Multi-Channel EEG Epileptic Spike Detection by a New Method of Tensor Decomposition

Le Trung Thanh[1], **Nguyen Thi Anh Dao**[1,2], **Nguyen Viet Dung**[1,3], **Nguyen Linh Trung**[1,*], **Karim Abed-Meraim**[4]

[1] Advanced Institute of Engineering and Technology (AVITECH), VNU University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam
[2] University of Technology and Logistics, Bac Ninh, Vietnam
[3] UMR6285 CNRS ENSTA Bretagne, Brest, France
[4] University of Orléans, Orléans, France
* Nguyen Linh Trung

E-mail: linhtrung@vnu.edu.vn

Received 17 October 2018, revised 25 June 2019, accepted 29 October 2019

Abstract. *Objective:* Epilepsy is one of the most common brain disorders. For epilepsy diagnosis or treatment, the neurologist needs to observe epileptic spikes from electroencephalography (EEG) data. Since multi-channel EEG records can be naturally represented by multi-way tensors, it is of interest to see whether tensor decomposition is able to analyze EEG epileptic spikes. *Approach:* In this paper, we first proposed the problem of simultaneous multilinear low-rank approximation of tensors (SMLRAT) and proved that SMLRAT can obtain local optimum solutions by using two well-known tensor decomposition algorithms (HOSVD and Tucker-ALS). Second, we presented a new system for automatic epileptic spike detection based on SMLRAT. *Main results:* We propose to formulate the problem of feature extraction from a set of EEG segments, represented by tensors, as the SMLRAT problem. Efficient EEG features were obtained, based on estimating the "eigenspikes" derived from nonnegative GSMLRAT. We compared the proposed tensor analysis method with other common tensor methods in analyzing EEG signal and compared the proposed feature extraction method with the state-of-the-art methods. Experimental results indicated that our proposed method is able to detect epileptic spikes with high accuracy. *Significance:* Our method, for the first time, makes a step forward for automatic detection EEG epileptic spikes based on tensor decomposition. The method can provide a practical solution to distinguish epileptic spikes from artifacts in real-life EEG datasets.

Keywords: Electroencephalography (EEG), epileptic spikes, multilinear low-rank approximation, tensor decomposition, nonegative Tucker decomposition, feature extraction, feature selection.

1. Introduction

According to estimates [1] in 2010, epilepsy affects about 50 million people worldwide in which nearly 40 million people live in developing countries. In electroencephalography (EEG) records of the brain, epilepsy biomarkers are seizures and epileptiforms (e.g. spikes, sharp waves and spike-wave complexes), which are resulted from abnormal and excessive electrical discharges of nerve cells.

For epilepsy diagnosis and treatment, one needs to observe epileptic seizures or epileptiforms to help identify the type of epilepsy and the affected area of the brain. Since epileptic spikes are interictal (i.e., they occur in between seizures) while seizures occur sparsely in time, one normally obtains EEG records which contain various spikes. To better detect the epileptic spikes in long EEG records or to reduce false-alarm detection (which is often the case since various other non-epileptic spikes also co-exist in EEG), automatic detection of spikes by software programs/systems is advantageous over visual reading by neurologists, and thus has been a subject of engineering and science studies for several decades [2–6].

For epileptic spike detection, most studies have focused on analysis of single-channel EEG signals, each of which is obtained from an EEG electrode. Especially, recent efforts are seen in developing multi-stage detection systems that take into account of various types of information (i.e., electrical, physiological and morphological) of the spikes [7–9]. However, each EEG record simultaneously collects signals from multiple electrodes, resulting in a *multi-channel* EEG signal. Since epilepsy is often caused by an affected area in the brain, several electrodes may be able to pick up the resulting epilepsy biomarkers around the same time at which the single-channel EEG signals are spatially correlated across the channels. Therefore, analysis of multi-channel EEG signals may enhance the detection of epileptic spikes. Currently, there exists only one study on spike detection that deals with multi-channel EEG signals [10].

Multi-channel EEG signals can be naturally represented by matrices which are two-way tensors (when considering time and channel domains), or multi-way tensors (when considering more than two domains, e.g. of time, frequency, space, trial, condition, subject and group). Many studies have used tensor decomposition for EEG signals in general and for epileptic seizures in particular [11–19]. To the best of our knowledge, there exists no study applying tensor decomposition to detect EEG epileptic spikes. Therefore, the aim of our study is to propose a new method of tensor decomposition to deal with this challenging problem, that is to separate two types of spikes in an EEG dataset, including real epileptic spikes and non-epileptic ones. Real epileptic spikes are those recognized and labeled by the neurologists, whereas non-epileptic spikes are not related to epilepsy but may be easily misdiagnosed as epileptic spikes. Non-epileptic spikes are large positive or negative voltage transients that can be confused as epileptic spikes by regular algorithms.

In many analysis and classification systems, low-rank matrix approximation (LRAM) and its multi-way extension – low-rank tensor approximation (LRAT) – play important role for dimensionality reduction, feature extraction and feature selection [20]. In this paper, we are interested in dealing with *a sequence* of matrices and tensors and, hence, the problems of *simultaneous* LRA of multiple matrices (SLRAM) and tensors (SLRAT), as well as their applications in classification. We now review simultaneous LRA approaches in general (i.e., not limited to EEG applications), while noting that the literature for LRA of a single matrix or tensor, and their applications can be found in recent reviews [21, 22].

The common idea in dimensionality reduction is to seek a linear or multilinear subspace embedded in a high-dimensional manifold which represents the dataset of interest. Then depending on the applications at hand, different approaches can be taken. For SLRAM, the resulting two-dimensional subspace methods include: two-dimensional singular value decomposition (2dSVD) [23, 24], two-dimensional principal component analysis (2dPCA) [25, 26], population value decomposition (PVD) [27], generalized low-rank approximation of matrices (GLRAM) [28–32], two-dimensional linear discriminant analysis (2dLDA) [33], and simultaneous component analysis (SCA) [34, 35]. These methods can be categorized in two main approaches: non-iterative-based and iterative-based algorithms. The former [23–27, 29] provides sub-optimal solutions, but is simple and efficient in practice. The latter [28, 30–32, 35] can yield optimal solutions, but follows procedures that are time-consuming.

For SLRAT, the resulting higher dimensional subspace methods include: manifold regularization non-

negative Tucker decomposition (MR-NTD) [36], concurrent subspace analysis (CSA) [37], multilinear discriminant analysis (MLDA) [38], multilinear PCA (MPCA) including unconstrained MPCA [39], nonnegative MPCA [40] and sparse MPCA [41]. In a perspective of data fusion perspective [42], the problem of SLRAT can be considered as a special type where data inputs have the same size. Data fusion is to look for common factors sharing meaningful information between datasets. When the data are represented by multiple matrices or tensors, *simultaneous* decomposition (a.k.a., joint factorization) methods are the key for fusing the data.

In light of the above literature, the work in this paper has two main contributions. Based on the preliminary result in [43], the first contribution of this paper is the proposal of a new method for SLRAT, generalizing SLRAM from matrices to tensors. In particular, we introduce *simultaneous multilinear* LRAT (SMLRAT) in which different tensors with identical dimensions are factorized so that (i) all tensors share common factor matrices and (ii) each tensor has it own core-tensor. By a theoretical analysis, we then show that SMLRAT can obtain local optimum solutions by using two well-known tensor decomposition algorithms: higher-order singular value decomposition (HOSVD) and higher-order orthogonal iteration (HOOI or Tucker alternating least-squares (Tucker-ALS). We further develop a local solution for *nonnegative* SMLRAT since our analysis aims to EEG signals for which the nonnegativity constraint plays an important role [44]. Finally, inspired by the feature extraction algorithm proposed in [45], we propose a generalized SMLRAT algorithm (GSMLRAT) to effectively solve the SMLRAT and nonnegative SMLRAT problems. We note that the SLRAT methods are considered as special cases of our proposed GSMLRAT, and is analyzed in conjunction with other methods, as shown in our technical report of [46]. Also, a similar approach to [45] can be found in [47, 48], which is based on the linear system with a CP decomposition constrained solution (LS-CPD) framework for data classification.

The second contribution of the paper is, for the first time, the successful application of tensor decomposition for detection of EEG epileptic spikes. We propose to formulate the problem of feature extraction from a set of EEG segments, represented by tensors, as the SMLRAT problem. We first obtain efficient EEG features, based on estimating the "eigenspikes" derived from nonnegative GSMLRAT. We then apply the Fisher score as the feature selection method for selecting significant features. These selected features are then fed into the widely used classifiers to evaluate their separability between epileptic and non-epileptic spikes. Due to artifacts, false detection of epileptic spikes often occurs, possibly leading to misdiagnosis. Our algorithm, however, can accurately distinguish real events from artifacts (See Table 2 for further details). Note that, nonnegative Tucker decomposition (NTD) has previously been applied for multi-domain feature extraction in EEG analysis [49]. In this paper, similar to [49], EEG segments were also represented by higher order tensors and then decomposed by NTD to derive their features. However, our work is different from [49] in both objective and method. Particularly, in [49], NTD was applied to extract the multi-domain feature of visual mismatch negativity (i.e., event-related potentials) for the cognitive research while the objective of this study is to detect epileptic spikes. In addition, the approach in [49] does not exploit the relation among inputs, while our method aims to investigate the relation among input tensors by looking for a "common" feature space of activities of interest (e.g. epileptic spikes). Besides, the core tensor in conjunction with factors are used as feature in our work instead of only core tensor as [49].

The paper is organized as follows. In Section 2, a brief review of tensors and related operators, as well as the multilinear LRAT problem will be presented. In Section 3, we present the SLRAM problem, and hence propose the SMLRAT problem that extends SLRAM from matrices to general tensors and nonnegative tensors. Section 4 describes a new tensor-based epileptic spike detection system and Section 5 shows experimental results.

2. Preliminaries

2.1. Tensor Notations and Definitions

Follow notations defined in [50], we use lowercase letters (e.g. a), boldface lowercase letters (e.g. $\mathbf{a}$), boldface capital letters (e.g. $\mathbf{A}$) and bold calligraphic letters (e.g. $\boldsymbol{\mathcal{A}}$) to denote scalars, vectors, matrices and tensors respectively. Moreover, we summarize here some useful tensor operators, to be used later.

The mode-k unfolding of a tensor $\boldsymbol{\mathcal{A}}$ is a matrix in vector space $\mathbb{R}^{I_k \times (I_1 \dots I_{k-1} I_{k+1} \dots I_n)}$, where I_k is the integer number presenting the dimension of the k-th vector space generating the tensor, denoted as $\mathbf{A}_{(k)}$, whose elements are defined by

$$\mathbf{A}_{(k)}(i_k, \overline{i_1 \dots i_{k-1} i_{k+1} \dots i_n}) = \boldsymbol{\mathcal{A}}(i_1, i_2, \dots, i_n),$$

where $\overline{i_1 i_2 \dots i_n}$ is a multi-index, which combines multiple indices $i_1, i_2, \dots, i_n$ together in a single index used regularly in vectorization/matricization for tensors [21], given by

$$\overline{i_1 i_2 \dots i_n} = i_1 + (i_2 - 1)I_1 + (i_3 - 1)I_1 I_2 + \dots + (i_n - 1)I_1 I_2 \dots I_N.$$

The k-mode product of $\boldsymbol{\mathcal{A}}$ with a matrix $\mathbf{U} \in \mathbb{R}^{r_k \times I_k}$, written as $\boldsymbol{\mathcal{A}} \times_k \mathbf{U}$, yields a new tensor $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \times \cdots \times I_{k-1} \times r_k \times I_{k+1} \cdots \times I_n}$ such that its k-mode unfolding is given by $\mathbf{B}_{(k)} = \mathbf{U}\mathbf{A}_{(k)}$. Useful properties for the k-mode product follow:

$$\begin{aligned} &\boldsymbol{\mathcal{A}} \times_k \mathbf{U} \times_l \mathbf{V} = \boldsymbol{\mathcal{A}} \times_l \mathbf{V} \times_k \mathbf{U} \quad \text{for} \quad k \neq l, \\ &\boldsymbol{\mathcal{A}} \times_k \mathbf{U} \times_k \mathbf{V} = \boldsymbol{\mathcal{A}} \times_k (\mathbf{V}\mathbf{U}). \end{aligned}$$

The inner product of two n-way tensors $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_n}$ is defined by

$$\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \rangle = \sum_{i_1=1}^{I_1} \cdots \sum_{i_n=1}^{I_n} \boldsymbol{\mathcal{A}}(i_1, i_2, \ldots, i_n) \boldsymbol{\mathcal{B}}(i_1, i_2, \ldots, i_n).$$

The Frobenius norm of a tensor $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_n}$ is defined by the inner product of $\boldsymbol{\mathcal{A}}$ with itself

$$\|\boldsymbol{\mathcal{A}}\|_F = \sqrt{\langle \boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{A}} \rangle}.$$

The concatenation of $\boldsymbol{\mathcal{A}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_n}$ and a tensor $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_{n-1}}$ yields a new tensor $\boldsymbol{\mathcal{C}} = \boldsymbol{\mathcal{A}} \boxplus \boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \times \cdots \times I_{n-1} \times (I_n+1)}$ such that

$$\boldsymbol{\mathcal{C}}(i_1, \ldots, i_n) = \begin{cases} \boldsymbol{\mathcal{A}}(i_1, \ldots, i_{n-1}, i_n), & \text{if } i_n \leqslant I_n, \\ \boldsymbol{\mathcal{B}}(i_1, \ldots, i_{n-1}), & \text{if } i_n = I_n + 1. \end{cases}$$

Remark that, a (n-1)-way tensor $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_{n-1}}$ can be represented by a n-way tensor $\boldsymbol{\mathcal{E}} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_{n-1} \times 1}$, so the operator can be used for concatenating the two n-way tensors.

