 0.5-50Hz using a Mexican-hat wavelet as the mother wavelet on each epoch. Wavelet coefficients are later used to observe the energy spread across these five frequency bands in each epoch. Let E_f be the estimate of the energy in frequency band f and E_T be the estimate for the total energy in all frequency bands computed as follows:

$$E_f = \sum_{i=1}^{N}\sum_{j=1}^{S} |c_{ij}|^2$$

$$E_T = \sum_{f=1}^{5} E_f$$

where c_{ij} denotes the wavelet coefficient corresponding to the i^{th} time sample in an epoch and j^{th} scale, N is the length of an epoch and S is the number of scales. We compute spectral entropy, H, using Shannon's entropy measure [26] as follows: $H = -\sum_{f=1}^{5} \frac{E_f}{E_T} log(\frac{E_f}{E_T})$, which is the seventh feature extracted from an epoch $\mathbf{s}$, $f_7(s)$.

The list of these features can be easily extended by adding vertical slices to the three-way dataset given in Figure 1.

IV. Data

Our dataset contains multi-channel scalp EEG recordings of 29 seizures from 8 patients suffering from focal epileptic seizures. The EEG data have been collected via scalp electrodes in the epilepsy monitoring unit of Yeditepe University Hospital and Albany Medical College. The recording of EEG with referential electrode Cz is used for computational analysis. The number of seizures per patient as well as sizes of Epilepsy Feature Tensors with modes: time epochs, features and channels, are given in Table I. EEG recordings are not preprocessed to remove artifacts. The data for the first patient are sampled at 200Hz and the data for other patients are sampled at 400Hz. The signals are filtered at 50 Hz (for the data from Yeditepe University) and 60 Hz (for the data from Albany Medical College) to remove the artifacts from the power source.

The data corresponding to a seizure of a patient contain a certain amount of data right before the seizure, the seizure period and a certain amount of data right after the seizure period. We try to include data from pre-seizure and post-seizure periods, each as long as the seizure duration. Each signal is divided into epochs of 10 sec. (an epoch typically contains 2000 or 4000 samples depending on the sampling frequency.). The epochs are

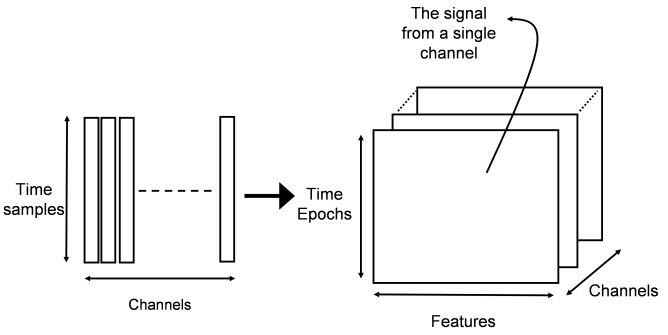

Fig. 5. Construction of an Epilepsy Feature Tensor from multi-channel EEG data.

TABLE I

EEG DATASET. MULTI-CHANNEL SCALP EEG SIGNALS FROM EPILEPSY PATIENTS WITH AT LEAST THREE RECORDED SEIZURES ARE INCLUDED IN OUR ANALYSIS. THE LAST COLUMN GIVES THE SIZE OF THE THIRD-ORDER TENSOR CONSTRUCTED FOR EACH SEIZURE OF A PATIENT. EACH SUCH TENSOR CONTAINS SOME DATA BEFORE AND AFTER SEIZURE AS WELL AS THE SEIZURE PERIOD.

