Vietnam National University, Hanoi
University of Engineering and Technology

Advanced Institute of Engineering and Technology

Technical Report

UET-AVITECH-2019002 April 18, 2019

Simultaneous Tensor Decomposition for EEG Epileptic Spike Detection

Le Trung Thanh, Nguyen Thi Anh Dao, Nguyen Viet Dung
Nguyen Linh Trung, Karim Abed-Meraim

Hanoi, Vietnam

Simultaneous Tensor Decomposition for EEG Epileptic Spike Detection

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

April 18, 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 report, 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 compared the proposed tensor analysis method with other common tensor methods in analyzing EEG signal and compared the proposed feature extraction method with Phan's method. Experimental results indicated that our proposed method is able to detect epileptic spikes with good performance.

Significance: To suitably deal with EEG spikes, we developed a local solution for nonnegative SMLRAT. For practical implementation, we proposed the generalized SMLRAT algorithm to effectively solve the SMLRAT and nonnegative SMLRAT problems. An efficient EEG feature extraction framework was proposed, based on estimating the "eigenspikes" from the nonnegative generalized SMLRAT algorithm.

Index Terms

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

[1] University of Engineering and Technology, Vietnam National University, Hanoi, Vietnam.
[2] University of Technology and Logistics, Bac Ninh, Vietnam.
[3] National Institute of Advanced Technologies of Brittany, Brest, France.
[4] University of Orléans, Orléans, France.
* Corresponding author, linhtrung@vnu.edu.vn.

Contents

Simultaneous Tensor Decomposition for EEG Epileptic Spike Detection

I. Introduction

Epilepsy is one of the most common brain disorders. 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. According to estimates [1] in 2010, epilepsy affects about 50 million people worldwide in which nearly 40 million people live in developing countries.

For epilepsy diagnosis and treatment, one often needs to observe epileptic seizures or epileptiforms in order 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 (locate) the epileptic spikes in long EEG records or to reduce false-alarm detection (which is often the case due to the fact that 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 (e.g. algorithms of signal processing in electrical engineering and of machine learning in computer science) 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 analyze EEG epileptic spikes. Therefore, the aim of our study is to seek for a new method of tensor decomposition which is able to analyze EEG epileptic spikes and hence facilitate automatic spike detection.

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 report, 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 nonnegative Tucker decomposition (MR-NTD) [36], concurrent subspaces analysis (CSA) [37], multilinear discriminant analysis (MLDA) [38], multilinear PCA (MPCA) including unconstrained MPCA [39], nonnegative MPCA [40] and sparse MPCA [41].

In view of the above literature, the work in this report has two main contributions. It is noted that a preliminary study has been preliminarily presented in a conference [42]. The first contribution of this report 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); the latter is also called 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 [43]. Finally, inspired by the feature extraction algorithm proposed by Phan and Cichocki in [44] decomposing a compound tensor formed by concatenating a set of data tensors, we propose a practical algorithm, namely generalized SMLRAT (GSMLRAT), to effectively solve the SMLRAT and nonnegative SMLRAT problems. We note here that the above reviewed SLRAT methods are considered as special cases of our proposed SLMRAT, and will be later analyzed in Section VI. Also, a similar approach to [44] can be found in [45], [46], which is based on the linear system with a CP decomposition constrained solution (LS-CPD) framework for data classification.

The second contribution of the report is the successful application of tensor decomposition to detection of EEG epileptic spikes, particularly proposing an SMLRAT-based system, thanks to the proposed SMLRAT method. We first propose an efficient EEG feature extraction framework, 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 were then fed into widely used "shallow" classifiers to evaluate their separability between epileptic and non-epileptic spikes.

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

II. Preliminaries

A. Tensor Notations and Definitions

Follow notations defined in [47], 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 $\boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \times \dots \times I_{k-1} \times r_k \times I_{k+1} \dots \times I_n}$ 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 $\boldsymbol{\mathcal{A}}, \boldsymbol{\mathcal{B}} \in \mathbb{R}^{I_1 \times I_2 \dots \times I_n}$ is defined by

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

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

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

Remark that, a $(n\text{-}1)$-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

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

B. 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:

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

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.