For operators on a matrix $\mathbf{A} \in \mathbb{R}^{I_1 \times I_2}$, $\mathbf{A}^T$ and $\mathbf{A}^{\#}$ denote the transpose and the pseudo-inverse of $\mathbf{A}$ respectively. The Kronecker product of $\mathbf{A}$ with a matrix $\mathbf{B} \in \mathbb{R}^{J_1 \times J_2}$, denoted by $\mathbf{A} \otimes \mathbf{B}$, yields a matrix $\mathbf{C} \in \mathbb{R}^{I_1 J_1 \times I_2 J_2}$ defined by

$$\mathbf{C} = \mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{1,1}\mathbf{B} & \ldots & a_{1,I_2}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{I_1,1}\mathbf{B} & \ldots & a_{I_1,I_2}\mathbf{B} \end{bmatrix}.$$

2.2. Multilinear Low-Rank Tensor Approximation

Multilinear LRAT (MLRAT) can be considered as a generalization of LRAM for tensors [21]. In particular, MLRAT of a tensor $\boldsymbol{\mathcal{X}}$ requires the following optimization problem:

$$\begin{aligned} \arg\min_{\widetilde{\boldsymbol{\mathcal{X}}}} \quad & f_{\text{MLRAT}} = \|\boldsymbol{\mathcal{X}} - \widetilde{\boldsymbol{\mathcal{X}}}\|_F^2, \\ \text{s.t.} \quad & \widetilde{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n, \\ & \text{rank}(\mathbf{U}_k) \leqslant r_k, \quad k = 1, 2, \ldots, n, \end{aligned} \tag{1}$$

where $\boldsymbol{\mathcal{G}}$ is called the core tensor of $\boldsymbol{\mathcal{X}}$, $\{\mathbf{U}_k\}_{k=1}^n$ are called factors of $\boldsymbol{\mathcal{X}}$, and a set of $\{r_k\}_{k=1}^n$ is the desired low multilinear rank. Next, we present the connection of MLRAT with several types of tensor decomposition.

2.2.1. CP decomposition

This decomposition can be considered as constrained MLRAT, where the core tensor is diagonal and the factors have the same rank. Specifically, f_{MLRAT} can be expressed to CP as

$$\begin{aligned} \arg\min_{\widetilde{\boldsymbol{\mathcal{X}}}} \quad & f_{\text{CP}} = \|\boldsymbol{\mathcal{X}} - \widetilde{\boldsymbol{\mathcal{X}}}\|_F^2, \\ \text{s.t.} \quad & \widetilde{\boldsymbol{\mathcal{X}}} = \sum_{i=1}^{r} \lambda_i \mathbf{U}_1(:, i) \circ \cdots \circ \mathbf{U}_n(:, i), \\ & \text{rank}(\mathbf{U}_k) = r, \quad k = 1, 2, \ldots, n, \end{aligned} \tag{2}$$

where "$\circ$" presents the outer product, the factors $\mathbf{U}_k \in \mathbb{R}^{I_k \times r}$ are full column-rank and $\{\lambda_i\}_{i=1}^r$ are diagonal entries of the core tensor $\boldsymbol{\mathcal{G}}$. To solve f_{CP} in (2), the "workhorse" algorithm is based on alternating least squares (ALS) [50, 51].

2.2.2. Tucker Decomposition

This decomposition is more flexible than CP, where the core-tensor are not required to be diagonal while the factors are orthogonal matrices, i.e,

$$\begin{aligned} \arg\min_{\widetilde{\boldsymbol{\mathcal{X}}}} \quad & f_{\text{Tucker}} = \|\boldsymbol{\mathcal{X}} - \widetilde{\boldsymbol{\mathcal{X}}}\|_F^2 \\ \text{s.t.} \quad & \widetilde{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n, \\ & \mathbf{U}_k^T \mathbf{U}_k = \mathbf{I}_{r_k}, \quad k = 1, 2, \ldots, n, \end{aligned} \tag{3}$$

where $\mathbf{I}_{r_k} \in \mathbb{R}^{r_k \times r_k}$ denotes the identity matrix. As a result, solution of f_{Tucker} in (3) is not unique in general, but the subspaces spanned by $\{\mathbf{U}_k\}_{k=1}^n$ are physically unique [50, Section IV]. Two well-known algorithms for solving f_{Tucker} are HOSVD and Tucker-ALS. Depending on applications, both HOSVD and Tucker-ALS can provide good approximation. Moreover, in many practical implementations, HOSVD is used as a starting point (i.e., initialization) to further accelerate the convergence of Tucker-ALS [52].

2.2.3. Nonnegative Tensor Decomposition

This type of decomposition is considered as a generalization of nonnegative matrix factorization for tensors, where the nonnegativity constraint is imposed on the factors and/or the core tensor [44]. Specifically, nonnegative tensor decomposition (NTD) can be seen as a nonnegative f_{MLRAT} as

$$\begin{aligned} \arg\min_{\widetilde{\boldsymbol{\mathcal{X}}}} \quad & f_{\text{NTD}} = \|\boldsymbol{\mathcal{X}} - \widetilde{\boldsymbol{\mathcal{X}}}\|_F^2, \\ \text{s.t.} \quad & \widetilde{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n, \\ & \boldsymbol{\mathcal{G}} \geq 0, \mathbf{U}_k \geq 0, \quad k = 1, 2, \ldots, n, \end{aligned} \tag{4}$$

where the notation $\geq$ means that all entries of the matrix/tensor are nonnegative.

3. Generalized Simultaneous Multilinear Low-Rank Approximation of Tensors

In this section, we first present connection between the SLRAM and three-way tensor decomposition methods (i.e., HOSVD and Tucker-ALS). Motivated by such connection, we then propose a generalized approach for SLRAT, where the tensors are general or constrained to be nonnegative.

3.1. SLRAM and Tensor Decomposition

SLRAM problem [28]: *Given a set of N matrices $\mathbf{X}_1, \ldots, \mathbf{X}_N \in \mathbb{R}^{I_1 \times I_2}$, find two orthogonal matrices $\mathbf{U}_1 \in \mathbb{R}^{I_1 \times r_1}$ and $\mathbf{U}_2 \in \mathbb{R}^{I_2 \times r_2}$ and N matrices $\mathbf{F}_1, \ldots, \mathbf{F}_N \in \mathbb{R}^{r_1 \times r_2}$ such that $\mathbf{U}_1 \mathbf{F}_i \mathbf{U}_2^T$, $i = 1, \ldots, N$, yield good approximates of $\mathbf{X}_i$.*

Solving SLRAM is equivalent to finding the solution of

$$\begin{aligned} \arg \min_{\{\mathbf{F}_i\}_{i=1}^N, \mathbf{U}_1, \mathbf{U}_2} \quad & f_{\text{SLRAM}} = \sum_{i=1}^{N} \|\mathbf{X}_i - \mathbf{U}_1 \mathbf{F}_i \mathbf{U}_2^T\|^2 \\ \text{s.t.} \quad & \mathbf{U}_1^T \mathbf{U}_1 = \mathbf{I}_{r_1}, \text{ and } \mathbf{U}_2^T \mathbf{U}_2 = \mathbf{I}_{r_2}. \end{aligned} \tag{5}$$

Let us define a three-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times N}$ concatenating inputs such that each slide $\boldsymbol{\mathcal{X}}_{:,:,i}$ of $\boldsymbol{\mathcal{X}}$ is the input matrix $\mathbf{X}_i$. Then tensor $\boldsymbol{\mathcal{X}}$ can be expressed as $\boldsymbol{\mathcal{X}} = \mathbf{X}_1 \boxplus \mathbf{X}_2 \cdots \boxplus \mathbf{X}_N$.

It is well-known that Tucker-ALS provides the local optimal solution [50, 52] of (3), while SLRAM was shown to be a special case of Tucker-ALS [53, Theorem 4.1]. Accordingly, we have the following proposition, showing the connection between SLRAM and Tucker-ALS, and hence providing good iterative-based approximation for SLRAM.

Proposition 1 ([53, Theorem 4.1])**.** *If $\mathbf{U}_1$ and $\mathbf{U}_2$ are the factors obtained from decomposing a three-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I \times J \times N}$ using Tucker-ALS, and let $\mathbf{F}_i = \mathbf{U}_1^T \mathbf{X}_i \mathbf{U}_2$, then $\mathbf{U}_1$, $\mathbf{U}_2$ and $\mathbf{F}_1, \ldots, \mathbf{F}_N$ form a (local) optimal solution of f_{SLRAM} in (5).*

It is also well-known that HOSVD gives a sub-optimal solution of (3) [50, 52]. Accordingly, we have the following connection between SLRAM and HOSVD, providing good non-iterative-based approximation for SLRAM.

Proposition 2 ([24, Section IV])**.** *If $\mathbf{U}_1$ and $\mathbf{U}_2$ are the factors obtained from decomposing a three-way tensor $\boldsymbol{\mathcal{X}}$ using HOSVD, and let $\mathbf{F}_i = \mathbf{U}_1^T \mathbf{X}_i \mathbf{U}_2$, then $\mathbf{U}_1$, $\mathbf{U}_2$ and $\{\mathbf{F}_i\}_{i=1}^N$ form a sub-optimal solution of f_{SLRAM} in (5).*

3.2. Generalized Simultaneous Multilinear LRAT

Inspired by results in Section 3.1, we first state the following simultaneous multilinear low-rank tensor approximation (SMLRAT) problem.

SMLRAT problem: *Given a set of N n-way tensors $\{\boldsymbol{\mathcal{X}}_i\}_{i=1}^N, \boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_n}$, find n common factors $\{\mathbf{U}_k\}_{k=1}^n, \mathbf{U}_k \in \mathbb{R}^{I_k \times r_k}$ and N core tensors $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^N, \boldsymbol{\mathcal{G}}_i \in \mathbb{R}^{r_1 \times r_2 \cdots \times r_n}$ such that $\boldsymbol{\mathcal{G}}_i \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n$, $i = 1, \ldots, N$, yield good approximates of $\boldsymbol{\mathcal{X}}_i$.*

The problem can be considered as a generalization of SLRAM for multi-way tensors and formulated as follows:

$$\begin{aligned} \arg \min_{\{\widetilde{\boldsymbol{\mathcal{X}}}_i\}_{i=1}^N} \quad & f_{\text{SMLRAT}} = \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i - \widetilde{\boldsymbol{\mathcal{X}}}_i\|_F^2 \\ \text{s.t.} \quad & \widetilde{\boldsymbol{\mathcal{X}}}_i = \boldsymbol{\mathcal{G}}_i \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n. \end{aligned} \tag{6}$$

To solve (6), we propose the following theorem which provides (local) optimal solution. Details of the proof of this theorem can be found in the technical report of [46].