PatientId	SeizureId	Size of Epilepsy Feature Tensor
1	1	$302 \times 7 \times 18$
	2	$386 \times 7 \times 18$
	3	$320 \times 7 \times 18$
	4	$398 \times 7 \times 18$
	5	$444 \times 7 \times 18$
2	1	$878 \times 7 \times 18$
	2	$866 \times 7 \times 18$
	3	$902 \times 7 \times 18$
	4	$986 \times 7 \times 18$
	5	$998 \times 7 \times 18$
3	1	$790 \times 7 \times 18$
	2	$746 \times 7 \times 18$
	3	$1034 \times 7 \times 18$
4	1	$1174 \times 7 \times 18$
	2	$1346 \times 7 \times 18$
	3	$1170 \times 7 \times 18$
5	1	$62 \times 7 \times 18$
	2	$74 \times 7 \times 18$
	3	$458 \times 7 \times 18$
6	1	$226 \times 7 \times 18$
	2	$186 \times 7 \times 18$
	3	$186 \times 7 \times 18$
	4	$186 \times 7 \times 18$
7	1	$638 \times 7 \times 18$
	2	$630 \times 7 \times 18$
	3	$578 \times 7 \times 18$
8	1	$866 \times 7 \times 18$
	2	$1082 \times 7 \times 18$
	3	$842 \times 7 \times 18$

formed using a sliding window approach such that consecutive epochs differ only in 100 samples. Seven features are computed for each epoch and a matrix of size *nb of time epochs* $\times$ 7 is created for a signal from a single channel. When all channels are included in the analysis, this forms a three-way array of *nb of time epochs* $\times 7 \times 18$ for each seizure (Figure 5).

The seizure period is visually identified by neurologists for each seizure of a patient. In accordance with the markings, the epochs are divided into two classes: epochs that belong to the seizure period and the ones outside the seizure period. The dependent variable, i.e., $\mathbf{y}$-vector in Algorithm 1, corresponding to the time epochs mode of an Epilepsy Feature Tensor is then constructed such that: $y_i = 1$ if i^{th} epoch is outside the seizure period and $y_i = 2$ if i^{th} epoch belongs to the seizure period. Since epochs are formed using a sliding window approach, some epochs contain samples from both pre-seizure and seizure periods or both seizure and post-seizure periods. These epochs are excluded from training and test sets so that the performance of the model is not affected by epochs containing the characteristics of different seizure dynamics.

V. SEIZURE RECOGNITION

We build our model on a training set constructed using all but one seizure of a patient together with the corresponding labels of the epochs. Once the training set is formed, we scale the three-way array within the features mode before the analysis since features have different ranges of magnitudes (See Figure 3 and Figure 4). Scaling a three-way array within one mode is different than scaling two-way datasets. Unlike matrices where columns or rows are scaled, in the three-way case, whole matrices need to be scaled [27]. Before the analysis, both independent and dependent data are also centered. We regress the data for all the seizures in the training set onto the $\mathbf{y}$-vector containing 1's and 2's (for non-seizure and seizure, respectively) using Multi-linear PLS regression and build a model based on Algorithm 1[1].

Since N-PLS is a regression method and we need a binary classifier to classify time epochs as seizure and non-seizure, we combine N-PLS with LDA. When we model the training set, $\underline{\mathbf{X}}_{train} \in \mathbb{R}^{I \times J \times K}$, using N-PLS, we extract the component matrices corresponding to each mode of a three-way array. Let $\mathbf{T}_{train} \in \mathbb{R}^{I \times N}$, $\mathbf{W}^J \in \mathbb{R}^{J \times N}$ and $\mathbf{W}^K \in \mathbb{R}^{K \times N}$ be the component matrices corresponding to the first, second and third modes, respectively. We can use this model to predict the labels of the time epochs in other EEG recordings of that particular patient; in other words the labels of the time epochs in our test set, which contains the left-out seizure and the recordings before and after that seizure (Figure 6). Let $\underline{\mathbf{X}}_{test} \in \mathbb{R}^{R \times J \times K}$ be a third-order tensor representing the time epochs in our test set. We can then compute $\mathbf{T}_{test} \in \mathbb{R}^{R \times N}$ using the component matrices $\mathbf{W}^J$ and $\mathbf{W}^K$ extracted from the training set based on the general formula derived in [22]:

$$\begin{aligned} \mathbf{R} &= [\mathbf{w}_1 \ \ (\mathbf{I} - \mathbf{w}_1\mathbf{w}_1^T)\mathbf{w}_2 \ \ldots \ (\prod_{n=1}^{N-1}(\mathbf{I} - \mathbf{w}_n\mathbf{w}_n^T))\mathbf{w}_N] \\ \mathbf{T}_{test} &= \mathbf{X}_{test\ (1)}\mathbf{R} \end{aligned}$$

where $\mathbf{X}_{test\ (1)}$ is the matrix formed by unfolding $\underline{\mathbf{X}}_{test}$ in the first mode and vector $\mathbf{w}_i$ equals to the Kronecker product of i^{th} column of matrices $\mathbf{W}^K$ and $\mathbf{W}^J$: $\mathbf{w}_i = \mathbf{w}_i^K \otimes \mathbf{w}_i^J$. Once we obtain the t-scores for the epochs in the test set, we can then determine the class (seizure or non-seizure) of each time epoch by comparing $\mathbf{T}_{test}$ with $\mathbf{T}_{train}$ through LDA using the discriminant function given in [28].

A. Feature Selection

Not every feature in our feature set may be a powerful discriminator between seizure and non-seizure dynamics. Therefore,

[1]Implementation of N-PLS in PLS_Toolbox (by Eigenvector Research Inc.) running under MATLAB is used for the analysis.

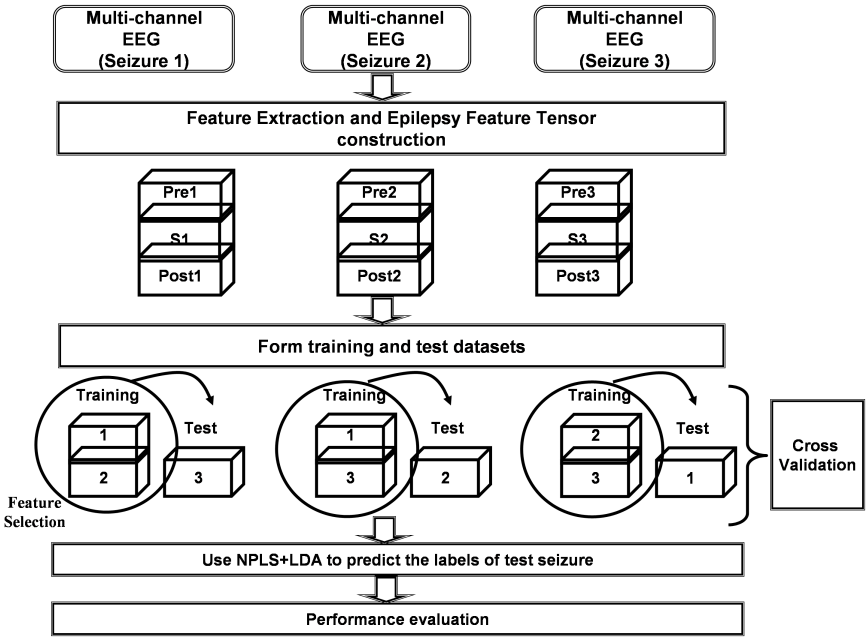

Fig. 6. Patient-specific Seizure Recognition Model. Multi-channel EEG signals corresponding to the data before, during and after each seizure of a patient are arranged as a third-order Epilepsy Feature Tensor. Then training and test sets are constructed by leaving out one seizure (together with data before and after that seizure period) at a time. The model built on the training set is used to predict the labels of the time epochs in the test set using NPLS and LDA. Final step is performance evaluation using the average performance of the model on test sets.

we identify the significant features for seizure recognition using a variable selection approach.

