Nonnegative Tensor Decomposition for EEG Epileptic Spike Detection

Nguyen Thi Anh-Dao[1,2], Le Trung Thanh[1], Nguyen Linh-Trung[1], Ha Vu Le[1]

[1] AVITECH, University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam
[2] Electronics and Telecommunications, University of Technology and Logistics, Bac Ninh, Vietnam

***Abstract*—Tensor decomposition can be used for analyzing multi-channel EEG signals in epilepsy diagnosis. We propose a new tensor-based approach to detect epileptic spikes in EEG data. Nonnegative Tucker decomposition was applied to analyze multi-domain features of EEG epileptic and non-epileptic spikes. An EEG feature extraction method was proposed, based on estimating the so-called "eigenspikes." The Fisher score was employed for feature selection. KNN and NB classifiers were used on the extracted features to separate epileptic spikes from non-epileptic spikes, and classification results were compared with those of the Phan-Cichoki method. Experimental results showed that our proposed method is efficient in detecting epileptic spikes.**

***Index Terms*—Nonegative Tucker decomposition, EEG, epileptic spikes, feature extraction, feature selection.**

I. Introduction

Epilepsy is one of the most serious disorders of the brain. Epilepsy affects about 50 million people around the world and is one of the global burden of disease. In developing countries, the number of people with epilepsy accounts for 80% of all epilepsies in the world and $80-90\%$ of them are untreated [1].

Electroencephalogram (EEG) is one of the most accessible tools for epilepsy diagnosis and treatment. EEG signals allow the neurologist to observe epileptic seizures and epileptiforms (spikes, sharp waves and spike-wave complexes), in order to help identify the type of epilepsy and the area of the affected brain. To aid epileptic diagnosis, automatic detection of epileptic spikes may help identify epileptic spikes faster than manual reading by the neurologist.

Over the last 40 years, there have been numerous studies on automatic detection of epileptic seizures and epileptiforms based on EEG analysis. The reader is referred to [2] for a good review on the topic. Several efficient multi-stage detection methods, combining multiple signal processing and machine learning techniques, have been proposed in order to take advantages of each specific technique (see [3] and references therein). One key stage of these methods *feature extraction*, which will be considered in this paper. Most of those works focused on analyzing a single channel of EEG data at a time, not simultaneous processing multiple channels. Indiradevi in [4] proposed a detection system based on multi-level, multi-resolution analysis, which aims to combine EEG information in time, frequency and space together, to automatically identify epileptiform activities.

A natural way to represent multi-dimensional data is to use tensor structure which is a multi-dimensional array. Hence, tensor decomposition becomes a useful tool to analyze high-dimensional data. With the rise of big data applications, tensor decomposition has become an attractive method. It can find many applications in signal processing [5]. Two widely used decompositions for tensors are parallel factor analysis (PARAFAC) (also referred to as canonical polyadic (CP) decomposition) and Tucker decomposition. Tensor decomposition-based methods for EEG signals in general [6] and for seizure in particular [7]–[9] have been studied. Tensor decomposition was also used for separating epileptic generalized spike waves from non-epileptic events in EEG signals [10].

In parallel, nonnegative tensor decomposition (i.e., imposing non-negativity constraints on the decomposition) can provide several interesting properties, including (i) the resulting expressions appear to be purely additive, not subtractive, (ii) the decomposed factors are often "sparse" and, hence, (iii) the decomposition may result in a "parts-based" representation that allows us to learn localized parts of the data [11]. These properties motivate us to look for a new approach for automatic detection of epileptic spike in the EEG using nonnegative tensor decomposition.

II. Nonnegative Tensor Decomposition

Nonnegative tensor decomposition of a k-way tensor is defined as follows:

$$\begin{aligned} &\underset{\boldsymbol{\mathcal{G}},\mathbf{U}_i}{\arg\min} && \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_k \mathbf{U}_k\|_F^2, \\ &\text{s.t.} && \boldsymbol{\mathcal{G}} \geq 0, \mathbf{U}_i \geq 0, i = 1, 2, \dots k, \end{aligned} \tag{1}$$

where $\boldsymbol{\mathcal{X}}$ is the underlying tensor, $\boldsymbol{\mathcal{G}}$ is called core tensor of $\boldsymbol{\mathcal{X}}$ and $\mathbf{U}_i$ are called factors of $\boldsymbol{\mathcal{X}}$, and the notation $\geq$ for a matrix/tensor means all elements of the matrix/tensor are nonnegative. There are two main models: nonnegative canonical polyadic decomposition (NCP) and nonnegative Tucker decomposition (NTD). NCP corresponds to the case when $\boldsymbol{\mathcal{G}}$ is a diagonal tensor, and NTD is when $\boldsymbol{\mathcal{G}}$ is sparse and $\mathbf{U}_i$ are orthogonal matrices.

