Seizure Recognition on Epilepsy Feature Tensor

Evrim Acar*, Canan Aykut Bingol†, Haluk Bingol‡, Rasmus Bro§ and Bülent Yener*
*Department of Computer Science, Rensselaer Polytechnic Institute, NY, USA
†Department of Neurology, Yeditepe University, Istanbul, Turkey
‡Department of Computer Engineering, Bogazici University, Istanbul, Turkey
§Faculty of Life Sciences, Copenhagen University, Copenhagen, Denmark

***Abstract*—With a goal of automating visual analysis of electroencephalogram (EEG) data and assessing the performance of various features in seizure recognition, we introduce a mathematical model capable of recognizing patient-specific epileptic seizures with high accuracy. We represent multi-channel EEG signals (recorded extracranially) using a set of features. These features expected to have distinct trends during seizure and non-seizure periods include features from both time and frequency domains. The contributions of this paper are threefold. First, we rearrange multi-channel EEG signals as a third-order tensor called an Epilepsy Feature Tensor with modes: time epochs, features and electrodes. Second, we model the Epilepsy Feature Tensor using a multilinear regression model, i.e., Multilinear Partial Least Squares regression, which is the generalization of Partial Least Squares (PLS) regression to higher-order datasets. This two-step approach facilitates EEG data analysis from multiple electrodes represented by several features from different domains. Third, we identify which features (in our feature set) are important for seizure recognition.**

Our results based on the analysis of 19 seizures from 5 epileptic patients demonstrate that multiway analysis of an Epilepsy Feature Tensor can detect (patient-specific) seizures with classification accuracy ranging between 77-96%.

I. Introduction

The identification of an epileptic seizure period based on visual analysis of multi-channel EEG signals is a time-consuming and subjective task. Automation of seizure recognition would, to a certain extent, remove the subjectivity introduced by visual analysis, which is often susceptible to poor judgments due to fatigue, etc. Extensive research has been dedicated to epileptic EEG analysis from diverse disciplines. In all these studies, the seizure onset and duration of a seizure are marked by neurologists. Thus, an automated seizure recognition method would not only remove the subjectivity in identifying a seizure period but also provide an objective and common basis for further research in this field.

A great deal of effort has also 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, etc.) as well as other features from time (e.g., higher-order statistics of the signal in time domain, Hjorth parameters, etc.) and frequency domain (e.g., spectral skewness, spectral entropy, etc.). An almost complete list of the features used in characterization of epileptic seizure dynamics can be found in recent studies ([1], [2]).

The procedure for feature extraction from multi-channel EEG data is often as follows: First, the EEG signal recorded at an electrode is divided into m time epochs (overlapping or non-overlapping). Then n features are extracted from each epoch. Consequently, a signal from a single channel can be represented as a matrix of size $m \times n$. In the literature, studies often make assumptions prior to the analysis and focus on the signal from a certain electrode by relying on the knowledge of seizure localization. For instance, in [1], seizure dynamics are analyzed solely on a specific electrode, which has the characteristics of an epileptogenic focus (seizure origin). Another approach commonly employed in literature is to analyze one feature at a time, as in [2] and [3], rather than analyzing the combined effect of several features. However, it is extremely important to realize that while a single feature may not be very affective in discriminating between seizure and non-seizure periods, linear or nonlinear combinations of several features may well be. In addition to the possibility of enhanced performance, simultaneous analysis of features also enables the performance comparison of features on the same data using the same classifier. Therefore, in this study, we introduce a new approach, which analyzes EEG data from all channels by characterizing the signals using a number of features. We propose to rearrange signals from p channels as a third-order tensor of size $m \times n \times p$ as shown in Figure 1 and model this tensor using multilinear regression analysis. To our knowledge, this is the first study on simultaneous analysis of EEG data from multiple electrodes based on several features from different domains.

In this study, we are particularly interested in distinguishing a seizure (ictal) period from a pre-seizure (preictal) and a post-seizure (post-ictal) period, which ends right before and starts right after a seizure period, respectively. Moreover, our goal is to identify how much each feature contributes to seizure recognition. This study, therefore, differs from the related work on seizure detection and prediction, e.g., [2], [4], [5]. They either focus on the identification of features distinguishing between interictal and preictal periods or aim to detect an epileptic seizure with minimum possible delay using features from a particular domain. Nevertheless, the multiway data construction and analysis approach introduced in this paper can easily be extended to seizure prediction and detection. How accurate such an approach would be is the focus of future studies.