Our variable selection method is an extension of Variable Importance in Projection (VIP) to three-way datasets. VIP is used in two-way regression analysis and based on the idea of factor models. In linear factor models, several components summarizing the data are extracted either to explain the variance in the data, e.g., as in PCA, or to capture the correlation between two datasets, e.g., as in PLS or in Canonical Correlation Analysis. The components extracted in these linear factor models are linear combinations of the variables in the data. The variable selection method, VIP, computes a VIP-score for each variable in order to quantify a variable's importance by using the coefficient of a variable in each component together with each component's significance in regression. Variables with a VIP-score under a certain threshold are then removed from the data since they are considered insignificant. Let $\mathbf{X} \in \mathbb{R}^{I \times J}$ and $\mathbf{y} \in \mathbb{R}^{I}$ be the independent and dependent variables, $\mathbf{T} \in \mathbb{R}^{I \times N}$ represents the lower dimensional space $\mathbf{X}$ is mapped to and $\mathbf{b} \in \mathbb{R}^{N}$ contains the regression coefficients such that we can write $\mathbf{y} = \mathbf{Tb} + \mathbf{e}$ and $\mathbf{X} = \mathbf{TW} + \mathbf{E}$, where $\mathbf{e}$ and $\mathbf{E}$ contain the residuals. The VIP-score of the i^{th} variable is then calculated as follows [29]:

$$VIP_i = \sqrt{I \times \frac{\sum_{n=1}^{N} b_n^2 \mathbf{t}_n^T \mathbf{t}_n (w_{in}/|\mathbf{w}_n|)^2}{\sum_{n=1}^{N} b_n^2 \mathbf{t}_n^T \mathbf{t}_n}}$$

where $\mathbf{w}_n$ and $\mathbf{t}_n$ correspond to the n^{th} column of matrix $\mathbf{W}$ and $\mathbf{T}$, respectively and w_{in} is the entry in the i^{th} row of the n^{th} column of matrix $\mathbf{W}$. b_n is the regression coefficient for the n^{th} component; in other words, the n^{th} entry of vector $\mathbf{b}$.

Similarly, in N-PLS we extract component matrices corresponding to each mode of a higher-order dataset. Each column of a component matrix contains the coefficients corresponding to the variables in a specific mode and represents a component, which is a linear combination of the variables. Let the independent and dependent variables be $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ and $\mathbf{y} \in \mathbb{R}^{I}$, respectively and let $\mathbf{T} \in \mathbb{R}^{I \times N}$, $\mathbf{W}^J \in \mathbb{R}^{J \times N}$ and $\mathbf{W}^K \in \mathbb{R}^{K \times N}$ be the component matrices corresponding to the first (time epochs), second (features) and third (channels) modes. In the computation of VIP scores for variables in one mode of a three-way array, we replace matrix $\mathbf{W}$ with the component matrix in the mode where we select variables, in our case with $\mathbf{W}^J$ corresponding to the features mode. In addition, we project the data $\underline{\mathbf{X}}$ onto $\mathbf{W}^J$: $\mathbf{F} = \mathbf{X}_{(2)}\mathbf{W}^J$ and use the columns of matrix $\mathbf{F}$, i.e., $\mathbf{f}_n$, instead of t-scores.

$$VIP_i = \sqrt{I \times \frac{\sum_{n=1}^{N} b_n^2 \mathbf{f}_n^T \mathbf{f}_n (w_{in}^J/|\mathbf{w}_n^J|)^2}{\sum_{n=1}^{N} b_n^2 \mathbf{f}_n^T \mathbf{f}_n}}$$

Since the average of squared VIP scores equals 1, a general criterion is to select the variables with VIP score greater than 1. On the other hand, we just want to remove insignificant variables and include most of the variables contributing to seizure recognition in our analysis. Therefore, we lower the threshold to 0.7 and this threshold is set to the same value for all patients.

When we analyze Epilepsy Feature Tensors with N-PLS, we have the chance to select features independent of the channels because N-PLS models the data by constructing different component matrices for each mode. On the other hand, if we matricized an Epilepsy Feature Tensor, then we would obtain a matrix of $time\ epochs$ by $features - channels$. In that case, we would not be able to select only features but rather a feature from a particular channel since each variable would be a combination of features and channels.

B. Parameter Selection

As seen in Algorithm 1, the number of components in N-PLS, N, is a user-defined parameter. In order to determine N, we use cross-validation on the training set. Each seizure of a patient in the training set is left out once and tested for different number of components ranging from 1 to 20. We then compare the predictions obtained by the model for all seizures in the training set with their actual labels. The component number, which gives the best overall classification performance in terms of both sensitivity and specificity, is selected to build the model to be used on the test set.