In this paper, we focus on NTD. In NTD, both the core tensor $\boldsymbol{\mathcal{G}}$ and the factors $\mathbf{U}_i$ are required to be nonnegative. Almost algorithms for NTD are inspired from nonnegative matrix factorization, in which a factor is optimized while the others are fixed. As a result, one can use another objective function for NTD, alternative to that in (1), as

$$F_{\text{NTD}} = \frac{1}{2} \sum_{i=1}^{k} \|\boldsymbol{\mathcal{X}}_{(i)} - \mathbf{U}_i \mathbf{S}_i\|^2, \quad \text{s.t. } \mathbf{U}_i \geq 0, \tag{2}$$

or its vectorization version as

$$F_{\text{NTD}} = \frac{1}{2} \|\text{vec}(\boldsymbol{\mathcal{X}}) - \mathbf{F}\text{vec}(\boldsymbol{\mathcal{G}})\|^2, \quad \text{s.t. } \boldsymbol{\mathcal{G}} \geq 0, \tag{3}$$

with $\mathbf{F} = \otimes \mathbf{U}_n$. In order to minimize the objective function, we can take gradients of F_{NTD} with respect to $\mathbf{U}_n$ and $\boldsymbol{\mathcal{G}}$ to obtain the following update rules:

$$\mathbf{U}_n = \mathbf{U}_n - \alpha \circledast \frac{\partial F_{\text{NTD}}}{\partial \mathbf{U}_n},$$
$$\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{G}} - \alpha \circledast \frac{\partial F_{\text{NTD}}}{\partial \text{vec}(\boldsymbol{\mathcal{G}})},$$

where the step size α can be determined by $\alpha = \mathbf{U}_n \oslash (\mathbf{U}_n \boldsymbol{\mathcal{X}}_{(n)} \boldsymbol{\mathcal{G}}_{(n)}^T)$; the notations $\circledast$ and $\oslash$ are the Hadamand product and the division of two matrices, respectively. More details about NTD algorithms can be found in [12].

978-1-5386-7983-8/18/$31.00 ©2018 IEEE

III. Proposed Methods

A. Feature Extraction

Given N_1 three-way tensors $\boldsymbol{\mathcal{X}}_i^{\text{ep}} \in \mathbb{R}_+^{I\times J\times K}|_{i=1,\dots,N_1}$ (I, J and K correspond to time, wavelet-scale and channel, respectively) representing the EEG segment containg epileptic spikes, and N_2 three-way tensors $\boldsymbol{\mathcal{X}}_j^{\text{nep}} \in \mathbb{R}_+^{I\times J\times K}|_{j=1,\dots,N_2}$ representing the EEG segment containing non-epileptic spikes. To extract features, a feature space, $\mathcal{F}_{\text{ep}}$, which spans the EEG epileptic spikes needs to be estimated, then both types of spikes are projected onto that space.

Our objective function can be expressed as

$$G = \sum_{i=1}^{N_1} \|\boldsymbol{\mathcal{X}}_i^{\text{ep}} - \boldsymbol{\mathcal{X}}_i^{\text{ep}} \times_1 \mathbf{A}\mathbf{A}^{\#} \times_2 \mathbf{B}\mathbf{B}^{\#} \times_3 \mathbf{C}\mathbf{C}^{\#}\|^2 \qquad (4)$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}$ are nonnegative factors.

In a method proposed by Phan and Cichocki in [13], the *complete* set of training tensors was used to form a single four-way tensor by concatenation, on which NTD was performed. Since epileptic spikes are abnormal activity, they can be assumed to be independent from the other activities. Thus, the feature space of epileptic spikes assumingly does not contain other activities. Therefore, we are interested in a feature space that spans only epileptic spikes. For that reason, in our proposed method, we concatenate only three-way epileptic tensors $\boldsymbol{\mathcal{X}}_i^{\text{ep}}$ into a single four-way tensor $\widetilde{\boldsymbol{\mathcal{X}}}^{\text{ep}} \in \mathbb{R}_+^{I\times J\times K\times N_1}$, and then perform NTD, as shown below, to obtain factors $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$:

$$\widetilde{\boldsymbol{\mathcal{X}}}^{\text{ep}} = \left[\boldsymbol{\mathcal{X}}_1^{\text{ep}}|\boldsymbol{\mathcal{X}}_2^{\text{ep}}|\dots|\boldsymbol{\mathcal{X}}_{N_1}^{\text{ep}}\right] = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \times_4 \mathbf{D}, \qquad (5)$$

where $\mathbf{A} \in \mathbb{R}_+^{I\times r_1}$, $\mathbf{B} \in \mathbb{R}_+^{J\times r_2}$, $\mathbf{C} \in \mathbb{R}_+^{K\times r_3}$ and $\mathbf{D} \in \mathbb{R}_+^{N_1\times N_1}$ span the parameter spaces respectively representing the time, wavelet-scale, channel, and epileptic spikes. The columns of $\mathbf{D}$ are considered as "eigenspikes," and its span forms the feature space $\mathcal{F}^{\text{ep}}$ of epileptic spikes.