Theorem 1. *A local optimum solution of the SMLRAT problem is given by*

$$\boldsymbol{\mathcal{G}}_i = \boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_n \mathbf{U}_n^T, \tag{7}$$

with $\mathbf{U}_k$, $k = 1, 2, \ldots, n$, including the principal r_k eigenvectors of the covariance matrix $\mathbf{R}_k$ defined by

$$\mathbf{R}_k = \sum_{i=1}^{N} \boldsymbol{\mathcal{X}}_{i(k)} \widetilde{\mathbf{U}}_k \widetilde{\mathbf{U}}_k^T \boldsymbol{\mathcal{X}}_{i(k)}^T, \tag{8}$$

where $\widetilde{\mathbf{U}}_k$ is given by

$$\widetilde{\mathbf{U}}_k = \big(\mathbf{U}_n \otimes \cdots \otimes \mathbf{U}_{k+1} \otimes \mathbf{U}_{k-1} \cdots \otimes \mathbf{U}_1\big). \tag{9}$$

Let $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times N}$ be the tensor formed by concatenating N multi-way tensors $\boldsymbol{\mathcal{X}}_1, \ldots, \boldsymbol{\mathcal{X}}_N$; that is, $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_1 \boxplus \boldsymbol{\mathcal{X}}_2 \cdots \boxplus \boldsymbol{\mathcal{X}}_N$. Inspired by Theorem 1, a practical solution for the problem of f_{SMLRAT} in (6) can be achieved, using the Tucker-ALS algorithm, given by the following proposition.

Proposition 3. *If $\{\mathbf{U}_k\}_{k=1}^n$ are the factors obtained from decomposing the $(n+1)$-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times N}$ using Tucker-ALS, and core tensors $\boldsymbol{\mathcal{G}}_i$ are defined by*

$$\boldsymbol{\mathcal{G}}_i = \boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_n \mathbf{U}_n^T, \tag{10}$$

then $\{\mathbf{U}_k\}_{k=1}^n$ and $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^N$ can be a (local) optimal solution of f_{SMLRAT} in (6).

This result contributes a connection between Theorem 1 and the well-known Tucker-ALS algorithm for Tucker decomposition, thus allowing us to exploit known characteristics of this algorithm to enhance performance and/or reduce computational complexity of implementation [54–56]. Therefore, we can also obtain an alternative solution of SMLRAT using the HOSVD algorithm, as given by the following proposition.

Algorithm 1: GSMLRAT: Generalized Simultaneous Multilinear LRA of Tensors

Input: N n-way tensors $\{\boldsymbol{\mathcal{X}}_i\}_{i=1}^N, \boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{I_1 \times I_2 \times \cdots \times I_n}$, multilinear rank $\{r_1, r_2, \ldots, r_n\}$.

Output: common factors $\{\mathbf{U}_k\}_{k=1}^n$, core tensors $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^N$.

function

  Initialization:

    Contruct a concatenated $(n+1)$-way tensor $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_1 \boxplus \boldsymbol{\mathcal{X}}_2 \cdots \boxplus \boldsymbol{\mathcal{X}}_N$;

    Compute covariance matrices $\{\widetilde{\mathbf{R}}_{(k)}\}_{k=1}^n$ over modes of tensors as $\widetilde{\mathbf{R}}_{(k)} = \sum_{i=1}^N \boldsymbol{\mathcal{X}}_{i(k)} \boldsymbol{\mathcal{X}}_{i(k)}^T$;

    $\{\mathbf{U}_k^{(0)}\}_{k=1}^n$ are initilized by selecting the first eigenvectors of $\widetilde{\mathbf{R}}_{(k)}$ and $\mathbf{U}_{n+1}^{(0)} = \mathbf{I}_N$;

  Tucker decomposition (HOSVD, Tucker-ALS, NTD):

  $\boldsymbol{\mathcal{G}}, \{\mathbf{U}_k\}_{k=1}^{n+1} = \text{decompose}(\boldsymbol{\mathcal{X}}, \{\mathbf{U}_k^{(0)}\}_{k=1}^{n+1})$;

  Obtain core tensors: $\boldsymbol{\mathcal{G}}_i = \boldsymbol{\mathcal{G}}(:, :, \ldots, i)$

Proposition 4. *If $\{\mathbf{U}_k\}_{k=1}^n$ are factors obtained from decomposing the $(n+1)$-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times N}$ using HOSVD, and*

$$\boldsymbol{\mathcal{G}}_i = \boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_n \mathbf{U}_n^T,$$

then $\{\mathbf{U}_k\}_{k=1}^n$ and $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^N$ can be a sub-optimal solution of f_{SMLRAT} in (6).

To deal with nonnegative tensors, we can propose the nonnegative SMLRAT as

$$\begin{aligned} \arg \min_{\{\widetilde{\boldsymbol{\mathcal{X}}}_i\}_{i=1}^N} \quad & f_{\text{NSMLRAT}} = \sum_{i=1}^N \|\boldsymbol{\mathcal{X}}_i - \widetilde{\boldsymbol{\mathcal{X}}}_i\|_F^2 \\ \text{s.t.} \quad & \widetilde{\boldsymbol{\mathcal{X}}}_i = \boldsymbol{\mathcal{G}}_i \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n. \\ & \{\mathbf{U}_k\}_{k=1}^n \geqslant 0. \end{aligned} \tag{11}$$

Similarly, we can obtain a practical solution for the NSMLRAT problem using NTD.

Proposition 5. *If $\boldsymbol{\mathcal{G}}, \{\mathbf{U}_k\}_{k=1}^n$ are core tensor and factors obtained from performing NTD on the $(n+1)$-way tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{I_1 \times \cdots \times I_n \times N}$, and $\boldsymbol{\mathcal{G}}_i = \boldsymbol{\mathcal{G}}(:, :, \ldots, i)$, then $\{\mathbf{U}_k\}_{k=1}^n$ and $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^N$ can be a local solution of f_{NSMLRAT} in (11).*

Based on Theorem 1, Propositions 3, 4 and 5, we propose Algorithm 1, namely Generalized SMLRAT (GSMLRAT). Depending on kinds of constraints being considered (e.g. orthogonality, sparsity or nonnegativity), we can apply the corresponding tensor decomposition (e.g. HOSVD, Tucker-ALS and NTD) to obtain the desired solution.

4. Proposed Epileptic Spike Detection System

In this section, we introduce a novel epileptic spike detection system based on the proposed SLRMAT method. This system, illustrated in Figure 1, is composed of four stages: data transformation, EEG feature extraction, feature selection and classification.

In the data representation stage, three-way EEG tensors (time, wavelet-scale and channel) are calculated by applying the continuous wavelet transform on multi-channel EEG segments simultaneously. Then, magnitude of the resulting wavelet coefficients is used to construct nonnegative EEG tensors. In EEG feature extraction, we propose to estimate the so-called "eigenspikes". We also propose to use HOSVD for determining the multilinear rank for the three-way EEG tensors. In the feature selection stage, we propose to apply the Fisher score for selecting significant EEG features. In the classification stage, we use several well-known classifiers to assess system performance. Now, we will focus on our contributions to the second and the third stages.

4.1. Feature Extraction

Consider N three-way EEG tensors, $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}_+^{I_1 \times I_2 \times I_3}$ (whose dimensions I_1, I_2 and I_3 correspond to time, wavelet-scale and channel), in which N_1 tensors represent EEG segment containing epileptic spikes, denoted as $\{\boldsymbol{\mathcal{X}}_i^{\text{ep}}\}_{i=1}^{N_1}$, and N_2 tensors represent EEG segment containing non-epileptic spikes, $\{\boldsymbol{\mathcal{X}}_j^{\text{nep}}\}_{j=1}^{N_2}$. For feature extraction, our idea is first to estimate a feature space, $\mathcal{F}_{\text{ep}}$, which spans the class of EEG epileptic spikes, and then to project both types of spikes onto the resulting space to derive discriminant features. In such a case, the objective function can be expressed as

$$f_{\text{EEG}} = \sum_{i=1}^{N_1} \|\boldsymbol{\mathcal{X}}_i^{\text{ep}} - \boldsymbol{\mathcal{G}}_i^{\text{ep}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}\|^2, \tag{12}$$

over nonnegative projection matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and N_1 core tensors $\{\boldsymbol{\mathcal{G}}_i^{\text{ep}}\}_{i=1}^{N_1}$.

Inspired by the proposed SMLRAT method and a method proposed by Phan and Cichocki in [45], we minimize f_{EEG} by concatenating all three-way epileptic tensors $\{\boldsymbol{\mathcal{X}}_i^{\text{ep}}\}_{i=1}^{N_1}$ into a single four-way tensor $\widetilde{\boldsymbol{\mathcal{X}}}^{\text{ep}} \in \mathbb{R}_+^{I_1 \times I_2 \times I_3 \times N_1}$, and then perform NTD of $\widetilde{\boldsymbol{\mathcal{X}}}^{\text{ep}}$, as given by

$$\begin{aligned} \widetilde{\boldsymbol{\mathcal{X}}}^{\text{ep}} &= \boldsymbol{\mathcal{X}}_1^{\text{ep}} \boxplus \boldsymbol{\mathcal{X}}_2^{\text{ep}} \cdots \boxplus \boldsymbol{\mathcal{X}}_{N_1}^{\text{ep}} \\ &\overset{\text{NTD}}{=} \boldsymbol{\mathcal{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \times_4 \mathbf{D}, \end{aligned} \tag{13}$$

to obtain the factors $\mathbf{A} \in \mathbb{R}_+^{I_1 \times r_1}$, $\mathbf{B} \in \mathbb{R}_+^{I_2 \times r_2}$, $\mathbf{C} \in \mathbb{R}_+^{I_3 \times r_3}$ and $\mathbf{D} \in \mathbb{R}_+^{N_1 \times N_1}$, which respectively span the

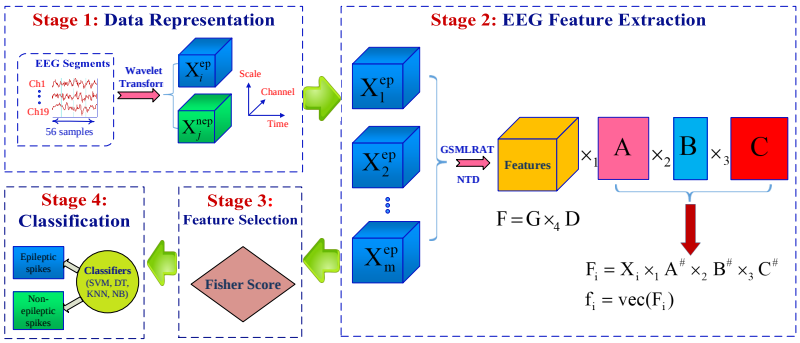


Figure 1: Proposed epileptic spike detection system.

spaces of parameters representing the domains of time, wavelet-scale, channel and epileptic spikes. Columns of $\mathbf{D}$ are considered as eigenspikes, the span of which forms the feature space $\mathcal{F}^{\text{ep}}$ of epileptic spikes.