Multilinear models have previously been employed in several applications in neuroscience. In [6], EEG data and data collected through experiments with different doses of a drug are arranged as a six-way array with modes: EEG, patients,

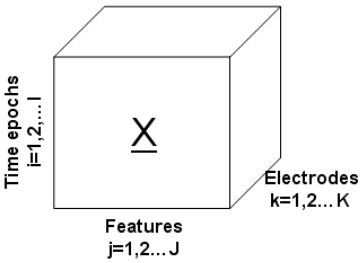

Fig. 1. Epilepsy Feature Tensor. $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ represents the multi-channel EEG data, which are transformed into the feature space by computing certain measures characterizing seizure dynamics. Each entry of $\underline{\mathbf{X}}$, x_{ijk}, corresponds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

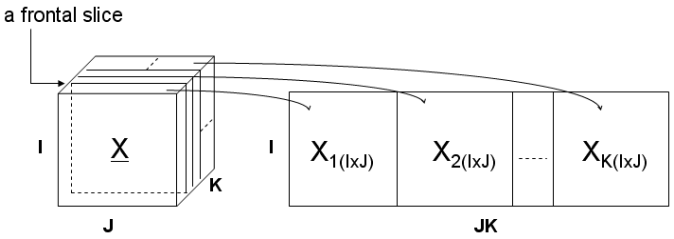

Fig. 2. Matricization of a three-way array in the first mode. A three-way array $\underline{\mathbf{X}} \in \mathbb{R}^{I \times J \times K}$ is unfolded in the first mode and a matrix of size $I \times JK$, denoted by $\mathbf{X}_{(1)}$ is formed. Subscript in $\mathbf{X}_{(i)}$ indicates the mode of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

doses, conditions, etc. The analysis of the six-way dataset demonstrates that significant information is successfully extracted from a complex drug dataset by a multilinear 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-frequency components [7]. The three-way array constructed from multi-channel EEG data in [7] with modes $time\ samples \times frequency \times electrodes$ is analyzed using a multilinear component model called Parallel Factor Analysis model (PARAFAC). Components extracted by PARAFAC 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 [8]. These studies have also motivated the application of multiway models for understanding the structure of epileptic seizures ([9], [10]). Similar to the three-way array constructed in [7], multi-channel ictal EEG data are arranged as a third-order tensor with modes $time\ samples \times frequency \times electrodes$ in [9]. Components extracted by multiway models, i.e., Tucker3 and PARAFAC, are used to explore the signatures of a seizure in 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 multilinear subspace analysis [10]. In addition to the applications of multilinear component models, multilinear regression models have also been previously employed in neuroscience, e.g., in [11] 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 as an alternative or a replacement for visual analysis of EEG data. We introduce a novel approach, which combines the recognition power of several features from different domains and classifies epochs of signals from multiple electrodes as seizure or non-seizure periods. Our contributions in this paper are as follows:

1) We rearrange multi-channel EEG data as a third order tensor, Epilepsy Feature Tensor, with modes: $time\ epochs \times features \times electrodes$. We extract seven features from both time and frequency domain and represent a signal using a set of feature vectors. We do not make any assumptions about the seizure origin but rather analyze the signals from all electrodes simultaneously.
2) We model Epilepsy Feature Tensors using a multilinear regression model called Multilinear PLS and develop a patient-specific seizure recognition method.
3) We observe that features from both the time and frequency domains contribute to seizure recognition. Among the features analyzed in this study, while median frequency and fractal dimension are the most insignificant features in almost all patients, spectral entropy, spectral skewness, Hjorth's activity, mobility and complexity have been observed to be important (with comparatively high regression coefficients).

The organization of this paper is as follows: In Section 2, we include a brief introduction on higher-order datasets and multilinear regression models. Features used in this study are described concisely in Section 3. Section 4 discusses the results, together with the characteristics of the EEG dataset. We conclude, in Section 5, with future objectives in seizure recognition.