The n-mode of any three-way EEG tensor $\boldsymbol{\mathcal{X}}$ can be expressed by a linear combination of eigenspikes as follows:

$$\underbrace{\boldsymbol{\mathcal{X}}_{(n)}}_{\text{input data}} = \underbrace{\mathbf{D}}_{\text{basic vectors}} \underbrace{\boldsymbol{\mathcal{G}}_{(n)}(\mathbf{C}\otimes\mathbf{B}\otimes\mathbf{A})^T}_{\text{coefficients}}. \qquad (6)$$

The feature space, $\mathcal{F}^{\text{ep}}$, is then obtained by a multiplication of the core tensor $\boldsymbol{\mathcal{G}}$ and the factor $\mathbf{D}$ of $\boldsymbol{\mathcal{X}}$, that is,

$$\boldsymbol{\mathcal{F}}^{ep} = \boldsymbol{\mathcal{G}} \times_4 \mathbf{D}. \qquad (7)$$

In order to investigate EEG epileptic spike features in a specific feature domain, such as time, wavelet-scale, or channel, we use a sub-space of the feature space corresponding to the parameter. For example, the eigenspike time-basis $\boldsymbol{\mathcal{F}}_{\text{time}}^{ep}$ is obtained by multiplying $\mathbf{A}$ with $\boldsymbol{\mathcal{F}}^{ep}$, i.e., $\boldsymbol{\mathcal{F}}_{\text{time}}^{ep} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A} \times_4 \mathbf{D}$, to yield the principal axes of variations of an epileptic spike across channel and wavelet-scale modes. Similarly, the scale-basis and channel-basis of the eigenspike space can be obtained, i.e, $\boldsymbol{\mathcal{F}}_{\text{scale}}^{ep} = \boldsymbol{\mathcal{G}} \times_2 \mathbf{B} \times_4 \mathbf{D}$ and $\boldsymbol{\mathcal{F}}_{\text{channel}}^{ep} = \boldsymbol{\mathcal{G}} \times_3 \mathbf{C} \times_4 \mathbf{D}$.

Given a training set of M EEG tensors (epileptic and non-epileptic) $\boldsymbol{\mathcal{X}}_m^{\text{train}}$, $m = 1, \dots, M$, the feature vector $\mathbf{f}_m^{\text{train}}$ is established as follows:

$$\begin{aligned} \boldsymbol{\mathcal{F}}_m^{\text{train}} &= \boldsymbol{\mathcal{G}}_m^{\text{train}} \times_4 \mathbf{D} = \boldsymbol{\mathcal{X}}_m^{\text{train}} \times_1 \mathbf{A}^{\#} \times_2 \mathbf{B}^{\#} \times_3 \mathbf{C}^{\#}, \\ \mathbf{f}_m^{\text{train}} &= \text{vec}(\boldsymbol{\mathcal{F}}_m^{\text{train}}). \end{aligned} \qquad (8)$$

Similarly, for any EEG tensor $\boldsymbol{\mathcal{X}}^{\text{test}}$ in the test set, its feature vector can be obtained by projecting the tensor onto $\mathcal{F}^{\text{ep}}$, as follows:

$$\begin{aligned} \boldsymbol{\mathcal{F}}^{\text{test}} &= \boldsymbol{\mathcal{X}}^{\text{test}} \times_1 \mathbf{A}^{\#} \times_2 \mathbf{B}^{\#} \times_3 \mathbf{C}^{\#}, \\ \mathbf{f}^{\text{test}} &= \text{vec}(\boldsymbol{\mathcal{F}}^{\text{test}}). \end{aligned} \qquad (9)$$

B. Feature Selection

Feature selection plays an important role in mining high-dimensional data in general and EEG data in particular. The key idea of feature selection is to find a subset of input features, such that it can span the data space. In this paper, we use Fisher score, which is one of the widely used methods for feature selection [14], for each feature to measure the effectiveness of the classification.

Given n features $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n\}$, extracted by NTD, the idea is to find a linear combination $\mathbf{w}^T\mathbf{F}$ such that the best separation between two classes can be achieved. In particular, the Fisher discriminant ratio is determined by maximizing the following ratio of between-class variation and within-class variation:

$$J(\mathbf{w}) = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}} = \frac{[\mathbf{w}(m_1 - m_2)]^2}{\mathbf{w}^T(\Sigma_1 + \Sigma_2)\mathbf{w}}. \qquad (10)$$

The Fisher score of each feature $\mathbf{f}_i$ can then be defined as the maximum separation $\mathbf{w}(i)$, that is,

$$\gamma(\mathbf{f}_i) \triangleq \mathbf{w}(i) = \frac{N_1(\mu_{i,1} - \mu_i)^2 + N_2(\mu_{i,2} - \mu_i)^2}{N_1\sigma_{i,1}^2 + N_2\sigma_{i,2}^2}. \qquad (11)$$

in which, $\mu_{i,c}$ and $\sigma_{i,c}$ are the mean and standard deviation of the i-th feature for class C_c, $c \in \{1, 2\}$, respectively, μ_i and σ_i are the mean and the standard deviation of the i-th feature in the whole training dataset, and (m_1, Σ_1) and (m_2, Σ_2) are the means and covariance matrices of two classes.

We only need to select l significant features, $l \leq n$, that can sufficiently approximate the space of the data

$$\mathbf{F}_i = \{\mathbf{f}_{(1)}, \mathbf{f}_{(2)}, \dots \mathbf{f}_{(l)} | \mathbf{f}_{(i)} \in \mathbf{F}, i = 1, 2, \dots l\}.$$

We can use l top Fisher scores to determine these significant features.