Therefore, given any three-way tensor $\boldsymbol{\mathcal{X}}$ of some EEG data, its k-mode unfolding can be expressed by a linear combination of eigenspikes as

$$\underbrace{\mathbf{X}_{(k)}}_{\text{input data}} = \underbrace{\mathbf{D}}_{\text{basic vectors}} \underbrace{\mathbf{G}_{(k)}(\mathbf{C}\otimes\mathbf{B}\otimes\mathbf{A})^T}_{\text{coefficients}}, \tag{14}$$

where $\mathbf{G}_{(k)}$ is the k-mode unfolding of the core tensor $\boldsymbol{\mathcal{G}}$. The core $\boldsymbol{\mathcal{G}}$ and the factor $\mathbf{D}$ now carry part of information of the EEG data which resides in $\mathcal{F}^{\text{ep}}$, i.e.,

$$\mathcal{F}^{\text{ep}} = \boldsymbol{\mathcal{G}} \times_4 \mathbf{D}. \tag{15}$$

To investigate multi-domain features of EEG epileptic spikes, we can choose different basis functions. For examples, we can define an eigenspike time-basis $\boldsymbol{\mathcal{F}}_{\text{time}}$, by multiplying $\mathbf{A}$ with $\mathcal{F}^{\text{ep}}$ to obtain $\mathcal{F}^{\text{ep}}_{\text{time}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A} \times_4 \mathbf{D}$, to yield the principal axes of variations of an epileptic spike across channel and wavelet-scale modes. Similarly, we can derive the channel-basis and scale-basis of the eigenspike space, i.e, $\mathcal{F}^{\text{ep}}_{\text{scale}} = \boldsymbol{\mathcal{G}} \times_2 \mathbf{B} \times_4 \mathbf{D}$ and $\mathcal{F}^{\text{ep}}_{\text{channel}} = \boldsymbol{\mathcal{G}} \times_3 \mathbf{C} \times_4 \mathbf{D}$.

Given a training set of M tensors $\boldsymbol{\mathcal{X}}^{\text{train}}_m$, $m = 1, 2 \ldots, M$, of EEG data (including both epileptic and non-epileptic spikes), we form a discriminant feature vector $\mathbf{f}^{\text{train}}_m$ as follows:

$$\begin{aligned} \boldsymbol{\mathcal{F}}^{\text{train}}_m &= \boldsymbol{\mathcal{G}}^{\text{train}}_m \times_4 \mathbf{D} = \boldsymbol{\mathcal{X}}^{\text{train}}_m \times_1 \mathbf{A}^{\#} \times_2 \mathbf{B}^{\#} \times_3 \mathbf{C}^{\#}, \\ \mathbf{f}^{\text{train}}_m &= \text{vec}(\boldsymbol{\mathcal{F}}^{\text{train}}_m). \end{aligned} \tag{16}$$

Similarly, for any tensor $\boldsymbol{\mathcal{X}}^{\text{test}}$ in the testing set of EEG data, its features can be extracted by projecting the tensor onto $\mathcal{F}^{\text{ep}}$, i.e.,

$$\begin{aligned} \boldsymbol{\mathcal{F}}^{\text{test}} &= \boldsymbol{\mathcal{X}}^{\text{test}} \times_1 \mathbf{A}^{\#} \times_2 \mathbf{B}^{\#} \times_3 \mathbf{C}^{\#}, \\ \mathbf{f}^{\text{test}} &= \text{vec}(\boldsymbol{\mathcal{F}}^{\text{test}}). \end{aligned} \tag{17}$$

Remark: Our formulation of the concatenated tensor in (13) differs from [45] in which we do not integrate the *complete* set of training tensors of both epileptic and non-epileptic spikes, but use only tensors of epileptic spikes to compute the factors and hence the feature space. It stems from the following observations. EEG signals are composed of several components, including epileptic spikes (which are abnormal brain activity), EEG background (which includes normal brain activities) and artifacts (which are non-brain activities, e.g. eye or muscle movements). Since epileptic spikes are abnormal activity, they can be considered independent from the other activities. Hence, we assume that the other activities do not belong to the feature space of epileptic spikes. Moreover, it is difficult to describe non-epileptic activities present in the EEG data because we do not have knowledge of all these activities. Furthermore, the number of non-epileptic spikes is huge in EEG datasets, hence, the concatenation of a *complete* set of training tensors results in a very big four-way tensor (e.g. more than 10^9 entries). This leads to two issues: (i) decomposition of such four-way tensor is difficult and the resulting factors are not guaranteed to be optimal; (ii) the imbalance problem which has emerged as one of the challenges in data science [57] (e.g. the ratio of epileptic spike class to non-epileptic class is 1:260 in our EEG dataset). Therefore, we aim to capture a feature space that covers only epileptic spikes. Our method is related to the one-class classification (OCC) which aims to find a decision boundary around a specific class of interest, namely "positive" class, in machine learning [58, 59]. Accordingly, data of no interest form the "negative" class. The OCC problem may be harder than the conventional classification with data from two or multiple classes. Since the "negative" data samples (i.e., belonging to the negative class) in such a case are limited (i.e., activities of non-interest such as collected non-epileptic spikes can not cover the whole feature space for the negative class in our case), so only one side of the decision boundary can be estimated

definitively by using the collected data. Our method is, thus, consistent with one of three learning frameworks of OCC, as categorized in [58, 59]: learning with only positive examples, learning with positive examples and some amount of poorly distributed negative examples, and learning with positive and unlabeled data. On the contrary, Phan-Cichocki's method [45] was proposed to solve the problem of binary/multi-class classification, concatenating all training tensors derived from multiple classes.

4.2. Feature Selection

The aim of feature selection is to find a subset of input features, such that it can span the space of data of interest. An EEG dataset usually includes different components: brain activities of interest such as epileptic spikes, and activities without interest such as artifacts and noise. In addition, tensor decomposition may result in a huge number of the features; for example, NTD would give $r = r_1 r_2 r_3$ features. Thus, the expected outputs (e.g. detected epileptic spikes) may not be determined by a *complete* set of the resulting features, but depends only on a subset of relevant features. In this stage, we use the Fisher score [60] of each feature to assess the effectiveness of classification. Assume that we have extracted n features from NTD, i.e., $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$. Denote N_1 and N_2 the numbers of epileptic spikes and non-epileptic spikes, respectively. Let $\mu_{i,c}$ and $\sigma_{i,c}$ be the mean and standard deviation of the i-th feature for class Ω_c, $c \in \{1, 2\}$, μ_i and σ_i be the mean and the standard deviation of the i-th feature in the whole training dataset, m_c and Σ_c be the mean and covariance matrix of class Ω_c.

The objective is to find a linear combination $\mathbf{w}^T\mathbf{f}$ such that the best separation can be achieved. In particular, the Fisher discriminant ratio is determined by maximizing the ratio of between-class variation and within-class variation, i.e.,

$$f_{\text{Fisher}}(\mathbf{w}) = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}} = \frac{[\mathbf{w}(\mu_1 - \mu_2)]^2}{\mathbf{w}^T(\Sigma_1 + \Sigma_2)\mathbf{w}}. \quad (18)$$

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

$$\gamma(\mathbf{f}_i) \triangleq \mathbf{w}(i) = \frac{N_1(\mu_{i,1} - \mu_i)^2 + N_2(\mu_{i,2} - \mu_i)^2}{N_1\sigma^2_{i,1} + N_2\sigma^2_{i,2}}. \quad (19)$$

We select l significant features with top Fisher scores,

$$\mathbf{F}_{\text{Fisher}} = \{\mathbf{f}_{(1)}, \mathbf{f}_{(2)}, \ldots, \mathbf{f}_{(l)} | \mathbf{f}_{(i)} \in \mathbf{F}, i = 1, 2, \ldots, l\}.$$

4.3. Number of Components

In tensor decomposition, determining rank of a tensor (or number of components) is an important issue, and it is also an NP-hard problem. In the literature, several popular methods for this task was surveyed in [61], such as DIFFIT, CORCONDIA and ARD.

To determine the number of components when decomposing an EEG tensor, we apply the truncated HOSVD algorithm, which can provide an effective solution for the best rank-$(r_1, \ldots, r_n)$ tensor approximation [52]. This selection is motivated by an observation that the "meaningful" components of each factor is often related to the underlying signal of interest (e.g. EEG spikes) and thus may be different the true rank of the data tensor. For our three-way EEG tensor, the numbers of components $\{r_k\}_{k=1}^3$ in the factors $\{\mathbf{U}_k\}_{k=1}^3$ can be estimated from their corresponding modes $\{\mathbf{X}_{(k)}\}_{k=1}^3$ using the truncated SVD, as follows:

$$\mathbf{X}_{(k)} \approx \mathbf{U}_k^{I \times r_k} \mathbf{\Lambda}_k^{r_k \times r_k} \mathbf{V}_k^{r_k \times JK}, \quad k = 1, 2, 3. \quad (20)$$

In the above SVDs, each number of components (e.g. r_1) in each tensor mode of the EEG tensor can be obtained by selecting r_1 principal singular values of the mode such that the total variance is maximized, i.e.,

$$\text{VAR}_{r_1} = \frac{\sum_{i=1}^{r_1} \lambda_i}{\sum_{j=1}^{I} \lambda_j} 100\%. \quad (21)$$

5. Experimental Results and Discussions

5.1. EEG Dataset and EEG Tensor Construction

The EEG data used in this study were recorded by using the international standard 10-20 system with 19 channels and the sampling rate of 256 Hz. The measurements were carried out on 17 patients (including 11 males and 6 females) who were clinically diagnosed to have epilepsy, with durations varying from 5 to 28 minutes. Details of the dataset are given in Table 1. Figure 2 illustrates some epileptic spikes from this dataset. Epileptic spikes were manually identified by a neurologist from Vietnam National Children's Hospital.

To obtain EEG signals within the desired frequency band and restrain artifacts and noise as well as "negligible" spikes, as shown in Figure 2, the following pre-processing was implemented [9]. We first used a digital Butterworth low-pass filter with the cutoff frequency 70 Hz, a notch filter with the cutoff frequency of 50 Hz associated with a bandwidth of 2 Hz, and a high-pass filter with the cutoff frequency of 0.5 Hz. After that, we removed negligible spikes by using a threshold criteria and three training perceptrons. The remaining set of possible spikes included either real epileptic or "non-epileptic".

From the EEG dataset, we extracted 1442 epileptic spikes and more than 375429 non-epileptic

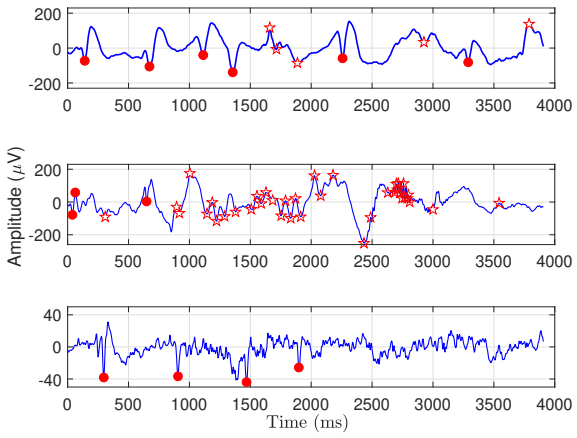


Figure 2: Some epileptic spikes (circle markers) and non-epileptic spikes (star markers) derived from three typical patients in our filtered EEG data.

spikes. Then, we constructed the corresponding tensors of the 19-channel EEG 56-point segments containing these spikes, with dimensions of time, wavelet-scale and channel, as follows. Denote Ω_1 and Ω_2 the classes of epileptic and non-epileptic tensors, respectively. Now, for each spike, an EEG data sample is first presented by a segment of 56 points around the location of a spike. After that, the continuous wavelet transform was used to obtain the time-frequency representation of the multi-channel EEG segments simultaneously. We enlarged the number of wavelet scales in the dominant range [4-8] to the size of 20, instead of 5 as used in [9]. Hence, we obtain 19 wavelet coefficient matrices of size 56×20 presenting EEG spectral features. Finally, we concatenate the 19 coefficient matrices into a tensor $\boldsymbol{\mathcal{X}} \in \mathbb{R}^{56\times 20\times 19}$ with three modes of time, wavelet-scale and channel.

Table 1: EEG Dataset

Pat.	Gen.	Age	Dur.	Spike
1	M	4	19m21s	8/15145
2	M	6	22m25s	635/20484
3	M	9	11m24s	6/14975
4	M	9	11m24s	16/30751
5	M	11	16m16s	351/25916
6	M	12	17m49s	22/44387
7	M	15	22m00s	2/2036
8	M	16	22m58s	11/29351
9	M	20	27m13s	1/3742
10	M	21	23m57s	8/2371
11	M	72	15m26s	2/1565
12	F	10	17m07s	3/53302
13	F	13	18m53s	5/69583
14	F	16	20m14s	8/6217
15	F	20	14m32s	324/11219
16	F	22	27m37s	28/23215
17	F	28	5m31s	12/21170

Pat. = Patient, Gen. = Gender (M=Male, F= Female), Dur. = Duration, Spike = Number of epileptic spikes / Number of non-epileptic spikes.