In addition to the parameter N, there are other parameters to be determined in our analysis. For instance, we set the duration of an epoch to 10 seconds. It has been set to different values in the literature, e.g., 1 second [3], 2 seconds ([1], [2]), 10 seconds [4] and around 20 seconds [5]. Besides, the duration of overlap between consecutive epochs, the maximum number of components in an N-PLS model and the threshold for a VIP score are other user-defined parameters. For each parameter, we use the same value for each patient. In future studies, we also plan to study the sensitivity of the performance of the model on each patient to each one of these parameters.

VI. Results and Discussions

We determine the performance of the model for each patient by computing the average performance over all seizures of the patient. We build a training set using all but one seizure of a patient and use the training set to determine the number of components in N-PLS and also to select a subset of features. We then test the model on the left-out seizure of that particular patient.

As a performance evaluation criterion, we use the geometric mean of sensitivity and specificity, which is called g-means. G-means is defined as $g = \sqrt{sensitivity \times specificity}$ [30]. Sensitivity indicates the proportion of the true-positives to the sum of true-positives and false-negatives, where true-positives are the time epochs that belong to the seizure period and are classified as seizure; false-negatives are the seizure epochs that are classified as non-seizure. Specificity, on the other hand, is the ratio of true-negatives to the sum of true-negatives and false-positives, where true-negatives are the time epochs that belong to non-seizure period and are classified as non-seizure; false-positives are the non-seizure epochs classified as seizure.

Table II demonstrates the performance of the model on eight patients. We show the average g-means for each patient both with feature selection and without feature selection. We observe that feature selection is especially useful for Patient 4, 5 and 6 to detect seizures. For instance in Patient 5, who has three seizures, first two seizures are not detected at all without feature selection and this results in very poor performance. On the other hand, when we select a subset of features based on the EEG signals of the patient in the training set, we refine the model and detect all seizures of the patient with average g-means around 83%. Table III shows the subset of features used in seizure-recognition for each patient. Since we form training sets by leaving-out one seizure at a time, different features can be selected from each training set. The features given in Table III correspond to the union of subsets of features selected from each training set. These subsets of features can be further used to understand the differences between patients. For instance, different seizure locations may result in differences in the features used for seizure recognition. Nevertheless, we should point out that feature selection may also result in overfitting the seizures in the training sets. Therefore, in the cases where there is variation among seizures of a patient, feature selection may degrade the performance.

We also assess the performance of the multi-modal data construction and modeling approach by comparing its performance with that of a two-way classification model. We unfold the Epilepsy Feature Tensor in the time epochs mode as shown in Figure 2 and then use SVM [31] to classify epochs as seizure and non-seizure. Similarly, [1] has previously proposed a patient-specific seizure detection model by representing each time epoch with a feature vector and then classifying the time epochs using SVMs. When we unfold the Epilepsy Feature Tensor in the time epochs mode, we have $7 \times 18 = 126$ features corresponding to each time epoch. We employ SVM[2] to classify the time epochs based on those 126 features. For each patient, we build a patient-specific model using all but one seizure of a patient and then test the model on the left-out seizure and recordings before and after that particular seizure. After each seizure is left-out once, we compute the average performance of the model for each patient. We use radial basis function kernel with a parameter adjusted for each patient. The parameter for each patient is determined using cross-validation on the training set in the same way the number of components for an N-PLS model is determined. Table II demonstrates the g-means corresponding to each patient obtained using a two-way approach. We observe that while SVM has a fairly good performance in terms of seizure detection, for the cases when it performs poorly, our multi-modal approach using feature selection improves the performance of the model. For example, in Patient 5, two-way analysis approach cannot detect one of the seizures at all and this results in low average g-means while NPLS+LDA with feature selection can capture all seizures. By preserving the multi-modality of the data, multiway data analysis keeps the model simple and makes the interpretation easier so that we can easily select features, which in turn would improve the performance resulting in some cases in much better performance than SVM.