C. Number of Components

Determining the rank of a tensor is an NP-hard problem. That means, we may only need to seek for a "good" tensor approximation, and can do so by determining the number of "good" components of each factor (i.e., number of columns in the factor). Readers are reffered to [5] for such methods, as DIFFIT, CORCONDIA and ARD.

Truncated higher-order singular value decomposition (HOSVD) is utilized in determining the number of components when decomposing an EEG tensor, to provide a best-rank tensor approximation [15]. The number of components, r_i, of the factor $\mathbf{U}_i$ can be estimated from mode $\boldsymbol{\mathcal{X}}_{(i)}$ as follows:

$$\boldsymbol{\mathcal{X}}_{(i)} \approx \mathbf{U}_i^{I\times r_i} \mathbf{\Lambda}_1^{r_i\times r_i} \mathbf{V}_1^{r_i\times JK} \qquad (12)$$

For a three-way tensor, $i = 1, 2, 3$. For example, the number of components, r_1, of the factor $\mathbf{U}_1$ in mode $\boldsymbol{\mathcal{X}}_{(1)}$ can be determined as the number of the top eigenvalues of the covariance matrix of the mode such that the eigengap or the total variance is maximized, i.e.,

$$\begin{aligned} \text{Eigengap}_{r_1} &= |\lambda_{r_1} - \lambda_{r_1+1}|, \\ \text{VAR}_{r_1} &= \frac{\sum_{i=1}^{r_1} \lambda_i}{\sum_{j=1}^{I} \lambda_j} 100\%. \end{aligned}$$

IV. Results and Discussions

A. EEG Dataset and EEG tensors

For experiments, we use an EEG dataset collected on 16 epilepsy patients, recorded by using a standard 10-20 system with 19 channels and the sampling rate of 256 Hz. Durations of the measurements vary from 6 to 28 minutes (Table I).

For pre-processing, we use a digital Butterworth low-pass filter with the cutoff frequency of 70 Hz; a notch filter with the cutoff

TABLE I: EEG Dataset

Pat.	Gen.	Age	Duration	Spikes	Pat.	Gen.	Age	Duration	Spikes
1	Male	21	23m57s	8	10	Female	13	18m53s	5
2	Male	6	22m25s	635	11	Female	20	14m32s	324
3	Male	9	11m24s	6	12	Female	16	20m14s	8
4	Male	4	19m21s	8	13	Female	28	5m31s	12
5	Male	15	22m0s	2	14	Male	9	11m24s	16
6	Male	12	17m49s	22	15	Male	11	16m16s	351
7	Male	72	15m26s	2	16a	Female	22	17m 56s	19
8	Male	16	22m58s	11	16b	Female	22	9m 41s	9
9	Male	20	27m13s	1					

Pat. = Patient, Gen = Gender

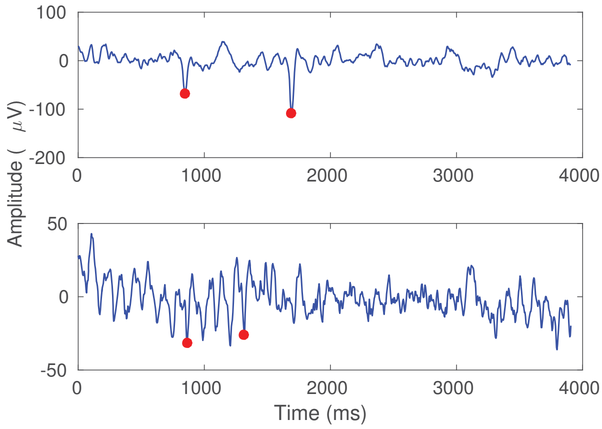


Fig. 1: Segments of single-channel EEG signals containing epileptic spikes of two patients after pre-processing.

frequency of 50 Hz and a bandwidth of 2 Hz; and a high-pass filter with the cutoff frequency of 0.5 Hz. Segments of single-channel EEG signals containing epileptic spikes of two patients after pre-processing are shown in Fig. 1.