5.2. Performance Metrics

To assess epileptic spike detection performance, three statistical metrics including Sensitivity (SEN, a.k.a., Recall), Specificity (SPE) and Accuracy (ACC) are widely used to evaluate performance of detection systems, see [7–9,18,62,63] for examples. Furthermore, boxplot, receiver operator characteristic (ROC) and its area under the ROC curve (AUC) are also used to illustrate the performance of the systems. When we assess the effectiveness of the system on the EEG dataset using cross validation methods, we may obtain different values of these metrics across different tests/patients. Inspired by the results on evaluating the average performance of EEG interictal spike detection algorithms in [64], the overall performance with respect to the metric ρ (e.g. SEN) of our system can be averaged in the following ways:

(i) Arithmetic mean: $\rho_{\mathrm{AM}} = \dfrac{1}{T}\sum_{i=1}^{T}\rho_i,$

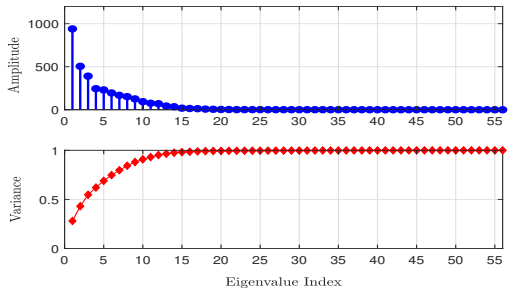
(ii) Time-weighted average: $\rho_{\mathrm{TWA}} = \dfrac{1}{\sum_{k=1}^{T} D_k}\sum_{i=1}^{T}\rho_i D_i,$

(iii) Total accuracy: $\rho_{\mathrm{TA}} = \dfrac{1}{\sum_{k=1}^{T} N_k}\sum_{i=1}^{T}\rho_i N_i$

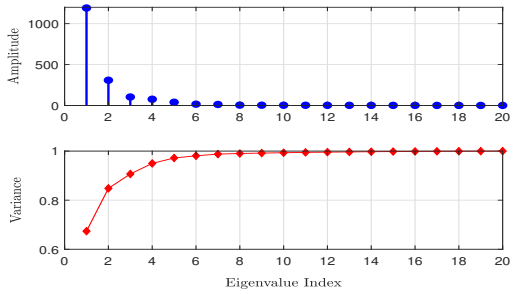
(iv) Time/event-weighting: $\rho_{\mathrm{TEW}} = \dfrac{1}{\sum_{k=1}^{T}\frac{D_k}{N_k}}\sum_{i=1}^{T}\rho_i\dfrac{D_i}{N_i},$

5.3. Experiment Setups and Results

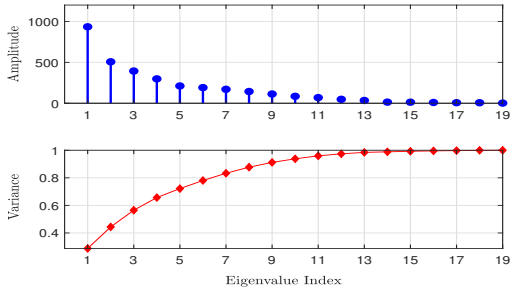
Our experiments are conducted to study the three stages, by: (i) performing feature extraction by estimating the eigenspikes and the corresponding features, (ii) performing feature selection to obtain the significant features for classification task, and (iii) performing classification by comparing the testing features with the training features using well-known classifiers. The EEG dataset is split into two groups, including a training set and a testing set using leave-one-out cross-validation (LOO-CV) method. In each test case, the classification model is fitted by using a training data composed of 16 patients and then is tested by a remaining patient. The evaluation is repeated until the last patient is done.



(a) Time mode



(b) Scale (freq.) mode



(c) Channel mode

Figure 3: Eigen-spectra of three modes of the epileptic tensor. For each mode, the first row (in blue) corresponds to the set of eigenvalues, the second row (in red) corresponds to their spectral variance.

5.3.1. Feature Extraction

The first task is to determine the multilinear rank (r_1, r_2, r_3) of EEG tensors. The spectra and total variances of three covariance matrices for epileptic tensor modes are illustrated in Figure 3. If we choose to have a significance level of 99%, which approximately corresponds to the sum of variances of the first 15 components in Figure 3(a), then we can have a good approximation for the time mode $\mathbf{A}$ by $\mathbf{A} \approx \sum_{i=1}^{15} \lambda_i \mathbf{u}_i \mathbf{v}_i^T$, where λ_i is the i-th singular-value associated with the right and left singular-vector, $\mathbf{u}_i$ and $\mathbf{v}_i$, of $\mathbf{A}$. In the same way, we also obtained 10 and 19 components for the frequency and spatial domains respectively, as shown in Figures 3(b) and 3(c).

By performing NTD of the training four-way epileptic tensor $\tilde{\boldsymbol{\mathcal{X}}}^{\text{ep}} \in \mathbb{R}_+^{56\times 20\times 19\times M}$, with M is the number of training three-way tensors, we obtained common factors $\mathbf{A} \in \mathbb{R}_+^{56\times 15}$, $\mathbf{B} \in \mathbb{R}_+^{20\times 10}$, $\mathbf{C} \in \mathbb{R}_+^{19\times 19}$. Similarly, we also obtain factors of time, scale and channel for the non-epileptic spike class. Comparison of the features between class Ω_1 and Ω_2 are shown in Figures 4 and 5, revealing some difference between the factors of epileptic tensors and non-epileptic tensors. In particular, considering first the factor $\mathbf{A}$, components of epileptic spikes were most localized in time; e.g. components #1, #2 and #3 were associated with the 30-th, 28-th and 3-th time sample, respectively. Meanwhile, the components of non-epileptic spikes seem to be spread, except from components #9, #11 and #12. Next, the factor $\mathbf{B}$ is shown in Figure 5. Since the behaviors of epileptic spikes and non-epileptic spikes are different, the resulting subspace of parameter representing wavelet-scale for class C_1 may not span non-epileptic spikes. Figure 6 shows common factor $\mathbf{C} \in \mathbb{R}_+^{19\times 19}$ of epileptic tensors obtained from NTD, showing that the factor of epileptic spikes was well localized in space, i.e., to specific regions on the head. Hence, it may lead to the ability of learning localized parts of epileptic spikes from the channel mode.

Next, we investigated the advantages of NTD over other types of tensor decomposition, e.g. unconstrained Tucker decomposition and nonnegative CP decomposition (NCP), for EEG epileptic spike analysis. The multilinear rank tensor used for this task is $r_1 = r_2 = r_3 = 15$. As shown in Figure 7, NTD and NCP can yield sparse basis vectors in $\mathbf{A}$ and $\mathbf{B}$ while this is not the case for Tucker decomposition (TD).

5.3.2. Feature Selection

To assess the effectiveness of the proposed feature selection stage for detecting EEG epileptic spikes, the extracted features are fed into the support vector machine (SVM) classifier. For illustration, we use the simplest SVM kernel, namely the linear kernel. In addition, we also use p-value [65] to provide the strength of ranked features derived by the Fisher score. A p-value in a statistical hypothesis test is that value of p, with $0 \leqslant p \leqslant 1$, such that given a significance level α, if $\alpha > p$ the test rejects the null hypothesis, H_0, otherwise the test does not reject H_0. By convention, α is commonly set to 0.05 [65]. The experimental results are shown in Figures 8, 9, and 10.

Figure 8 shows 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. Specifically, among the features there were the top 500 features having p-value close to 0, meaning that we can reject the null hypothesis H_0 completely. Hence, these 500 features have stronger discrimination

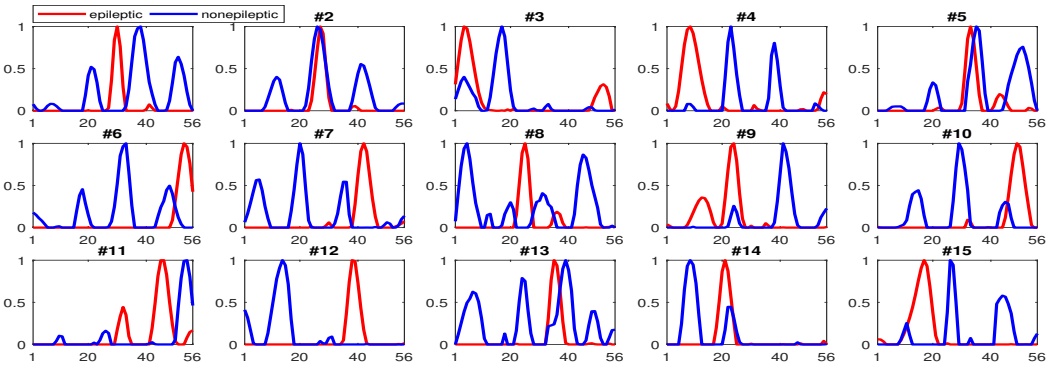
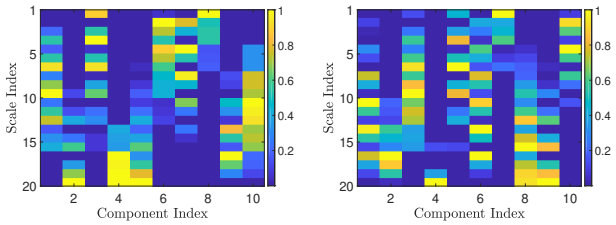


Figure 4: Common time factor $\mathbf{A} \in \mathbb{R}_{+}^{56\times 15}$ derived from NTD.



(a) Epileptic spike (b) Non-epileptic spike

Figure 5: Common scale 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].

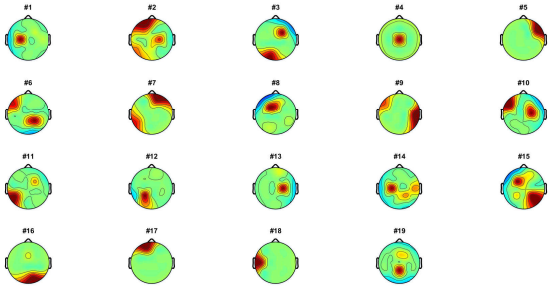
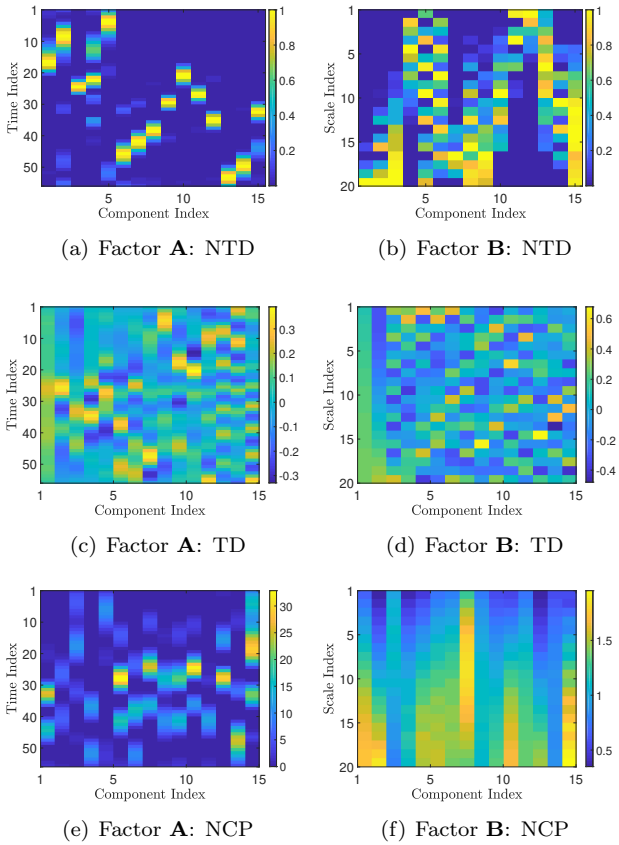


Figure 6: Common channel factor $\mathbf{C} \in \mathbb{R}_{+}^{19\times 19}$ of the epileptic tensor derived from NTD.