1) CP decomposition:

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

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

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 $\boldsymbol{\mathcal{G}}$. In order to solve f_{CP} in (2), the "workhorse" algorithm is based on alternating least squares (ALS) [47], [48].

2) Tucker Decomposition:

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

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

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 [47, 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 [49].

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 [43]. 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.

III. 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.

A. 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 [47], [49] of (3), while SLRAM was shown to be a special case of Tucker-ALS [50, 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 ([50, 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)*.*

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

It is also well-known that HOSVD gives a sub-optimal solution of (3) [47], [49]. 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).*

B. Generalized Simultaneous Multilinear LRAT

Inspired by results in Section III-A, 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.

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*

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

Proof Sketch: We first show that minimizing the objective function f_{SMLRAT} in (6) corresponds to maximizing the following objective function

$$f'_{\text{SMLRAT}}\big(\{\mathbf{U}_k\}_{k=1}^n\big) = \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \cdots \times_n \mathbf{U}_n^T\|_F^2.$$

After that, we exploit that maximizing f'_{SMLRAT} is equivalent to

$$\begin{aligned} \arg \max_{\mathbf{U}_j \in \mathbb{R}^{I_j \times r_j}} \quad & \operatorname{trace} \mathbf{U}_k^T \mathbf{R}_k \mathbf{U}_k \\ \text{s.t.} \quad & \mathbf{U}_k^T \mathbf{U}_k = \mathbf{I}_{r_k}. \end{aligned} \tag{10}$$

As a result, the solution of maximizing f_{SMLRAT} is obtained from the r_k principal eigenvectors of $\mathbf{R}_k$. Finally, given n orthogonal matrices $\{\mathbf{U}_k\}_{k=1}^n, \mathbf{U}_k \in \mathbb{R}^{I_k \times r_k}$, if tensors $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^N$ are determined by

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

then $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^N$ are optimal tensors for f_{SMLRAT}. The detailed proof is deferred to the supplementary materials because of space limit.

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*

$$\boldsymbol{\mathcal{G}}_i = \boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \cdots \times_n \mathbf{U}_n^T, \tag{11}$$

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 gives 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 [51]–[53].

Therefore, we can also obtain an alternative solution of SMLRAT using the HOSVD algorithm, as given by the following proposition.

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).*

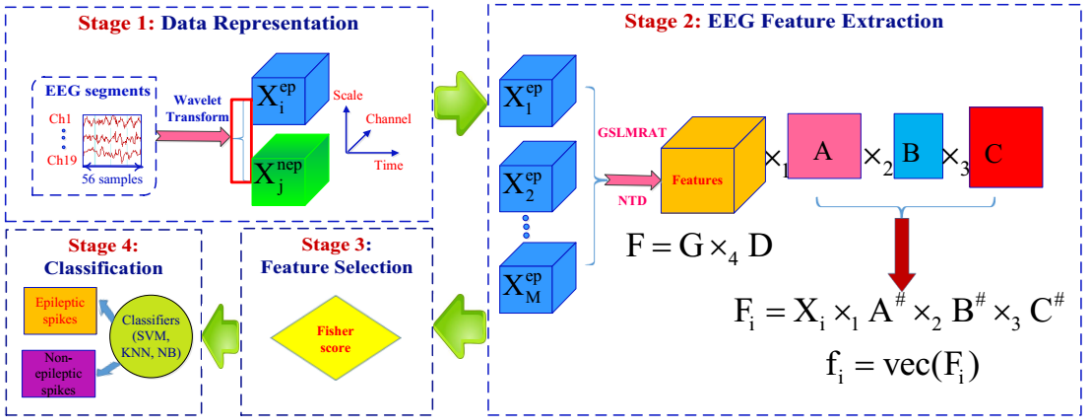


Fig. 1: Proposed epileptic spike detection system.