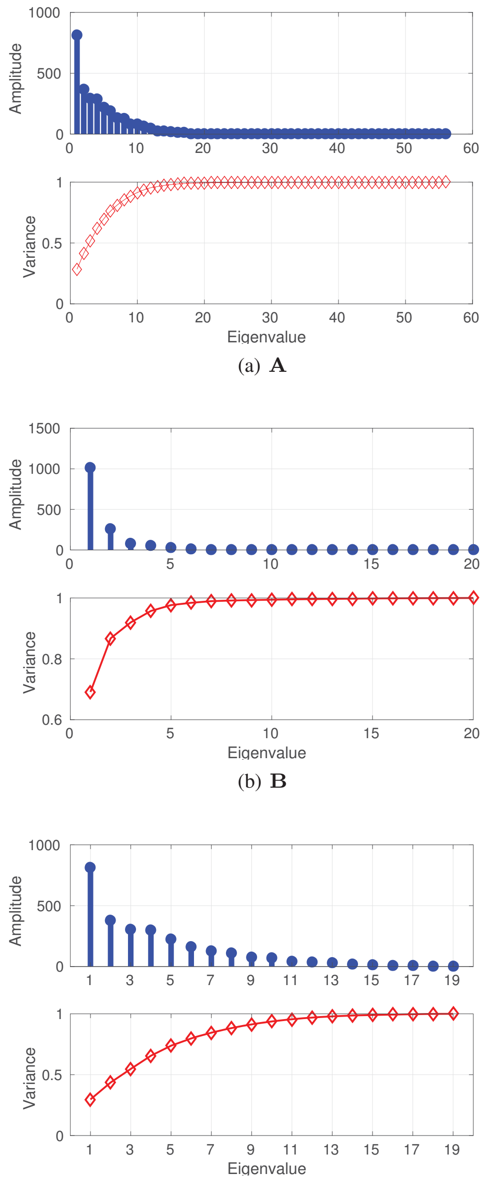
From this dataset, 1439 epileptic tensors and 6888 non-epileptic tensors are first constructed. The continuous wavelet transform (CWT) was performed on each multi-channel EEG segment of 19 channels and 56 samples containing a spike, as in [3]. We increase the number of wavelet scales in the dominant range $[4-8]$ to 20, instead of 5 as in [3]. Hence, 19 wavelet coefficient matrices of sizes 56×20 are obtained, representing EEG wavelet features. These 19 coefficient matrices are concatenated into an epileptic tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}_+^{56\times 20\times 19}$. The non-epileptic tensor is constructed in a similar manner from 6888 randomly selected non-epileptic spikes.

B. Experiments and Results

In the experiments, these three tasks are carried out: (i) extracting features, (ii) selecting features, and (iii) performing feature classification using KNN and NB classifiers. The training set is composed of randomly selected 1000 epileptic tensors and 2000 non-epileptic tensors, while the remaining tensors are used for testing.

1) Feature Extraction: The first task is to determine the multilinear rank (r_1, r_2, r_3) of EEG tensors.

Fig. 2 shows the spectra and variances of three covariance matrices for three tensor modes. As presented in the figure, there are about 20 significant eigenvalues for $\mathbf{A}$ (time), since of $\text{VAR}_{r_1} = 100\%$, with $r_1 \geq 20$. Among them, the first 15 eigenvalues are more significant than the others, with their variance almost reaching 100% of the full variance (Fig. 2(a)). So the number of components for mode time is set as 15. Similarly, the numbers of components for modes scale and channel are set as 10 and 19, respectively.



(a) $\mathbf{A}$

(b) $\mathbf{B}$

(c) $\mathbf{C}$

Fig. 2: Spectra of three modes of epileptic tensor.

The four-way epileptic tensor $\tilde{\boldsymbol{\mathcal{X}}}^{\text{ep}} \in \mathbb{R}_+^{56\times 20\times 19\times 1000}$ is constructed by concatenating 1000 epileptic tensors of the training set, from which the common factors $\mathbf{A} \in \mathbb{R}_+^{56\times 15}$, $\mathbf{B} \in \mathbb{R}_+^{20\times 10}$, $\mathbf{C} \in \mathbb{R}_+^{19\times 19}$ were derived with NTD. The factors for the non-epileptic tensor are also obtained in a similar way.

Experimental results showed significant differences between the NTD factors of epileptic tensor and non-epileptic tensor. For factor $\mathbf{A}$ (Fig. 3), all components of epileptic spikes were well localized, whereas most those of non-epileptic tensor seemed either not well localized or spread (except from components #1, #7 and #8). For

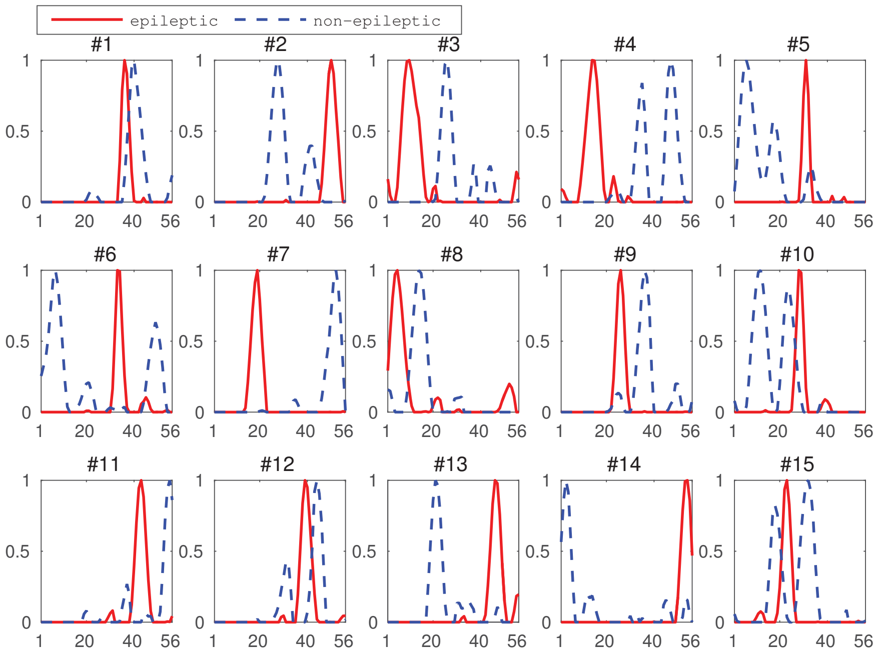


Fig. 3: Common factor $\mathbf{A} \in \mathbb{R}_{+}^{56\times 15}$ derived from NTD of epileptic tensor and non-epileptic tensor.