(a) Factor **A**: NTD (b) Factor **B**: NTD

(c) Factor **A**: TD (d) Factor **B**: TD

(e) Factor **A**: NCP (f) Factor **B**: NCP

Figure 7: A comparison of obtained loading factors between using three different tensor decompositions (NTD, TD, NCP) of the epileptic tensor. The x-axis denotes the number of components (column vectors), while the y-axis presents 56 time samples.

power than others. Furthermore, these features are different to that of non-epileptic class Ω_2, as illustrated in Figure 9. Thus, they are efficient in detecting epileptic spikes. Performance of the SVM model using the first 500 significant features achieved the top SEN at approximately 0.9 and overall ACC around 0.92 with the first 200 features (Figure 10), while the corresponding values of the area under ROC curves (AUC) were always higher than 0.9, thus ranked as excellent result of classification‡.

‡ 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.

5.3.3. Classification

To investigate how effective concatenation of input tensors, we compare the proposed method against Phan-Cichocki method in [45]. Second, we use other tensor-based approaches which were successfully

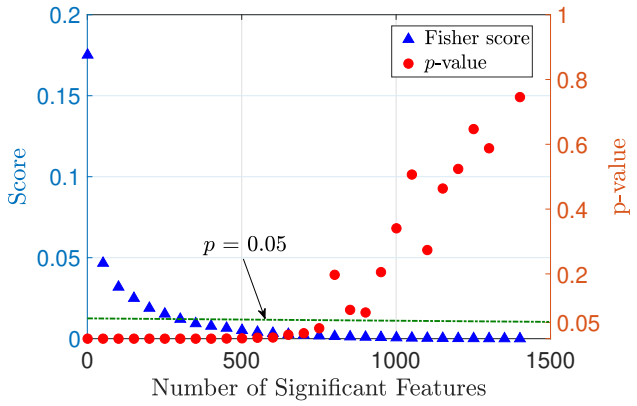


Figure 8: Fisher scores (blue marks) and p-values (red mark) of features. Features are ordered based on their Fisher score. The significant level $p = 0.05$ is to make decision for rejecting H_0, removing features with $p > 0.05$.

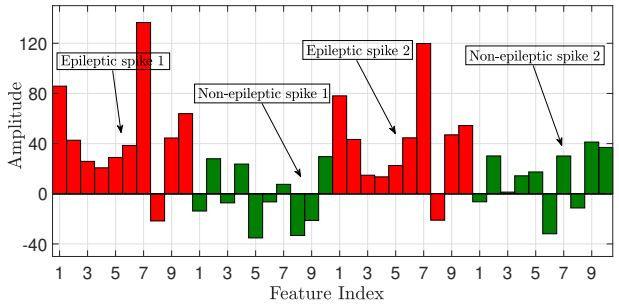


Figure 9: Top 10 selected features of two typical epileptic spikes and two non-epileptic spikes. Behavior of features derived from epileptic spikes are similar, unlike non-epileptic spikes.

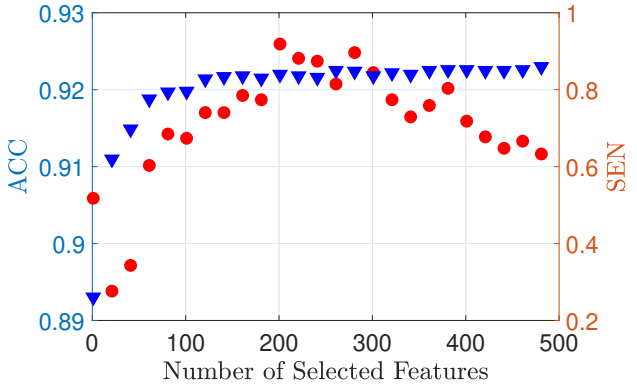
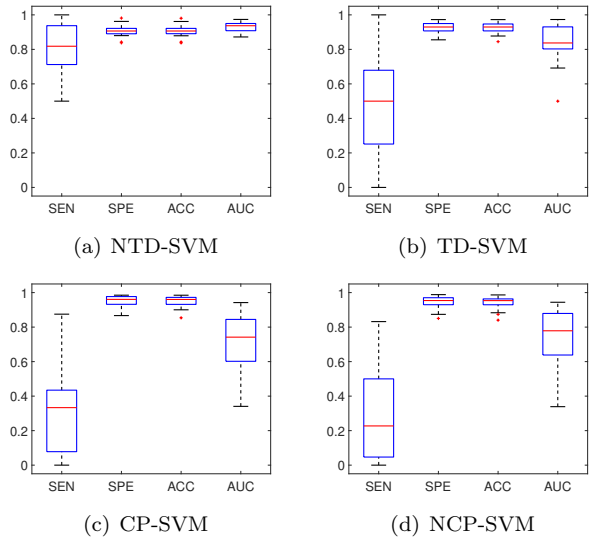


Figure 10: Classification performance vs. number of selected features

applied to detect brain activities of interest in EEG signals as comparative methods, including Tucker-based [15, 18], CP-based [14, 16], NCP-based [66].

To evaluate the separability of the extracted features, we applied three widely used classification models in the platform WEKA [67] to classify EEG epileptic spikes out of non-epileptic spikes, including k-nearest neighbors (KNN), naive Bayes (NB), decision tree (DT) and SVM. Parameters of the classifiers were set by default. In particular, the distance metric used in the KNN was the Euclidean distance and the size of the neighborhood was automatically obtained by setting the cross-validation option. For NB, we selected the Gaussian distribution as predictor distribution to compute the posterior probability for the two classes and then made decision for the class with higher probability. For DT, the standard CART algorithm was selected as the predictor selection technique, the tree depth equaled the size of training set and each node in the training tree had 10 observations. Meanwhile, we used the linear SVM kernel, similar to the previous task. We report here results of the SVM model trained with our features across 17 patients using the LOOCV method. The detailed results of other classifiers can be found in the technical report of [46], due to space limit.



(a) NTD-SVM (b) TD-SVM

(c) CP-SVM (d) NCP-SVM

Figure 11: Detection performance of SVM when using our features against different tensor-based approaches (CP, NCP, TD and NTD).

Feature extraction is key for EEG epileptic spike detection, and our approach outperforms the baselines on all evaluation metrics. First, the NTD-SVM model yielded strong results in terms of all measurements (SEN, SPE, ACC and AUC) (see Table 2). Table 2 presents a quantitative statistic of EEG epileptic spike detection performance of our NTD-SVM model using

Table 2: Detection performance of the NTD-SVM model using Leave-One-Out Cross-Validation (LOOCV)

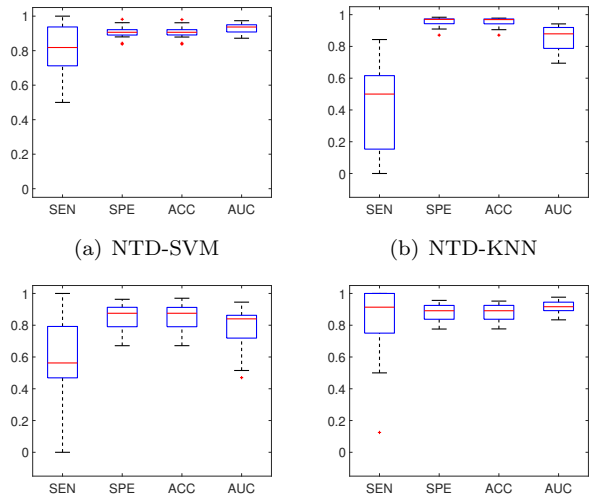
Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	7	1	13546	1599	0.8750	0.8944	0.8944	0.9435
2	635	20484	525	110	18639	1845	0.8268	0.9099	0.9074	0.9267
3	6	14975	6	0	13314	1661	**1.0000**	0.8891	0.8891	0.9488
4	16	30751	15	1	29587	1164	0.9375	0.9621	0.9617	0.9370
5	351	25916	329	22	23088	2828	0.9373	0.8909	0.8915	0.9633
6	22	44387	16	6	40967	3420	0.7273	0.9230	0.9229	0.9191
7	2	2036	2	0	1791	245	**1.0000**	0.8797	0.8789	0.9536
8	11	29351	9	2	26437	2914	0.8182	0.9007	0.9007	0.8822
9	1	3742	1	0	3447	259	**1.0000**	0.9212	0.9212	0.9546
10	8	2371	5	3	2327	44	0.6250	0.9814	0.9802	0.9036
11	2	1565	1	1	1407	158	**0.5000**	0.8990	0.8985	0.9476
12	3	53302	2	1	48318	4984	0.6667	0.9065	0.9065	0.9098
13	5	69583	4	1	66534	3049	0.8000	0.9562	0.9562	0.9736
14	8	6217	5	3	5691	526	0.6250	0.9154	0.9150	0.9406
15	324	11219	271	53	9401	1818	0.8364	0.8380	0.8379	0.8976
16	28	35495	21	7	19563	3652	0.7500	0.8427	0.8426	0.8720
17	12	21170	9	3	19262	1908	0.7500	0.9099	0.9098	0.9212

Average Performance:	Mean ± S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho}$ ± S.D.
SEN :	0.8044 ± 0.1468	**0.8044**	0.8516	0.8042	0.8593	**0.8299 ± 0.0297**
SPE :	0.9071 ± 0.0371	0.9071	0.9145	0.9066	0.9137	0.9105 ± 0.0042
ACC :	0.9067 ± 0.0369	0.9067	0.9142	0.9062	0.9118	0.9097 ± 0.0039
AUC :	0.9291 ± 0.0287	0.9291	0.9291	0.9272	0.9436	**0.9323 ± 0.0076**

the LOOCV method. Accordingly, the percentage of correctly detected epileptic spikes (i.e., SEN) varied from patient to patient. In particular, the highest SEN achieved 100% in the cases of patients with a few of epileptic spikes (e.g. the 3-rd, 7-th and 9-th patient), while the worst case with SEN of 0.5 was from the 11-th patient. The NTD-SVM model achieved over 80% SEN in 10 out of 17 patients. In spite of the variation, the average metrics for SEN were still good, e.g. the arithmetic mean $\text{SEN}_{\text{AM}} = 0.8044$, $\text{SEN}_{\text{TA}} = 0.8516$, and $\overline{\text{SEN}} = 0.8299 \pm 0.0297$. The metrics to the non-epileptic class, including SPE and ACC were all high with small standard deviations (i.e., $\bar{\rho} \geqslant 0.9$ and S. D. ≈ 0.04). The key metric AUC to measure of separability of the classifier was also excellent, i.e., NTD-SVM obtained over 90% AUC in 14 out of 17 patients and the mean $\overline{\text{AUC}} = 0.9323 \pm 0.0076$ on average. These results indicate that the features extracted by our method are able to detect epileptic spikes with good performance.

Our detection system outperforms the three other tensor-based approaches, including CP, NCP and unconstrained TD decompositions (see Figure 11 and 13(a)). Figure 11 illustrates a number of boxplots to demonstrate the performance improvement of our system over others. Each boxplot for a specific metric (e.g. SEN) presents the distribution of evaluation performance across 17 patients in our EEG dataset. A box is based on the five summary numbers, including the "minimum", first quartile (Q1), second quartile, third quartile (Q3) and the "maximum". For instance, across the 17 patients, the highest median SEN achieved 0.8182 from the NTD-SVM model, while the value was low (i.e., $\leqslant 0.5$) when using other tensor decompositions. In addition, the interquartile range (i.e., IQR = Q3 − Q1) measuring the variability of the NTD-SVM were lower than that of TD-SVM, CP-SVM and NCP-SVM for each evaluation metric. The results were also verified by Figure 13(a) that shows ROC curves to illustrate overall performance of the four models. The ROC curve is drawn by plotting the true positive rate (TPR equivalent to SEN) and false positive rate that can be computed as 1 − SPE. Thus, the ROC curve allows us to derive a cost/benefit analysis for making decision. We can observe from the two figures that the NTD-based feature extraction provided a better classification accuracy than the CP decomposition (i.e., unconstrained CP and NCP decomposition) and unconstrained Tucker decomposition based approaches in this work. According to the Table 3, the average AUC of the CP-based and NCP-based models were always lower than 0.9. That means there were less than 90% chance that the models will be able to distinguish between epileptic spikes and non-epileptic spikes. The worst result was from the NCP-NB model which had much less discrimination capacity to detect EEG epileptic spikes, i.e., AUC = 0.574 ± 0.25. The results of TD-based models were similar to that of CP-based models. Although TD-SVM might provide a good performance in terms of AUC (i.e., 0.836 ± 0.113), the resulting SEN was not good enough, around 0.5. Hence, a half of the total number of epileptic spikes were detected incorrectly and labeled as non-epileptic activities. Meanwhile, the NTD-based models yielded a 10% to 30% better performance than that of other tensor decompositions. The two best overall accuracy belonged to the NTD-

based models, including NTD-SVM and NTD-DT (i.e., AUC $= 0.929 \pm 0.029$ and 0.914 ± 0.041 respectively, while SEN $\geqslant 0.8$ in both cases). Furthermore, the NTD-based models also detected non-epileptic spikes successfully, which more than 95% activities of non-interest were rejected correctly by the NTD-KNN model. The percentage was 90% when using the NTD-SVM model. The experiments shows that the NTD-based feature extraction can provide good features to enhance the separation between epileptic spikes and non-epileptic spikes.