VII. Conclusion

We have introduced a multi-modal data construction and analysis approach for patient-specific seizure detection using multi-channel scalp EEG signals. Multi-modality of the data enables us to represent EEG signals from multiple channels using various features from different domains as a third-order tensor called Epilepsy Feature Tensor with modes: time epochs, features and channels. We analyze these multiway arrays using a multi-linear discriminant analysis based on N-PLS in order to classify time epochs as seizure or non-seizure. We combine this multi-modal approach with a variable selection method to identify a subset of features with discriminative power in terms of seizure detection for each patient. Our results demonstrate that multiway data analysis can detect patient-specific seizures with high performance and improve our understanding of different seizure structures by giving us the chance to compare seizures of patients through the features used in seizure detection.

In this study, we have tried to extract various features that can differentiate between seizure and non-seizure periods. While

[2]Implementation of support vector machines called SVM^{light} [32] is used in the analysis.

TABLE II

SEIZURE VS. NON-SEIZURE. PERFORMANCE OF THREE-WAY (NPLS-BASED) AND TWO-WAY (SVM-BASED) APPROACHES IN TERMS OF GEOMETRIC MEANS OF SENSITIVITY AND SPECIFICITY OF THE MODEL. THE ROW CORRESPONDING TO NPLS + LDA SHOWS THE RESULTS WITHOUT FEATURE SELECTION WHILE THE ROW CORRESPONDING TO NPLS + LDA (FS) DEMONSTRATES THE RESULTS OF THE MODEL WITH FEATURE SELECTION.

Seizure vs. Non-seizure	Patient1	Patient2	Patient3	Patient4	Patient5	Patient6	Patient7	Patient8	MEAN
NPLS + LDA	85.3%	97.6%	91.3%	75.0%	28.6%	72.3%	97.0%	86.0%	79.1%
NPLS + LDA (FS)	86.6%	96.7%	91.1%	77.3%	83.1%	89.3%	92.1%	78.4%	86.8%
SVM	86.9%	98.6%	98.4%	76.3%	44.8%	88.3%	98.2%	94.0%	85.7%

TABLE III

SUBSETS OF FEATURES USED IN THE PATIENT-SPECIFIC SEIZURE RECOGNITION MODEL OF EACH PATIENT. PATIENT 1, 2, 7 AND 8 HAVE RIGHT TEMPORAL SEIZURES. PATIENT 3 AND 4 SUFFER FROM LEFT FRONTAL AND LEFT TEMPORAL SEIZURES, RESPECTIVELY. PATIENT 5 IS BILATERAL CENTRAL FRONTAL AND PATIENT 6 IS BILATERAL OCCIPITAL. WHILE SUBSETS OF FEATURES TEND TO BE SIMILAR BASED ON SEIZURE ORIGINS, IT IS NOT POSSIBLE TO MAKE GENERALIZATIONS ON A SMALL SET OF PATIENTS.

	Activity	Mobility	Complexity	Mean Abs. Slope	Spatial Info.	Median Freq.	Spectral Entropy
Patient1	✓	✓	✓	✓	✓	×	✓
Patient2	✓	✓	✓	✓	✓	×	✓
Patient3	✓	✓	✓	×	✓	×	✓
Patient4	✓	✓	✓	×	✓	×	✓
Patient5	✓	✓	✓	×	×	×	✓
Patient6	✓	✓	✓	×	✓	×	×
Patient7	✓	✓	✓	✓	✓	×	✓
Patient8	✓	✓	✓	✓	✓	×	×

TABLE IV

PRE-SEIZURE VS. POST-SEIZURE (BINARY CLASSIFICATION WITHIN NON-SEIZURE EPOCHS). EACH ENTRY SHOWS THE PERFORMANCE OF THE MODEL WHEN IT IS TRAINED ON NON-SEIZURE EPOCHS BEFORE/AFTER SOME SEIZURES OF A PATIENT AND TESTED ON NON-SEIZURE EPOCHS BEFORE/AFTER ANOTHER SEIZURE OF THAT PARTICULAR PATIENT.