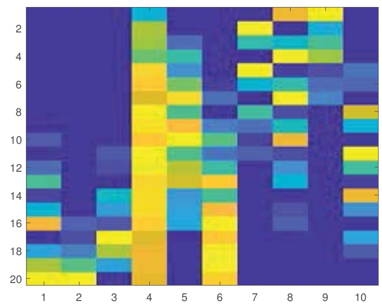
factor $\mathbf{B}$ (Fig. 4), the epileptic tensor behaved differently from the non-epileptic tensor, however it was not easy to draw any remark on the difference. For factor $\mathbf{C}$ (Fig. 5), epileptic tensor was well localized to specific regions on the head (we excluded factor $\mathbf{C}$ of non-epileptic tensor due to the limit of space, but it was not localized). As a result, the localization property in components of epileptic tensor may lead to the ability of learning localized parts.

2) Feature Selection: In order to evaluate how effective the feature selection stage is at detecting EEG epileptic spikes, the extracted features are fed into the k-nearest neighbor (KNN) classifier. The distance metric used in the KNN is the Euclidean distance. The number of nearest neighbors is fixed to 5 in all experiments. We also use the p-value to provide the strength of ranked features derived by the Fisher score. In particular, the null hypothesis H_0 is that there is no difference between the mean of epileptic class and non-epileptic class. A p-value smaller than 0.05 indicates that one has strong evidence against the null hypothesis H_0, so H_0 can be rejected. With a p-value larger than 0.05, one may fail to reject the H_0. The higher the Fisher score is, the smaller the p-value, and hence the more significant feature is selected.

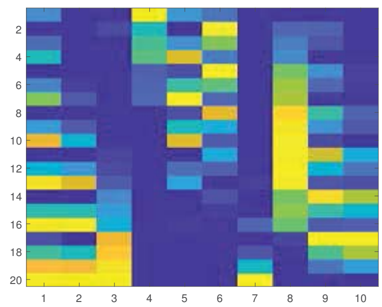
It can be seen from Fig. 7 that more than 600 significant features with largest Fisher scores had p-values smaller than 0.05, corresponding to 45% of the original 1425 features.

Among the features were the top 500 features having p-value close to 0, meaning that one can reject the null hypothesis H_0 completely. As a result, the 500 features had stronger discrimination power than others, as illustrated in Fig. 6.

Using these 500 features, the classification performance of KNN is given in Fig. 8. ACC (for accuracy) presents the proportion of the correctly predicted samples over the total number of EEG data samples, and SEN (for sensitivity) indicates the proportion of epileptic spikes that are correctly identified over the total number of labelled spikes. ACC and SEN reached their maximum of 90% (with



(a) Epileptic spike



(b) Non-epileptic spike

Fig. 4: Common factor $\mathbf{B} \in \mathbb{R}_{+}^{20\times 10}$ derived from NTD. The x-axis denotes the number of components (column vectors), while the y-axis presents 20 wavelet scales in the range of $[4-8]$.