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

$$\begin{aligned} \arg\min_{\{\widetilde{\boldsymbol{\mathcal{X}}}_i\}_{i=1}^N} \quad & f_{\text{NSMLRAT}} = \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i - \widetilde{\boldsymbol{\mathcal{X}}}_i\|_F^2 \\ \text{s.t.} \quad & \widetilde{\boldsymbol{\mathcal{X}}}_i = \boldsymbol{\mathcal{G}}_i \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n. \\ & \{\mathbf{U}_k\}_{k=1}^n \geqslant 0. \end{aligned} \tag{12}$$

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 (12).*

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.

IV. 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 feature selection stage, we propose to apply Fisher score for feature selection, being able to select EEG features significant for the purpose of epileptic spike detection. In the classification stage, we only use well-known "shallow" classifiers are

applied, as opposed to recent classifiers based on "deep" learning. Now, we will focus on our contributions to the second and the third stages.

A. 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

$$f_{\text{EEG}} = \sum_{i=1}^{N_1} \|\boldsymbol{\mathcal{X}}_i^{\text{ep}} - \boldsymbol{\mathcal{G}}_i^{\text{ep}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}\|^2, \tag{13}$$

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 [44], 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

$$\begin{aligned} \widetilde{\boldsymbol{\mathcal{X}}}^{\text{ep}} &= \boldsymbol{\mathcal{X}}_1^{\text{ep}} \boxplus \boldsymbol{\mathcal{X}}_2^{\text{ep}} \cdots \boxplus \boldsymbol{\mathcal{X}}_{N_1}^{\text{ep}} \\ &\overset{\text{NTD}}{=} \boldsymbol{\mathcal{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \times_4 \mathbf{D}, \end{aligned} \tag{14}$$

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 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

$$\underbrace{\mathbf{X}_{(k)}}_{\text{input data}} = \underbrace{\mathbf{D}}_{\text{basic vectors}} \underbrace{\mathbf{G}_{(k)}(\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})^T}_{\text{coefficients}}, \tag{15}$$

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.

$$\mathcal{F}^{\text{ep}} = \boldsymbol{\mathcal{G}} \times_4 \mathbf{D}. \tag{16}$$

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 $\mathcal{F}_{\text{time}}^{\text{ep}} = \boldsymbol{\mathcal{G}} \times_1 \mathbf{A} \times_4 \mathbf{D}$, to yield the principal axes of variations of an epileptic spike across channel and wavelet-scale modes. Similarly, we can derive the channel-basis and scale-basis of the eigenspike space, i.e, $\mathcal{F}_{\text{scale}}^{\text{ep}} = \boldsymbol{\mathcal{G}} \times_2 \mathbf{B} \times_4 \mathbf{D}$ and $\mathcal{F}_{\text{channel}}^{\text{ep}} = \boldsymbol{\mathcal{G}} \times_3 \mathbf{C} \times_4 \mathbf{D}$.

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

$$\begin{aligned} \boldsymbol{\mathcal{F}}_m^{\text{train}} &= \boldsymbol{\mathcal{G}}_m^{\text{train}} \times_4 \mathbf{D} = \boldsymbol{\mathcal{X}}_m^{\text{train}} \times_1 \mathbf{A}^{\#} \times_2 \mathbf{B}^{\#} \times_3 \mathbf{C}^{\#}, \\ \mathbf{f}_m^{\text{train}} &= \text{vec}(\boldsymbol{\mathcal{F}}_m^{\text{train}}). \end{aligned} \tag{17}$$

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.

$$\begin{aligned} \boldsymbol{\mathcal{F}}^{\text{test}} &= \boldsymbol{\mathcal{X}}^{\text{test}} \times_1 \mathbf{A}^{\#} \times_2 \mathbf{B}^{\#} \times_3 \mathbf{C}^{\#}, \\ \mathbf{f}^{\text{test}} &= \text{vec}(\boldsymbol{\mathcal{F}}^{\text{test}}). \end{aligned} \tag{18}$$

Remark: Our formulation of the concatenated tensor in (14) differs from [44] 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 activities are very huge in EEG datasets, 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 the 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 [54] (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 [55], [56]. 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 [55], [56]: 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 [44] was proposed to solve the problem of binary/multi-class classification, concatenating all training tensors derived from multiple classes.