Pre-seizure vs. Post-seizure	Patient1	Patient2	Patient3	Patient4	Patient5	Patient6	Patient7	Patient8
NPLS + LDA	93.1%	98.0%	94.2%	88.4%	8.1%	64.7%	85.9%	85.3%

these features can reflect the differences between seizure and non-seizure dynamics to a certain extent, we also explore whether these features can capture the differences between pre-seizure and post-seizure periods. Table IV shows that if we only analyze the data from pre-seizure and post-seizure periods, we can classify epochs into pre-seizure and post-seizure classes with very high performance for most of the patients. These results suggest that we indeed have a multi-class classification problem at hand or we should extract features such that they will be different only in seizure period in order to improve the performance of the model.

Throughout the paper, we have mainly focused on classification of time epochs and selection of features. On the other hand, we have one more mode that we can consider: channels mode. The components in the channels mode can further be explored to see whether seizure localization can be achieved. Besides, in this study we have addressed only patient-specific seizure detection. On the other hand, patient non-specific seizure detection is the ideal seizure detection approach since it would be much more efficient to build a model using the previously recorded seizures of other patients and use that model to detect seizures of new patients. However, patient non-specific seizure recognition is quite challenging considering that patients suffer from seizures with different morphologic and topographic characteristics and training on one type and testing on another may not perform well if the right features are not identified.

REFERENCES

[1] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. T. Treves, and J. Guttag, "Patient-specific seizure onset detection," in *Proc. of the 26th Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, vol. 1, 2004, pp. 419–422.

[2] M. E. Saab and J. Gotman, "A system to detect the onset of epileptic seizures in scalp eeg," *Clinical Neurophysiology*, vol. 116, no. 2, pp. 427–442, 2005.

[3] N. Paivinen, S. Lammi, A. Pitkanen, J. Nissinen, M. Penttonen, and T. Gronfors, "Epileptic seizure detection: A nonlinear viewpoint," *Computer Methods and Programs in Biomedicine*, vol. 79, no. 2, pp. 151–159, 2005.

[4] M. van Putten, T. Kind, F. Visser, and V. Lagerburg, "Detecting temporal lobe seizures from scalp eeg recordings: A comparison of various features," *Clinical Neurophysiology*, vol. 116, no. 10, pp. 2480–2489, 2005.

[5] F. Mormanna, T. Kreuza, C. Riekea, R. G. Andrzejaka, A. Kraskovc, P. Davidb, C. E. Elgera, and K. Lehnertz, "On the predictability of epileptic seizures," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 569–587, 2005.

[6] H. R. Mohseni, A. Maghsoudi, and M. B. Shamsollahi, "Seizure detection in eeg signals: A comparison of different approaches," in *Proc. of the 28th Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, vol. Supplement, 2006, pp. 6724–6727.

[7] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *J. Machine Learning Research*, vol. 3, no. 7-8, pp. 1157–1182, 2003.

[8] W. Chaovalitwongse, P. M. Pardalos, and O. A. Prokopyev, "Electroencephalogram (eeg) time series classification: Applications in epilepsy," *Annals of Operations Research*, vol. 148, no. 1, pp. 227–250, 2006.

[9] F. Estienne, N. Matthijs, D. L. Massart, P. Ricoux, and D. Leibovici, "Multi-way modelling of high-dimensionality electroencephalographic

data," *Chemometrics Intell. Lab. Systems*, vol. 58, no. 1, pp. 59–72, 2001.

[10] F. Miwakeichi, E. Martnez-Montes, P. Valds-Sosa, N. Nishiyama, H. Mizuhara, and Y. Yamaguchi, "Decomposing eeg data into space-time-frequency components using parallel factor analysis," *NeuroImage*, vol. 22, no. 3, pp. 1035–1045, 2004.

[11] R. A. Harshman, "Foundations of the parafac procedure: Models and conditions for an explanatory multi-modal factor analysis," *UCLA working papers in phonetics*, vol. 16, pp. 1–84, 1970.