II. Methodology

Regression models, e.g., multiple linear regression, PLS and Principal Component Regression (PCR), etc., 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 is a third-order array (Figure 1). This section briefly introduces higher-order arrays and the regression model, i.e., Multilinear Partial Least Squares (N-PLS), developed for higher-order data analysis.

A. Notation and Background

Multiway arrays, often 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.

We denote higher-order arrays using underlined boldface letters, e.g., $\underline{\mathbf{X}}$, following the standard notation in the multiway literature [12]. Matrices and vectors are represented by

Algorithm 1 Multilinear PLS($\underline{\mathbf{X}}$, $\mathbf{y}$, N)

1: $\mathbf{y}_0 = \mathbf{y}$, $\mathbf{X}_0 = \mathbf{X}_{(1)}$
2: **for** $i = 1$ to N **do**
3: $\mathbf{z} = \mathbf{y}_{i-1}^T \mathbf{X}_{i-1}$
Reshape z as a matrix $\mathbf{Z} \in \mathbb{R}^{J \times K}$ such that $\mathbf{Z}(m,n) = \mathbf{z}(m + J * (n-1))$
4: {Compute singular value decomposition of matrix Z}
$\mathbf{Z} = \mathbf{U}\mathbf{S}\mathbf{V}^T$
5: $\mathbf{w}^J = \mathbf{U}(:,1), \quad \mathbf{w}^K = \mathbf{V}(:,1)$
$\mathbf{W}^J(:,i) = \mathbf{w}^J, \mathbf{W}^K(:,i) = \mathbf{w}^K$
6: $\mathbf{T}(:,i) = \mathbf{X}_{i-1}(\mathbf{w}^K \otimes \mathbf{w}^J)$
7: $\mathbf{X}_i = \mathbf{X}_{i-1} - \mathbf{T}(:,i)(\mathbf{w}^K \otimes \mathbf{w}^J)'$
8: $\mathbf{b}_i = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{y}_{i-1} = \mathbf{T}^+\mathbf{y}_{i-1}$
9: {Regression and Deflation}
$\mathbf{y}_i = \mathbf{y}_{i-1} - \mathbf{T}\mathbf{b}_i = (\mathbf{I} - \mathbf{T}\mathbf{T}^+)\mathbf{y}_{i-1}$
10: **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 mode, respectively. $\mathbf{T}^+$ stands for pseudo-inverse defined as $\mathbf{T}^+ = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T$. $\otimes$ indicates Kronecker product [14].

boldface capital, e.g., $\mathbf{X}$, and boldface lowercase letters, e.g., $\mathbf{x}$, respectively. Scalars are denoted by lowercase letters, e.g., x.

In higher-order array terminology, each dimension of a multiway array is called a mode (or a way) and the number of variables in each mode is used to indicate the dimensionality of 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., first mode as shown in Figure 2. This operation is called matricization. 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 low prediction accuracy, information loss and misinterpretation of the results especially if the data are noisy [13]. Therefore, we preserve the multimodality of the dataset and employ a generalized version of a regression model, i.e., PLS, to higher-order arrays.

B. Multilinear Partial Least Squares (N-PLS)

Multilinear PLS is introduced as a generalization of PLS to multiway datasets [15]. 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 class assignments of time epochs (first mode). Multilinear 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 mode, respectively are also formed. The steps of the algorithm are briefly summarized in Algorithm 1 [16].

Once component matrices $\mathbf{T}$, $\mathbf{W}^J$ and $\mathbf{W}^K$ are formed using Algorithm 1 and the model is built on the available data points, the labels for the new samples, $\mathbf{y}_{new}$, can also be predicted. The predictions for new samples, $\underline{\mathbf{X}}_{new}$, are obtained by computing $\mathbf{y}_{new} = \mathbf{T}\mathbf{b}_t = \mathbf{X}_{new(1)}\mathbf{b}_{pls}$. How $\mathbf{b}_{pls}$ and $\mathbf{b}_t$ are related and computed using $\mathbf{W}^J$ and $\mathbf{W}^K$ are given in detail in [14]. Usually, the predictions are based on centered $\underline{\mathbf{X}}$ and $\mathbf{y}$.