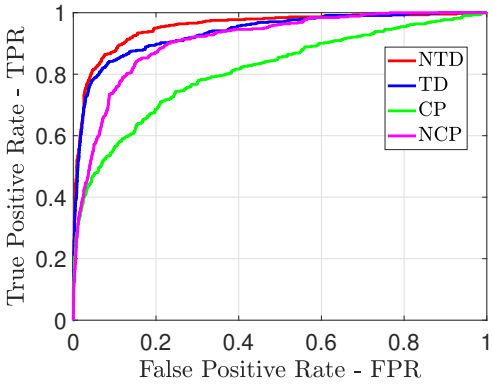


(a) NTD-SVM (b) NTD-KNN

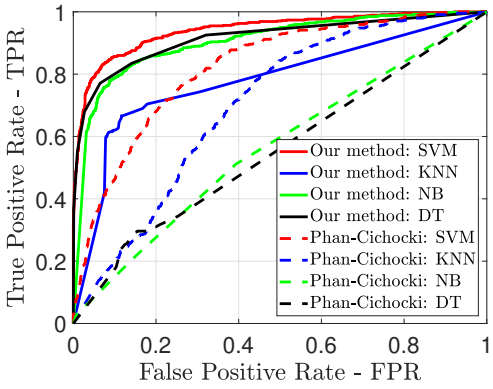
(c) NTD-NB (d) NTD-DT

Figure 12: Detection performance of four classifiers using our features.

Our NTD-SVM outperforms three widely used classifiers (i.e., KNN, NB and DT) in the classification task. The performance comparison between using difference machine learning models is shown statistically in the Table 3 and Figure 12. The Table 3 shows the overall performance of 16 models in terms of all evaluation metrics. SVM-based models performed better than others both in cases using features extracted from different tensor decompositions. As mentioned above, the two average area under ROC curves of NTD-SVM are AUC $= 0.929 \pm 0.029$ and $\overline{\text{AUC}} = 0.932 \pm 0.008$ in terms of arithmetic mean and overall mean respectively. The values were higher than that of NTD-KNN (e.g. $\overline{\text{AUC}} = 0.856 \pm 0.002$), NTD-NB (e.g. $\overline{\text{AUC}} = 0.756 \pm 0.070$) and NTD-DT (e.g. $\overline{\text{AUC}} = 0.911 \pm 0.006$). The number of correctly detected epileptic spikes of NTD-SVM (i.e., $\overline{\text{SEN}} = 0.830 \pm 0.003$) was also higher than that of KNN and NB (i.e., 0.402 ± 0.213 and 0.539 ± 0.212). Moreover, the results were verified by boxplots across 17 epileptic patients, as shown in Figure 12. Results from TD-based feature extraction also indicated that the SVM model took more advance of tensor decompositions than the three classifiers. The $\overline{\text{AUC}}$ of TD-SVM was 0.836 ± 0.113 compared to 0.524 ± 0.229, 0.702 ± 0.243 and 0.612 ± 0.171 of TD-KNN, TD-NB and TD-DT respectively. In spite of that, the average SEN of the four models using the features were not good enough. Specifically, neither one of the TD-based models could detect more than 50% of total epileptic spikes in our EEG dataset (i.e., SEN < 0.5, see the second column of Table 3). In the cases of using features extracted from two types of CP decompositions, the detection performances were bad, except the DT classifier. However, the resulting AUC of the DT classifier (i.e., < 0.9) were not good enough compared to that of the NTD-SVM where four evaluation metrics for AUC of NTD were all higher 0.9.



(a) Tensor-based epileptic spike detection methods



(b) Concatenation of input tensors

Figure 13: Performance comparison in terms of averaged ROC.

The concatenation of training input tensors is key for EEG epileptic spike detection. Accordingly, our method provided a better performance than Phan-Cichocki method. Table 4 and Figure 13(b) present a performance comparison of epileptic spike detection between using our method and Phan-Cichocki method.

Table 3: Detection performance comparison between using difference machine learning models

Method	SEN	SPE	ACC	AUC	$\overline{\text{SEN}}$	$\overline{\text{SPE}}$	$\overline{\text{ACC}}$	$\overline{\text{AUC}}$
CP-KNN	0.106 ± 0.132	0.966 ± 0.027	0.963 ± 0.028	0.690 ± 0.124	0.126 ± 0.107	0.966 ± 0.003	0.966 ± 0.004	0.682 ± 0.012
CP-NB	0.546 ± 0.275	0.710 ± 0.138	0.709 ± 0.137	0.662 ± 0.216	0.627 ± 0.085	0.709 ± 0.003	0.713 ± 0.020	0.693 ± 0.050
CP-DT	0.826 ± 0.246	0.829 ± 0.046	0.829 ± 0.045	0.857 ± 0.112	0.831 ± 0.057	0.829 ± 0.008	0.826 ± 0.005	0.856 ± 0.001
CP-SVM	0.311 ± 0.270	0.950 ± 0.034	0.948 ± 0.035	0.742 ± 0.155	0.341 ± 0.142	0.952 ± 0.005	0.952 ± 0.005	0.728 ± 0.008
NCP-KNN	0.162 ± 0.194	0.955 ± 0.031	0.951 ± 0.030	0.733 ± 0.150	0.189 ± 0.118	0.956 ± 0.005	0.952 ± 0.003	0.726 ± 0.011
NCP-NB	$\mathbf{0.366 \pm 0.358}$	0.742 ± 0.147	0.741 ± 0.148	$\mathbf{0.574 \pm 0.250}$	0.472 ± 0.111	0.736 ± 0.011	0.733 ± 0.012	$\mathbf{0.587 \pm 0.023}$
NCP-DT	0.871 ± 0.149	0.835 ± 0.052	0.834 ± 0.051	0.888 ± 0.049	0.850 ± 0.042	0.834 ± 0.004	0.834 ± 0.004	0.892 ± 0.007
NCP-SVM	0.288 ± 0.255	0.941 ± 0.040	0.939 ± 0.040	0.734 ± 0.188	0.324 ± 0.203	0.941 ± 0.007	0.940 ± 0.006	0.695 ± 0.074
TD-KNN	$\mathbf{0.098 \pm 0.135}$	0.984 ± 0.030	0.980 ± 0.031	$\mathbf{0.524 \pm 0.229}$	0.081 ± 0.029	0.985 ± 0.003	0.983 ± 0.005	$\mathbf{0.476 \pm 0.090}$
TD-NB	0.333 ± 0.286	0.856 ± 0.160	0.857 ± 0.161	0.702 ± 0.243	0.278 ± 0.062	0.865 ± 0.008	0.872 ± 0.024	0.667 ± 0.062
TD-DT	0.240 ± 0.207	0.897 ± 0.042	0.894 ± 0.043	0.612 ± 0.171	0.211 ± 0.077	0.898 ± 0.004	0.900 ± 0.007	0.621 ± 0.014
TD-SVM	0.490 ± 0.281	0.929 ± 0.034	0.927 ± 0.035	$\mathbf{0.836 \pm 0.113}$	0.470 ± 0.115	0.903 ± 0.062	0.903 ± 0.063	$\mathbf{0.811 \pm 0.048}$
NTD-KNN	0.404 ± 0.274	0.957 ± 0.023	0.956 ± 0.030	0.855 ± 0.079	0.402 ± 0.213	0.958 ± 0.001	0.958 ± 0.005	0.856 ± 0.002
NTD-NB	0.560 ± 0.313	0.850 ± 0.087	0.851 ± 0.087	0.794 ± 0.141	0.539 ± 0.212	0.851 ± 0.008	0.854 ± 0.005	0.756 ± 0.070
NTD-DT	0.826 ± 0.247	0.877 ± 0.061	0.877 ± 0.060	0.914 ± 0.041	0.847 ± 0.015	0.879 ± 0.003	0.877 ± 0.001	0.911 ± 0.006
NTD-SVM	$\mathbf{0.804 \pm 0.147}$	$\mathbf{0.907 \pm 0.037}$	$\mathbf{0.907 \pm 0.037}$	$\mathbf{0.929 \pm 0.029}$	$\mathbf{0.830 \pm 0.030}$	$\mathbf{0.910 \pm 0.004}$	$\mathbf{0.910 \pm 0.004}$	$\mathbf{0.932 \pm 0.008}$

Results expressed as Mean ± S.D.

Table 4: Concatenation of input tensors, with SVM, KNN, NB and DT using first 500 significant features.

	Our method				Phan-Cichocki method			
Metric	SVM	NB	KNN	DT	SVM	NB	KNN	DT
SEN	$\mathbf{0.830 \pm 0.030}$	0.402 ± 0.274	0.539 ± 0.213	$\mathbf{0.847 \pm 0.015}$	$\mathbf{0.379 \pm 0.047}$	$\mathbf{0.346 \pm 0.060}$	$\mathbf{0.092 \pm 0.036}$	$\mathbf{0.217 \pm 0.022}$
SPE	0.911 ± 0.004	$\mathbf{0.958 \pm 0.001}$	0.851 ± 0.010	0.878 ± 0.004	0.931 ± 0.013	0.779 ± 0.012	0.954 ± 0.008	0.866 ± 0.006
ACC	0.910 ± 0.004	$\mathbf{0.958 \pm 0.005}$	0.857 ± 0.006	0.877 ± 0.003	0.927 ± 0.012	0.783 ± 0.006	0.951 ± 0.006	0.865 ± 0.003
AUC	$\mathbf{0.932 \pm 0.008}$	0.856 ± 0.002	0.756 ± 0.070	$\mathbf{0.911 \pm 0.006}$	0.817 ± 0.011	0.590 ± 0.082	0.622 ± 0.028	0.521 ± 0.003

We note that, according to Phan-Cichocki method, the *complete* set of training tensors was used to concatenate a single four-way tensor. However, the number of non-epileptic spikes is very huge in our EEG dataset (i.e., more than 375000 spikes). Therefore, taking NTD decomposition of the resulting four-way tensor $\tilde{\boldsymbol{\mathcal{X}}}_{\text{train}}$ may be difficult, while the decomposed factors were not guaranteed to be optimal, because of the very big tensor (i.e., the number of entries in $\tilde{\boldsymbol{\mathcal{X}}}_{\text{train}}$ is more than 7.10^9 for each testing case using LOOCV). This could be a weakness of Phan-Cichocki method in this work. For the ease of implementation as well as avoiding the imbalanced problem, we applied the random under-sampling technique for the non-epileptic spike class to balance two class distributions, which is a widely used technique to handle imbalance dataset [57]. As a result, around 6000 non-spikes were selected to form the training four-way tensor in our experimental setup. The results showed that evaluation metrics measuring the four classifiers using our method were higher than that of Phan-Cichocki method. Specifically, the our method obtained the best classification accuracy, i.e., $\overline{\text{AUC}}$ of 0.932 ± 0.008, achieved the highest $\overline{\text{ACC}} = 0.910 \pm 0.004$ and the highest $\overline{\text{SEN}} = 0.830 \pm 0.030$. The separability of our features was also validated by applying the classifiers KNN, NB and DT. In contrast to our method, both classifiers using features extracted by Phan-Cichocki method did not work well. In all test cases, the average $\overline{\text{SEN}}$ across 17 patients of four classifiers were low, (i.e., $\overline{\text{SEN}} < 0.4$). That means more than 60% epileptic spikes in our EEG dataset could not be detected by these classifiers. Our NTD-SVM and NTD-DT models provided much better performance in terms of SEN in which they detected more than 80% the number of epileptic spike correctly. The metrics to non-epileptic class (i.e., SPE and ACC) of both four classifiers were also effective, e.g. the overall SPE of SVM and KNN were 0.931 ± 0.013 and 0.954 ± 0.008 respectively. However, three of four classifiers resulted in a poor AUC on average (i.e., $0.5 \leqslant \overline{\text{AUC}} \leqslant 0.6$) which indicates that these models failed to detect EEG epileptic spikes, except the SVM classifier. However, the AUC of SVM was lower 11% than that of our method. We refer the reader to the technical report [46] for further detailed results of the four classifiers using features extracted by Phan-Cichocki method.