B. 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 include 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. As a consequence, 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 Fisher score [57] for each feature to assess the effectiveness of the 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, that is

$$f_{\text{Fisher}}(\mathbf{w}) = \frac{\sigma^2_{\text{between}}}{\sigma^2_{\text{within}}} = \frac{[\mathbf{w}(\mu_1 - \mu_2)]^2}{\mathbf{w}^T(\Sigma_1 + \Sigma_2)\mathbf{w}}. \tag{19}$$

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

$$\gamma(\mathbf{f}_i) \triangleq \mathbf{w}(i) = \frac{N_1(\mu_{i,1} - \mu_i)^2 + N_2(\mu_{i,2} - \mu_i)^2}{N_1\sigma^2_{i,1} + N_2\sigma^2_{i,2}}. \tag{20}$$

We select l significant features with top Fisher scores:

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

C. 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 [58], 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 a reasonable solution for the best rank-$(r_1, r_2, \ldots, r_n)$ tensor approximation [49]. 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_1, r_2, r_3) in the factors $(\mathbf{U}_1, \mathbf{U}_2, \mathbf{U}_3)$ can be estimated from their corresponding modes $(\mathbf{X}_{(1)}, \mathbf{X}_{(2)}, \mathbf{X}_{(3)})$ using the truncated SVD, as follows:

$$\mathbf{X}_{(k)} \approx \mathbf{U}_k^{I \times r_k} \mathbf{\Lambda}_k^{r_k \times r_k} \mathbf{V}_k^{r_k \times JK}, k = 1, 2, 3. \tag{21}$$

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.

$$\text{VAR}_{r_1} = \frac{\sum_{i=1}^{r_1} \lambda_i}{\sum_{j=1}^{I} \lambda_j} 100\%. \tag{22}$$

V. Experimental Results and Discussions

A. 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 I. Figure 2 illustrates some epileptic spikes from this dataset. Epileptic spikes were manually identified by a neurologist from Vietnam National Children's Hospital.

In order 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 perceptron to obtain "possible" spikes. The remaining set of possible spikes now included spikes which are either real epileptic or "non-epileptic". These non-epileptic spikes often cause misdiagnosis; they are not related to epilepsy but may be easily diagnosed as epileptic spikes. Non epileptic spikes are large positive or negative voltage transients that can be confused as epileptic spikes by regular algorithms. The goal of a spike detection algorithm is to separate these two types of spikes.

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

TABLE I: 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	22m0s	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	17m7s	3/53302
13	F	13	18m53s	5/69583
14	F	16	20m14s	8/6217
15	F	20	14m32s	324/11219
16	F	22	17m 56s	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.