This prediction step enables us to build a model on available seizures of a patient and then use the model to predict the labels of epochs in new EEG recordings of a patient as seizure or non-seizure.

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 time 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 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 [1] 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.

2) Fractal Dimension (FD): In order to quantify the signal complexity and self-similarity, we compute FD of each epoch using Higuchi's algorithm [17] briefly described below:

- Given a time series $\mathbf{s}$, k new time series are generated with different initial times ($m = 1, ..k$) denoted by $\mathbf{s}_m^k$, where $\mathbf{s}_m^k = \{s(m), s(m+k), ..., s(m + \lfloor (N-m)/k \rfloor * k)\}$ and N is the total number of samples in series $\mathbf{s}$.

- The length of each time series, $L_m(k)$ is computed: $L_m(k) = \frac{1}{k}\{\frac{N-1}{\lfloor (N-m)/k \rfloor * k} * (\sum_{i=1}^{\lfloor (N-m)/k \rfloor} |s(m+i*k) - s(m+(i-1)*k)|)\}$
- $L(k)$ is calculated by taking the average of $L_m(k)$ over m.
- If $L(k)$ is proportional to k^{-D}, this indicates that the time series is fractal-like with dimension D, called the fractal dimension. The slope of $log(L(k))$ vs. $log(\frac{1}{k})$ for $k = 1, 2, ..k_{max}$[1] is used as an estimator of D.

$$FractalDimension: \quad f_4(s) = D$$

B. Frequency domain

1) Frequency Spectrum: We reduce the time series to a stationary time series by taking the first difference of the signal before computing the amplitude spectrum. Given a time series **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 used to extract the fifth ($f_5(s)$) and sixth ($f_6(s)$) features, which are median frequency and skewness of the amplitude spectrum, respectively.

2) Spectral Entropy: The last feature in this study is a measure of spectral entropy used to quantify the uncertainty in the frequency domain. Five frequency bands in accordance with the traditional EEG frequency bands 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 (CWT) 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[2] and $|c_{ij}|^2 = c_{ij}c_{ij}^*$. N is the length of an epoch and S is the number of scales.

We then compute spectral entropy, H, using Shannon's entropy measure [18] 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.

$$SpectralEntropy: \quad f_7(s) = H$$

The list of these features can easily be extended to a larger set and the approach proposed in this paper will still be valid.

[1]Maximum interval time (k_{max}) is chosen to be 6.

[2]Scale is not the same as frequency but contains frequency information (inversely proportional to frequency).

IV. Results and Discussions

A. Data

Our dataset contains multi-channel EEG recordings of 19 seizures from 5 patients with different pathology substrates: 3 mesial temporal sclerosis (MTS); 1 dysembroyoplastic neuroepithelial tumor (DNET); 1 nonlesional. The EEG data have been collected via scalp electrodes in the epilepsy monitoring unit of Yeditepe University Hospital. The recording of EEG with referential electrode Cz is used for computational analyses. The number of seizures per patient, sampling frequencies, epoch sizes as well as sizes of Epilepsy Feature Tensors are given in Table I. EEG recordings in the dataset are not preprocessed to remove artifacts and we know that eye and muscle artifacts are present based on our previous study on some of the patients used in this paper [10][3]. The only filter applied on the data is the notch filter at 50 Hz to remove the artifact from the power source.

The data corresponding to a seizure of a patient contain a certain amount of data before the seizure, the seizure period and a certain amount of data after the seizure period. Each signal is divided into epochs of 10 sec. (each epoch typically contains 2000 or 4000 samples depending on the sampling frequency.). The epochs are formed using a sliding window approach such that consecutive epochs differ only in 100 samples[4]. Seven features are computed for all epochs and a matrix of size *nb of time epochs* $\times$ 7 is created for the signal from a single electrode. When all electrodes are included in the analysis, this forms a three-way array of *nb of time epochs*$\times 7 \times 18$ for each seizure (Figure 3). Once the tensor is built, we scale the three-way array within the feature mode before the analysis since features have different ranges of magnitudes. Scaling a three-way array within one mode is different than scaling in two-way datasets. Unlike matrices where columns or rows are scaled, in three-way case, whole matrices need to be scaled [19]. For instance, while scaling $\underline{\mathbf{X}}$ (Figure 1) within feature mode, vertical slices are scaled.