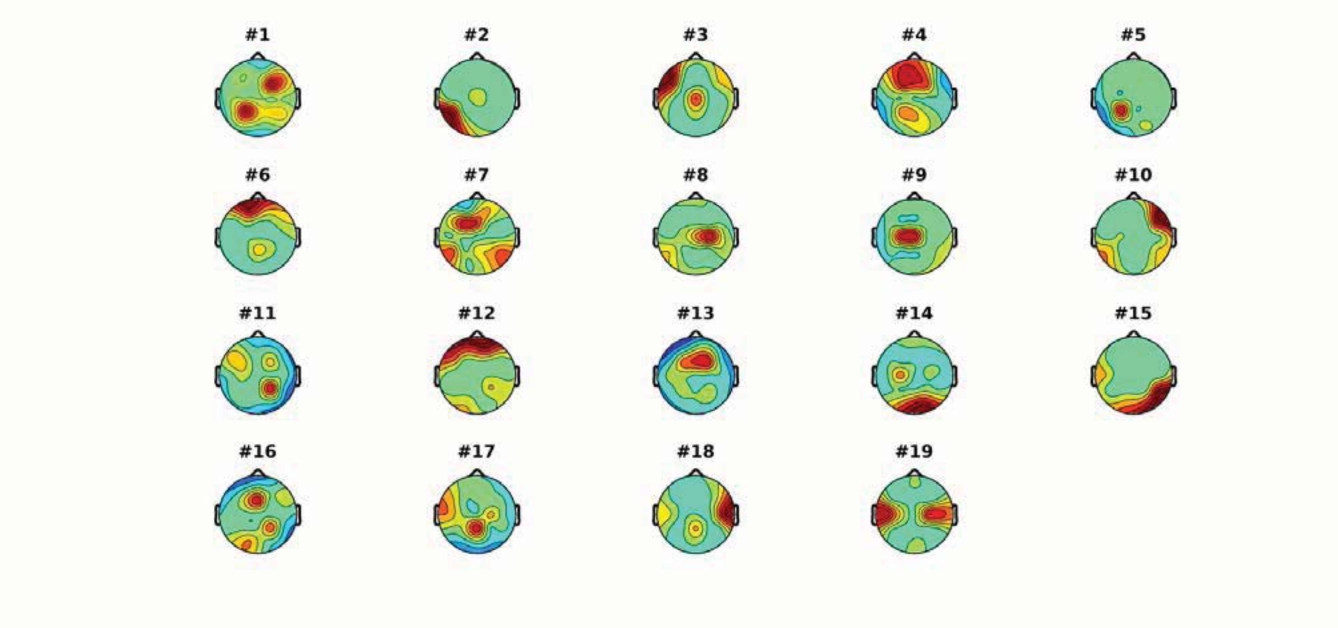


Fig. 5: Common factor $\mathbf{C} \in \mathbb{R}_{+}^{19\times 19}$ of the epileptic tensor derived from NTD of epileptic tensor.

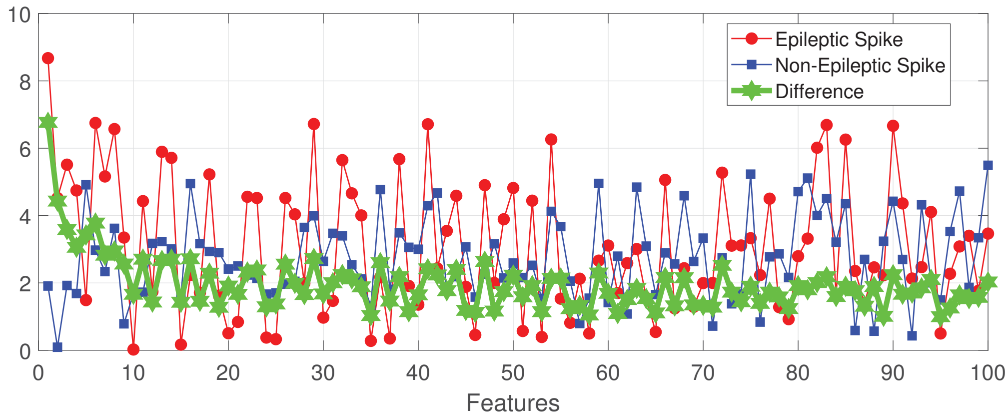


Fig. 6: Top 100 significant features of epileptic class and non-epileptic class.

180 features) and 93%, respectively. It indicates that that best results can be obtained at low dimensionality, i.e. around 180 features.

3) Classification: Performance comparison of epileptic spike detection between our method and Phan-Cichocki's method [13] is provided via numerical simulation, using two well-known classifiers: KNN and Naive Bayes (NB).

For KNN, the distance metric is the Euclidean distance and the size of the neighborhood was automatically obtained by setting the cross-validation option. For NB, the Gaussian distribution was used as the predictor distribution to compute the posterior probability for the two classes, and then make decision for the class with higher probability.

Apart from the performance metrics ACC and SEN, we also use SPE (for specificity). SPE provides similar information as SEN but for non-epileptic spikes. The receiver operating characteristic (ROC) curve and the area under this curve (AUC) are also used to illustrate the performance of the system. The performance results are shown in Table II and Fig. 9. It is clear that all the evaluation metrics of two classifiers, when training with our extracted features, were much better than those by Phan-Cichocki's method. Overall, our method yielded an excellent classfication result[1], since AUC of KNN and NB were both greater than 0.9.

[1]Performance ranking based on AUC: $[0.9-1]$ is excellent, $[0.8-0.9]$ is good, $[0.7-0.8]$ is fair, $[0.6-0.7]$ is poor, $[0.5-0.6]$ is fail.

TABLE II: Classification performance, top 100 significant features

	Our features		Phan-Cichocki features	
Metric	KNN	NB	KNN	NB
SEN	**84.05%**	**84.05%**	65.38%	65.38%
SPE	**92.94%**	82.53%	87.66%	75.68%
ACC	91.85%	82.75%	85.53%	75.95%
AUC	0.9339	0.9097	0.8263	0.8481