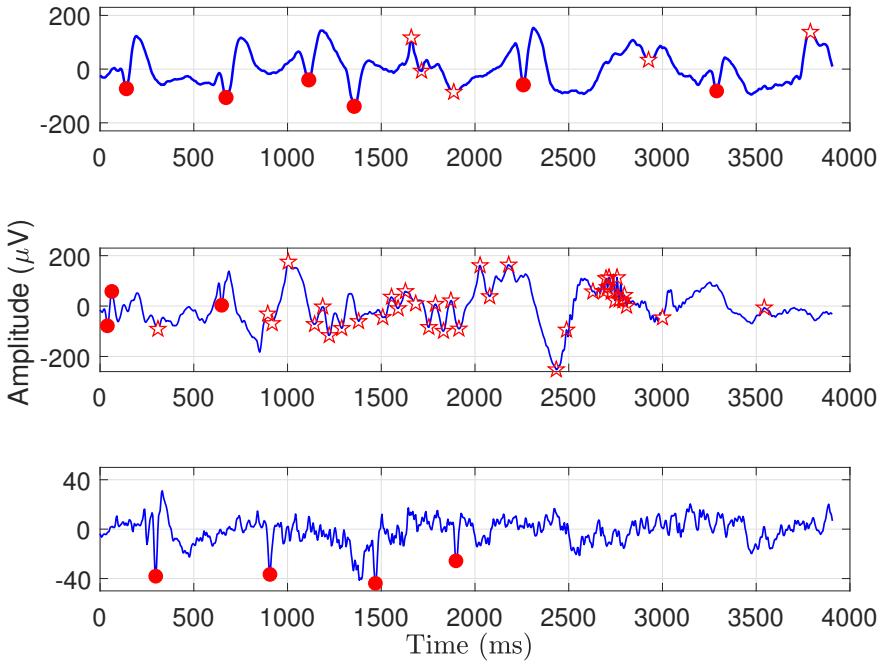


Fig. 2: Some epileptic spikes (circle markers) and non-epileptic spikes (star markers) derived from three typical patients in our filtered EEG data.

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 $[4\text{-}8]$ to the size of 20, instead of 5 as used in [9]. As a result, 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.

B. Performance Metrics

In the evaluation task of the EEG epileptic spike detection problem, three statistical metrics including Sensitivity (SEN, aka Recall), Specificity (SPE) and Accuracy (ACC) are widely used to evaluate performance of detection systems, see [9], [18], [59]–[62] 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 of results on evaluating the average performance of EEG interictal spike detection algorithms [63], the overall performance with respect to the metric ρ (e.g. SEN) of our system can be averaged in the following ways:

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

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

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

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

where T is the number of patients and the i-th patient has a recording duration D_i and N_i desired events (e.g. epileptic spikes).

C. Experiment Setups and Results

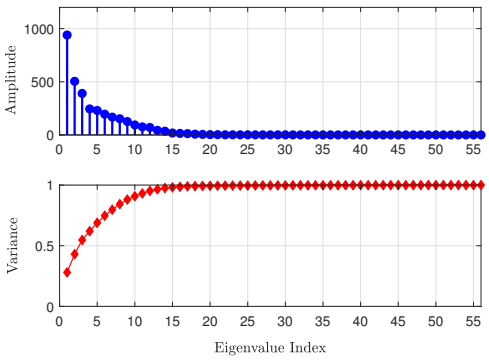
Our experiments are conducted to study the three stages, by: (i) performing feature extraction by estimating the eigenspikes and the corresponding features for the EEG data samples, (ii) performing feature selection to obtain the significant features for classification task, and (iii) performing classification 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 particular, the LOO-CV method is a well-known cross validation method for evaluating the performance of the classifiers. 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.

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

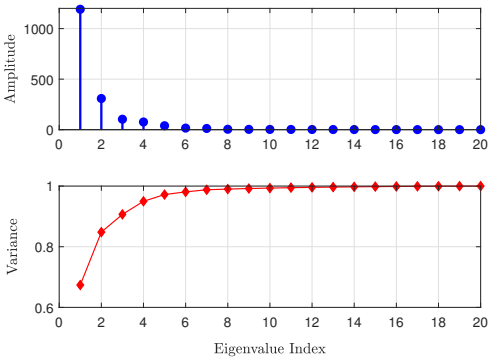
$$\mathbf{A} \approx \sum_{i=1}^{15} \lambda_i \mathbf{u}_i \mathbf{v}_i^T,$$

where λ_i is the i-th eigenvalue 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).