[12] M. Mørup, L. K. Hansen, C. S. Hermann, J. Parnas, and S. M. Arnfred, "Parallel factor analysis as an exploratory tool for wavelet transformed event-related EEG," *NeuroImage*, vol. 29, no. 3, pp. 938–947, 2006.

[13] E. Acar, C. A. Bingol, H. Bingol, and B. Yener, "Computational analysis of epileptic focus localization," in *Proc. of The Fourth IASTED Int. Conf. Biomedical Engineering*, 2006, pp. 317–322.

[14] E. Acar, C. A. Bingol, H. Bingol, R. Bro, and B. Yener, "Multiway analysis of epilepsy tensors," *Bioinformatics*, vol. 23, no. 13, pp. i10–i18, 2007.

[15] M. de Vos, A. Vergult, L. D. Lathauwer, W. D. Clercq, S. V. Huffel, P. Dupont, A. Palmini, and W. V. Paesschen, "Canonical decomposition of ictal scalp eeg reliably detects the seizure onset zone," *Neuroimage*, vol. 37, no. 3, pp. 844–854, 2007.

[16] E. Martinez-Montes, P. A. Valdes-Sosa, F. Miwakeichi, R. I. Goldman, and M. S. Cohen, "Concurrent eeg/fmri analysis by multiway partial least squares," *Neuroimage*, vol. 22, no. 3, pp. 1023–1034, 2004.

[17] E. Acar, C. A. Bingol, H. Bingol, R. Bro, and B. Yener, "Seizure recognition on epilepsy feature tensor," in *Proc. of the 29th Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, vol. 1, 2007, pp. 4273–4276.

[18] R. Bro, "Multi-way analysis in the food industry: models, algorithms, and applications," Ph.D. dissertation, University of Amsterdam, Holland, 1998.

[19] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemometrics*, vol. 14, no. 3, pp. 105–122, 2000.

[20] R. Bro, "Multiway calibration. multilinear pls," *J. Chemometrics*, vol. 10, no. 1, pp. 47–61, 1996.

[21] R. Bro, A. K. Smilde, and S. D. Jong, "On the difference between low-rank and subspace approximation improved model for multi-linear pls regression," *Chemometrics Intell. Lab. Systems*, vol. 58, no. 1, pp. 3–13, 2001.

[22] A. K. Smilde, R. Bro, and P. Geladi, *Multi-way Analysis. Applications in the chemical sciences*. England: Wiley, 2004.

[23] W. W. S. Wei, *Time Series Analysis: Univariate and Multivariate Methods (2nd edn)*. USA: Addison Wesley, 2006.

[24] K. Schindler, H. Leung, C. E. Elger, and K. Lehnertz, "Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial eeg," *Brain*, vol. 130, no. 1, pp. 65–77, 2007.

[25] O. A. Rosso, M. T. Martin, and A. Plastino, "Evidence of self-organization in brain electrical activity using wavelet-based information tools," *Physica A*, vol. 347, pp. 444–464, 2005.

[26] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 1948.

[27] R. Bro and A. K. Smilde, "Centering and scaling in component analysis," *J. Chemometrics*, vol. 17, no. 1, pp. 16–33, 2003.

[28] L. Nørgaard, R. Bro, and S. B. Engelsen, "A modification of canonical variates analysis to handle highly collinear multivariate data," *J. Chemometrics*, vol. 20, no. 8-10, pp. 425–435, 2006.

[29] I. G. Chong and C. H. Jun, "Performance of some variable selection methods when multicollinearity is present," *Chemometrics Intell. Lab. Systems*, vol. 78, no. 1-2, pp. 103–112, 2005.

[30] M. Kubat and S. Matwin, "Addressing the curse of imbalanced training datasets: One sided selection," in *Proc. of Int. Conf. Machine Learning*, vol. 30, no. 2-3, 1997.

[31] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Berlin: Springer, 1995.

[32] T. Joachims, *Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning*, B. Schlkopf, C. Burges, and A. Smola, Eds. MIT-Press, 1999.