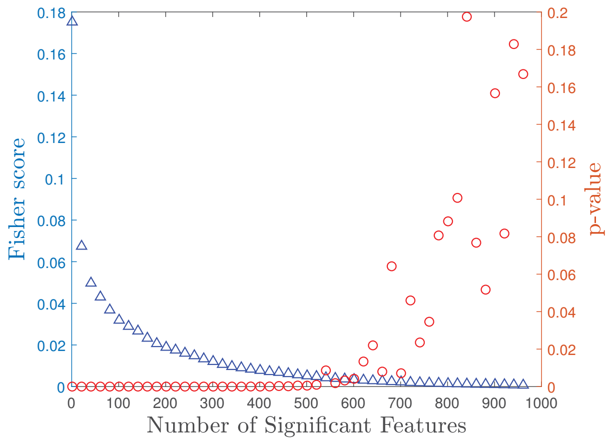


Fig. 7: Fisher Score

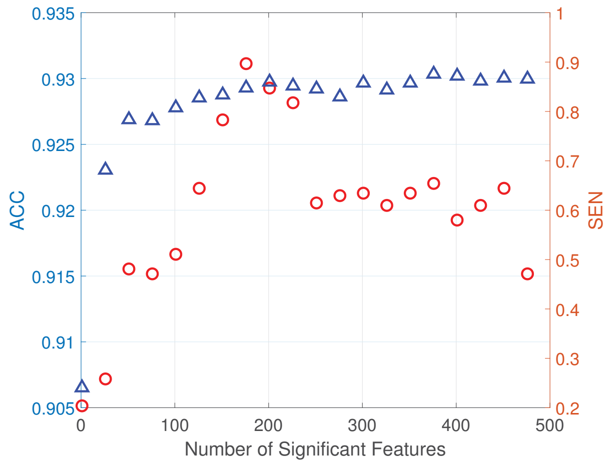


Fig. 8: Classification performance vs. number of selected features.

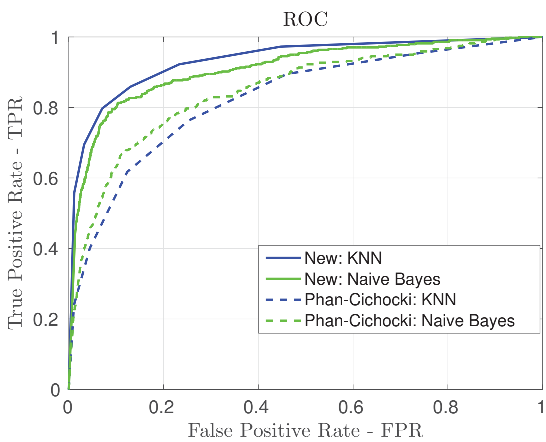


Fig. 9: ROC: A performance comparison of epileptic spike detection between using our method and Phan-Cichocki's method.

V. Conclusions

In this paper, inspired by the advantages of NTD, we have proposed a new approach to detect epileptic spikes in EEG data. We first derived a new feature space that can span EEG epileptic spikes from sparse loading factors of NTD. A new discriminant set of features was learned from NTD which can distinguish between epileptic class and non-epileptic class with high accuracy. In order to reduce feature dimensionality as well as to achieve the best separability between these classes, we used the Fisher score as the EEG feature selection model. The numerical experiments indicated that EEG multi-way analysis using NTD allowed us to not only extract multi-domain features of epileptic spikes but also provide high classification accuracy only with "shallow" classifiers such as KNN and NB.

Acknowledgments

This research was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under grant number 102.02-2015.32.

References

[1] L. Prilipko, *Atlas: Epilepsy care in the world*. Switzerland: WHO Press, 2005.

[2] A. T. Tzallas, M. G. Tsipouras, D. G. Tsalikakis, E. C. Karvounis, L. Astrakas, S. Konitsiotis, and M. Tzaphlidou, "Automated epileptic seizure detection methods: A review study," in *Epilepsy-histological, electroencephalographic and psychological aspects*, D. Stevanovic, Ed. InTech, 2012, ch. 4, pp. 75–98.

[3] N. T. Anh-Dao, N. Linh-Trung, L. V. Nguyen, T. Tran-Duc, N. T. H. Anh, and B. Boashash, "A multistage system for automatic detection of epileptic spikes," *REV Journal on Electronics and Communications*, vol. 8, no. 1–2, pp. 1–13, 2018.

[4] K. Indiradevi, E. Elias, P. Sathidevi, S. Dinesh Nayak, and K. Radhakrishnan, "A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram," *Computers in Biology and Medicine*, vol. 38, no. 7, pp. 805–816, 2008.

[5] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa, and H. A. Phan, "Tensor decompositions for signal processing applications: From two-way to multiway component analysis," *IEEE Signal Processing Magazine*, vol. 32, no. 2, pp. 145–163, 2015.

[6] F. Cong, Q.-H. Lin, L.-D. Kuang, X.-F. Gong, P. Astikainen, and T. Ristaniemi, "Tensor decomposition of eeg signals: a brief review," *Journal of neuroscience methods*, vol. 248, pp. 59–69, 2015.

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

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

[9] W. Deburchgraeve, P. J. Cherian, M. De Vos, R. M. Swarte, J. H. Blok, G. H. Visser, P. Govaert, and S. Van Huffel, "Neonatal seizure localization using parafac decomposition," *Clinical Neurophysiology*, vol. 120, no. 10, pp. 1787–1796, 2009.

[10] E. Pippa, V. G. Kanas, E. I. Zacharaki, V. Tsirka, M. Koutroumanidis, and V. Megalooikonomou, "EEG-based classification of epileptic and non-epileptic events using multi-array decomposition," *International Journal of Monitoring and Surveillance Technologies Research (IJMSTR)*, vol. 4, no. 2, pp. 1–15, 2016.

[11] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: Applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[12] G. Zhou, A. Cichocki, Q. Zhao, and S. Xie, "Nonnegative matrix and tensor factorizations: An algorithmic perspective," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 54–65, May 2014.

[13] A. H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets," *Nonlinear theory and its applications*, vol. 1, no. 1, pp. 37–68, 2010.

[14] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[15] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(r_1, r_2,..., r_n) approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.