6. Conclusions

In this paper, we have proposed a new model, abbreviated as GSMLRAT, for solving the problem of simultaneous multilinear low-rank approximation of tensors. Inspired by the advantages of GSMLRAT and NTD, we proposed, for the first time, a new tensor-based system to detect epileptic spikes in EEG data. We have first derived a new feature space that can span EEG epileptic spikes from sparse loading factors of NTD. A new discriminant set of features, learned from NTD, which can distinguish between epileptic spike class and non-epileptic spike class with high accuracy. To reduce feature dimensionality as well as to achieve

the good separability between these classes, we have applied the Fisher score in EEG feature selection. The numerical experiments have indicated that EEG multi-way analysis using NTD allows us to extract multi-domain features of epileptic spikes and provide high classification accuracy only with well-known "shallow" classifiers such as KNN, NB, DT and SVM.

Acknowledgments

This research was funded by Vietnam National Foundation for Science and Technology Development (NAFOSTED) under Grant No. 102.04-2019.14.

References

[1] Berg A T and Panayiotopoulos C P 2010 *Atlas of Epilepsies* 1st ed (Springer-Verlag London)
[2] Gotman J 1982 *Electroencephalography and Clinical Neurophysiology* **54** 530–540
[3] Gotman J 1999 *Journal of Clinical Neurophysiology* **16** 130–140
[4] Tzallas A T, Tsipouras M G, Tsalikakis D G, Karvounis E C, Astrakas L, Konitsiotis S and Tzaphlidou M 2012 Automated epileptic seizure detection methods: A review study *Epilepsy: Histological, Electroencephalographic and Psychological Aspects* ed Stevanovic D chap 4, pp 75–98
[5] Orosco L 2013 *Journal of Medical and Biological Engineering* **33** 526–537
[6] Acharya U R, Sree S V, Swapna G, Martis R J and Suri J S 2013 *Knowledge-Based Systems* **45** 147–165
[7] Liu H S, Zhang T and Yang F S 2002 *IEEE Transactions on Biomedical Engineering* **49** 1557–1566
[8] Acır N, Öztura İ, Baklan B and Güzeliş C 2005 *IEEE Transactions on Biomedical Engineering* **52** 30–40
[9] Anh-Dao N T, Linh-Trung N, Nguyen L V, Tran-Duc T, Anh N T H and Boashash B 2018 *REV Journal on Electronics and Communications* **8** 1–13
[10] Indiradevi K, Elias E, Sathidevi P, Dinesh Nayak S and Radhakrishnan K 2008 *Computers in Biology and Medicine* **38** 805–816
[11] Acar E, Aykut-Bingol C, Bingol H, Bro R and Yener B 2007 *Bioinformatics* **23** i10–i18
[12] De Vos M, Vergult A, De Lathauwer L, De Clercq W, Van Huffel S, Dupont P, Palmini A and Van Paesschen W 2007 *NeuroImage* **37** 844–854
[13] Deburchgraeve W, Cherian P J, De Vos M, Swarte R M, Blok J H, Visser G H, Govaert P and Van Huffel S 2009 *Clinical Neurophysiology* **120** 1787–1796
[14] Ontivero-Ortega M, Garcia-Puente Y and Martínez-Montes E 2015 Comparison of classifiers to detect epileptic seizures via PARAFAC decomposition *VI Latin American Congress on Biomedical Engineering, 29-31 October, 2014, Parana, Argentina. IFMBE Proceedings.* vol 49 (Springer) pp 500–503
[15] Spyrou L, Kouchaki S and Sanei S 2015 Multiview classification of brain data through tensor factorisation *International Workshop on Machine Learning for Signal Processing* (IEEE) pp 1–6
[16] Aldana Y R, Hunyadi B, Reyes E J M, Rodriguez V R and Huffel S V 2018 *IEEE Journal of Biomedical and Health Informatics* 1–12
[17] Cong F, Lin Q H, Kuang L D, Gong X F, Astikainen P and Ristaniemi T 2015 *Journal of neuroscience methods* **248** 59–69
[18] Pippa E, Kanas V G, Zacharaki E I, Tsirka V, Koutroumanidis M and Megalooikonomou V 2016 *International Journal of Monitoring and Surveillance Technologies Research* **4** 1–15
[19] Hunyadi B, Dupont P, Van Paesschen W and Van Huffel S 2017 *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **7** e1197
[20] Markovsky I 2012 *Low Rank Approximation* (Springer)
[21] Cichocki A, Lee N, Oseledets I, Phan A H, Zhao Q, Mandic D P *et al.* 2016 *Foundations and Trends® in Machine Learning* **9** 249–429
[22] Kishore Kumar N and Schneider J 2017 *Linear and Multilinear Algebra* **65** 2212–2244
[23] Ding C and Ye J 2005 2-dimensional singular value decomposition for 2D maps and images *SIAM International Conference on Data Mining* pp 32–43
[24] Inoue K and Urahama K 2005 DSVD: A tensor-based image compression and recognition method *IEEE International Symposium on Circuits and Systems* (IEEE) pp 6308–6311
[25] Yang J, Zhang D, Frangi A F and yu Yang J 2004 *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26** 131–137
[26] Zhang D and Zhou Z H 2005 *Neurocomputing* **69** 224–231
[27] Crainiceanu C M, Caffo B S, Luo S, Zipunnikov V M and Punjabi N M 2011 *Journal of the American Statistical Association* **106** 775–790
[28] Ye J 2005 *Machine Learning* **61** 167–191
[29] Liu J and Chen S 2006 *Pattern Recognition Letters* **27** 1002–1008
[30] Liu J, Chen S, Zhou Z and Tan X 2010 *IEEE Transactions on Neural Networks* **21** 621–632
[31] Inoue K, Hara K and Urahama K 2012 Symmetric generalized low rank approximations of matrices *IEEE International Conference on Acoustics, Speech and Signal Processing* pp 949–952
[32] Jiarong, Wei Y and Shi Z X 2015 *PLOS ONE* **10** 1–23
[33] Ye J, Janardan R and Li Q 2005 Two-dimensional linear discriminant analysis *Advances in Neural Information Processing Systems* pp 1569–1576
[34] Van Deun K, Smilde A K, van der Werf M J, Kiers H A and Van Mechelen I 2009 *Bmc Bioinformatics* **10** 246
[35] Stegeman A 2018 *Psychometrika* **83** 21–47 ISSN 1860-0980
[36] Li X, Ng M K, Cong G, Ye Y and Wu Q 2017 *IEEE Transactions on Neural Networks and Learning Systems* **28** 1787–1800
[37] Xu D, Yan S, Zhang L, Lin S, Zhang H J and Huang T S 2008 *IEEE Transactions on Circuits and Systems for Video Technology* **18** 36–47
[38] Yan S, Xu D, Yang Q, Zhang L, Tang X and Zhang H 2007 *IEEE Transactions on Image Processing* **16** 212–220
[39] Lu H, Plataniotis K N and Venetsanopoulos A N 2008 *IEEE Transactions on Neural Networks* **19** 18–39 ISSN 1045-9227
[40] Panagakis Y, Kotropoulos C and Arce G R 2010 *IEEE Transactions on Audio, Speech, and Language Processing* **18** 576–588
[41] Lai Z, Xu Y, Chen Q, Yang J and Zhang D 2014 *IEEE Transactions on Neural Networks and Learning Systems* **25** 1942–1950
[42] Sorber L, Van Barel M and De Lathauwer L 2015 *IEEE Journal of Selected Topics in Signal Processing* **9** 586–600
[43] Anh-Dao N T, Thanh L T, Linh-Trung N and Le H V 2018 Nonnegative tensor decomposition for EEG epileptic spike detection *NAFOSTED Conference on Information and Computer Science (NICS)* (IEEE)
[44] Cichocki A, Zdunek R, Phan A H and ichi Amari S 2009 *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and*

blind source separation (John Wiley & Sons)
[45] Phan A H and Cichocki A 2010 *Nonlinear theory and its applications, IEICE* **1** 37–68
[46] Thanh L T, Dao N T A, Dung N V, Trung N L and Abed-Meraim K 2019 Simultaneous tensor decomposition for EEG epileptic spike detection Tech. Rep. UET-AVITECH-2019002 University of Engineering and Technology, Vietnam National University, Hanoi Vietnam
[47] Boussé M, Vervliet N, Domanov I, Debals O and De Lathauwer L *Numerical Linear Algebra with Applications* **25** e2190
[48] Boussé M, Goovaerts G, Vervliet N, Debals O, Huffel S V and Lathauwer L D 2017 Irregular heartbeat classification using kronecker product equations *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* pp 438–441
[49] Cong F, Phan A H, Astikainen P, Zhao Q, Wu Q, Hietanen J K, Ristaniemi T and Cichocki A 2013 *International Journal of Neural Systems* **23**
[50] Kolda T G and Bader B W 2009 *SIAM Review* **51** 455–500
[51] Harshman R A 1970 *UCLA Working Papers in Phonetics* **16**
[52] De Lathauwer L, De Moor B and Vandewalle J 2000 *SIAM Journal on Matrix Analysis and Applications* **21** 1324–1342
[53] Sheehan B N and Saad Y 2007 Higher order orthogonal iteration of tensors (HOOI) and its relation to PCA and GLRAM *SIAM International Conference on Data Mining* (SIAM) pp 355–365
[54] Jeon I, Papalexakis E E, Kang U and Faloutsos C 2015 HaTen2: Billion-scale tensor decompositions *2015 IEEE 31st International Conference on Data Engineering* pp 1047–1058
[55] Austin W, Ballard G and Kolda T G 2016 Parallel tensor compression for large-scale scientific data *2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS)* pp 912–922
[56] Oh J, Shin K, Papalexakis E E, Faloutsos C and Yu H 2017 S-HOT: Scalable High-Order Tucker Decomposition *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (ACM) pp 761–770
[57] Galar M, Fernandez A, Barrenechea E, Bustince H and Herrera F 2012 *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **42** 463–484
[58] Tax D 2001 *PhD thesis, Delft University of Technology*
[59] Khan S S and Madden M G 2014 *The Knowledge Engineering Review* 345–374
[60] Duda R O, Hart P E and Stork D G 2012 *Pattern Classification* (John Wiley & Sons)
[61] Cichocki A, Mandic D, Lathauwer L D, Zhou G, Zhao Q, Caiafa C and Phan H A 2015 *IEEE Signal Processing Magazine* **32** 145–163
[62] Liu Y C, Lin C C, Tsai J J and Sun Y N 2013 *Sensors* **13** 12536–12547
[63] Xuyen L T, Thanh L T, Viet D V, Long T Q, Trung N L and Thuan N D 2018 *VNU Journal of Science: Computer Science and Communication Engineering* **33**
[64] Casson A J, Luna E and Rodriguez-Villegas E 2009 *Journal of neuroscience methods* **177** 479–487
[65] Van Belle G, Fisher L D, Heagerty P J and Lumley T 2004 *Biostatistics: A Methodology for the Health Sciences* vol 519 (John Wiley & Sons)
[66] Wang D, Zhu Y, Ristaniemi T and Cong F 2018 *Journal of Neuroscience Methods* **308** 240–247
[67] Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P and Witten I H 2009 *ACM SIGKDD Explorations Newsletter* **11** 10–18