The seizure period is marked by two 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., **y**-vector in Algorithm 1, corresponding to the *time epoch* 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 in both training and test sets so that the performance of the model is not affected by epochs containing the characteristics of different seizure dynamics.

[3]In this study, we have only chosen the patients with more than two seizures.

[4]We keep this number the same for all patients regardless of the sampling frequency. We may also keep the overlap duration constant instead (We did try for 50 samples for the patient with 200Hz sampling frequency and the classification accuracy slightly changes, i.e., 83.73%).

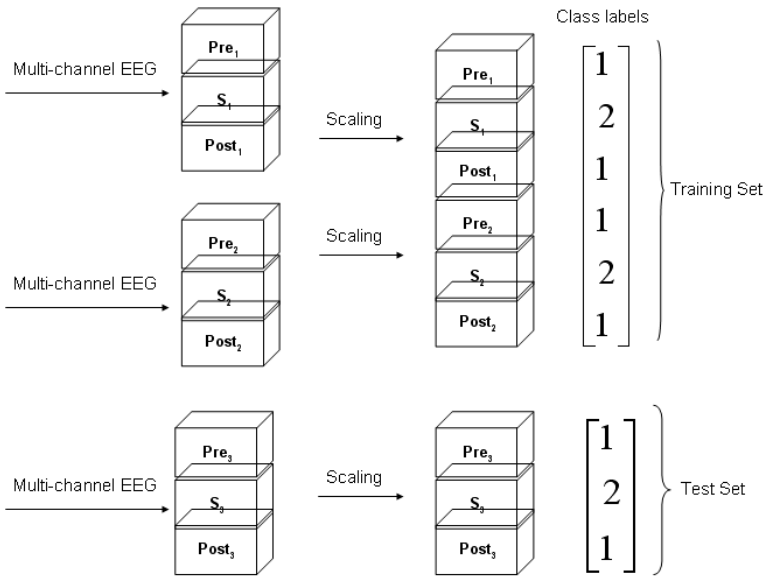

Fig. 3. Construction of training and test sets for a patient with three seizures. S_i indicates the data for the i^{th} seizure of a patient while Pre_i and $Post_i$ indicate the recordings in pre-seizure and post-seizure periods corresponding to the i^{th} seizure.

B. Results

1) Seizure Recognition: In order to assess the performance of the model, we form a training set using all but one seizure of a patient together with the corresponding labels of the epochs in the training set (Before the analysis, both independent and dependent data are centered). We regress the data for all the seizures in the training set onto the y-vector containing 1's and 2's (for non-seizure and seizure, respectively) using Multilinear PLS regression and build a model based on Algorithm 1[5]. The model is then tested on the test dataset, which contains the left-out seizure (Figure 3). Predicted classes for the epochs in the test set are real numbers. A simple approach that rounds the predictions to the nearest integer (1 or 2) is used to determine the class of an epoch. This approach is not the optimal way and it can possibly be improved by a classifier like Linear Discriminant Analysis (LDA), etc.

As seen in Algorithm 1, the number of components, N, is a user-defined parameter. In order to determine N, we use an approach based on cross-validation. Each seizure of a patient is left out once and tested for different number of components ranging from 1 to 10. After all seizures are tested once, we compare the predictions obtained by the model for all seizures with the actual labels. We finally pick the component number, which gives the best overall classification accuracy[6]. The classification accuracy is the percent of the number of epochs correctly assigned to their actual classes. The last column of Table I shows the performance of the model for five patients analyzed in this study and demonstrates that we obtain promising classification accuracies ranging between 77% and 96%. It is also possible to increase N beyond 10 and obtain slightly better classification. However, as the model gets complex, the interpretation of features (discussed in the next section) becomes harder.

In [3], the performance of different approaches in seizure detection has been summarized by presenting the classification accuracies given in the literature for the publicly available EEG dataset described in [21]. We would like to point out that comparison of our results with those would be misleading due to major differences in the type of the data. In this study, we aim to differentiate between non-seizure and seizure phases using multi-channel EEG data recorded *extracranially*. We have also mentioned that non-seizure phases correspond to pre-seizure and post-seizure periods. Therefore, our goal is to mark the seizure period. On the other hand, in previous work ([3] and references therein), even if the problem definition is presented as the differentiation of non-seizure and seizure periods, the concept of non-seizure is defined differently. Epochs that belong to a non-seizure period include seizure-free data from healthy patients recorded extracranially as well as seizure-free data from epilepsy patients recorded *intracranially*. Consequently, in our case, it is more challenging to differentiate a few seconds before and after a seizure period from the seizure compared to differentiating EEG of a healthy patient from the seizure. Besides, we obtain these results without placing electrodes within the scalp but rather use the recordings collected outside the cranial cavity. In addition to these differences, we currently focus on patient-specific seizure recognition and do not model the variation among

[5]Implementation of N-PLS in PLS_Toolbox [20] running under MATLAB is used for the analysis.

[6]It is also possible to fully automate the approach for picking the component number. When a seizure of a patient is left out as a test case, the component number can be determined in the training set using cross validation. The component number which gives the highest accuracy in the training set can be chosen as the component number to be used on the test set. We do not use this approach for the time being since some patients have only 3 seizures.

TABLE I
EEG DATASET

Patient-ID	Seizures-ID	Tensor Size	Sampling Freq.(Hz)	Epoch Size (Samples)	**Classification Accuracy**
1	1	$322 \times 7 \times 18$	200	2000	
1	2	$406 \times 7 \times 18$	200	2000	
1	3	$202 \times 7 \times 18$	200	2000	84.00%
1	4	$202 \times 7 \times 18$	200	2000	
1	5	$262 \times 7 \times 18$	200	2000	
2	1	$938 \times 7 \times 18$	400	4000	
2	2	$934 \times 7 \times 18$	400	4000	
2	3	$946 \times 7 \times 18$	400	4000	95.97%
2	4	$974 \times 7 \times 18$	400	4000	
2	5	$978 \times 7 \times 18$	400	4000	
3	1	$690 \times 7 \times 18$	400	4000	
3	2	$710 \times 7 \times 18$	400	4000	76.94%
3	3	$742 \times 7 \times 18$	400	4000	
4	1	$830 \times 7 \times 18$	400	4000	
4	2	$786 \times 7 \times 18$	400	4000	94.05%
4	3	$922 \times 7 \times 18$	400	4000	
5	1	$1214 \times 7 \times 18$	400	4000	
5	2	$1294 \times 7 \times 18$	400	4000	81.66%
5	3	$1210 \times 7 \times 18$	400	4000	

different patients.

2) Interpretation of Features: In order to understand the contribution of each feature to seizure recognition, we model all the seizures of a patient using N-PLS. We, first, combine all seizures of a patient in a single Epilepsy Feature Tensor and then regress onto the actual labels using the optimal number of components chosen in the previous section. We determine the regression coefficients, i.e., $\mathbf{b}_{pls} \in \mathbb{R}^{JK \times 1}$, which indicate the significance of each variable in the prediction of time epoch classes. There are J features and K electrodes. Consequently, there is a regression coefficient corresponding to each feature recorded at each electrode. We rearrange $\mathbf{b}_{pls}$ as a matrix of electrodes by features and each entry in the matrix represents the regression coefficient corresponding to the feature recorded at a particular electrode. The mean across the electrodes is then used to evaluate the overall significance of a single feature.

When we inspect the mean absolute regression coefficients of the features in Table II, we observe that:

- Some features, in particular spectral entropy, spectral skewness, Hjorth's activity, mobility and complexity, have relatively higher mean absolute regression coefficients. Consequently, they are more significant compared to the remaining features.
- On the other hand, regression coefficients corresponding to fractal dimension and median frequency are lower in magnitude. Therefore, these features have the least contribution in seizure recognition.

In addition to these observations, we detect that Patient2 and Patient4, who have comparatively higher classification accuracies (94.05% and 95.97%, respectively), also have almost the same pattern in terms of regression coefficients of the features. On the other hand, the patterns in other patients are different.

TABLE II
MEAN ABSOLUTE REGRESSION COEFFICIENTS CORRESPONDING TO FEATURES FOR DIFFERENT PATIENTS

Patient-ID	f_1	f_2	f_3	f_4	f_5	f_6	f_7[a]
1	0.69	1.07	0.48	0.10	0.58	0.52	1.55
2	0.78	0.28	0.49	0.15	0.12	0.99	0.96
3	1.83	2.20	2.11	0.25	0.96	1.42	2.73
4	0.83	0.44	0.66	0.10	0.21	0.89	1.06
5	1.25	3.87	1.59	0.40	0.32	1.32	1.73

[a] f_1: Activity, f_2: Mobility, f_3: Complexity, f_4: FD, f_5: Median Freq., f_6: Spectral Skewness, f_7: Spectral Entropy

While spectral entropy, spectral skewness and activity are the most significant features in Patient2 and Patient4, we observe that complexity and mobility are in the top three significant features in other patients. Relatively lower seizure recognition accuracies in some patients may be attributed to artifacts and features reflecting the effects of artifacts or even to subjectivity in visual analysis. We also acknowledge that when compared with the clinical findings, classification accuracies are observed not to be correlated with lateralization or underlying etiology for these five patients. However, we need a larger dataset to generalize these observations.

V. CONCLUSION

We have introduced a multimodal approach with a goal of automatically differentiating a seizure period from pre-seizure and post-seizure periods in multi-channel ictal EEG. The proposed approach enables the analysis and comparison of a multitude of features from different domains. In addition to that, signals from multiple electrodes can be analyzed

simultaneously by constructing a third-order tensor called an Epilepsy Feature Tensor. We model the data using Multilinear PLS regression and develop a mathematical model for patient-specific seizure recognition with promising classification accuracies on five epileptic patients. Our approach characterizes a patient's seizure dynamics with a set of features so it is an initial but important step in terms of understanding the differences in seizures of different patients.

Nevertheless, there are many research directions to explore. First of all, our model focuses on patient-specific seizure recognition. On the other hand, a more generalizable approach, which understands and models the variation among patients is also significant in terms of seizure recognition and patient treatment. Secondly, this study has only focused on capturing the linear relations in the feature set. However, whether modeling nonlinear relations between features (also suggested in [4]) would improve the classification accuracy is an interesting question. Finally, we hope to work on a larger dataset where each patient has many seizures. In our dataset, origins of all seizures for a particular patient are the same. In a larger dataset, with many seizures from a patient, we would like to explore the performance of the model in the cases where some seizures of a patient have different seizure origins.

REFERENCES

[1] 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.

[2] 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.

[3] H. R. Mohseni, A. Maghsoudi, and M. B. Shamsollahi, "Seizure detection in eeg signals: A comparison of different approaches," in *Proc. of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. Supplement, 2006.

[4] A. Shoeb, H. Edwards, J. Connolly, B. Bourgeois, S. T. Treves, and J. Guttag, "Patient-specific seizure onset detection," in *Proc. of the 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, vol. 1, 2004.

[5] W. Chaovalitwongse, P. M. Pardalos, and O. A. Prokopyev, "Electroencephalogram (eeg) time series classification: Applications in epilepsy," *Annals of Operations Research*, vol. 148, pp. 227–250, 2006.

[6] 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.

[7] 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.

[8] 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.

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

[10] E. Acar, C. A. Bingol, H. Bingol, R. Bro, and B. Yener, "Multiway analysis of epilepsy tensors," *Accepted to Bioinformatics*, 2007.

[11] 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.

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

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

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

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

[16] 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, pp. 3–13, 2001.

[17] A. Accardo, M. Affinito, M. Carrozzi, and F. Bouquet, "Use of the fractal dimension for the analysis of electroencephalographic time series," *Biological Cybernetics*, vol. 77, no. 5, pp. 339–350, 1997.

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

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

[20] Eigenvector, "Plstoolbox," in *http://www.eigenvector.com/*, 2007.

[21] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state," *Physical Review E*, vol. 64, no. 6, p. 061907, 2001.