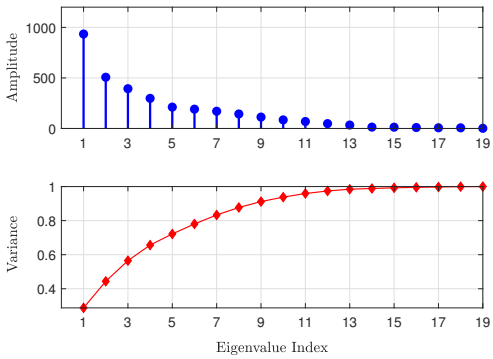
By performing NTD of the training four-way epileptic tensor $\widetilde{\boldsymbol{\mathcal{X}}}^{\text{ep}} \in \mathbb{R}_{+}^{56\times 20\times 19\times M}$, with M is the number of training three-way tensors, we obtained common factors $\mathbf{A} \in \mathbb{R}_{+}^{56\times 15}$, $\mathbf{B} \in \mathbb{R}_{+}^{20\times 10}$, $\mathbf{C} \in \mathbb{R}_{+}^{19\times 19}$. 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



(a) Time mode



(b) Scale (freq.) mode



(c) Channel mode

Fig. 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.

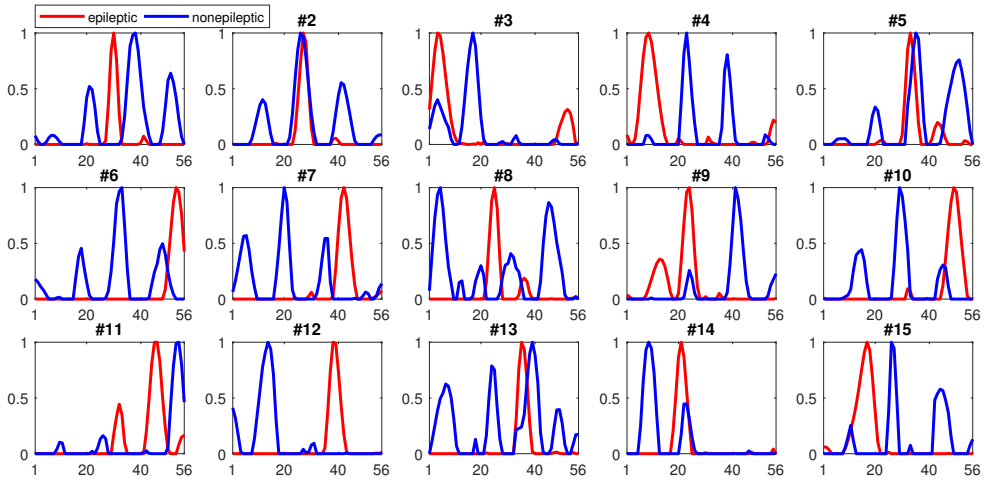
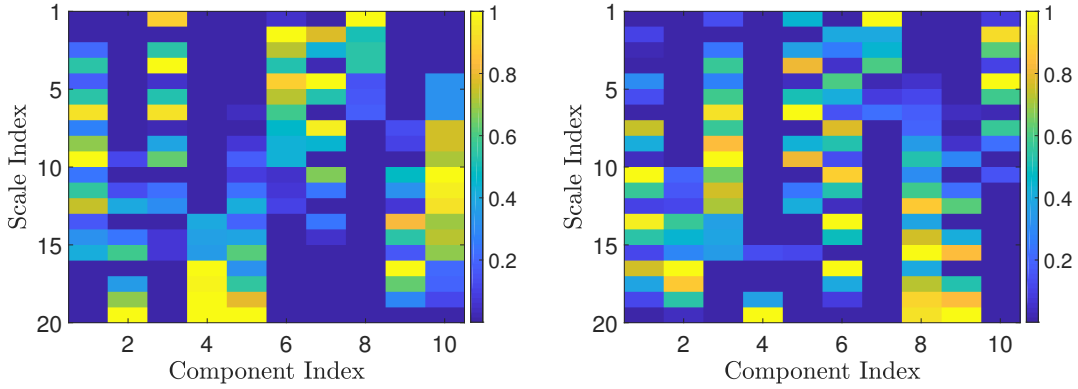


Fig. 4: Common time factor $\mathbf{A} \in \mathbb{R}_{+}^{56\times 15}$ derived from NTD.

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. As a result, it may lead to the ability of learning localized parts of epileptic spikes from the channel mode.



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

Fig. 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 $[4\text{-}8]$.

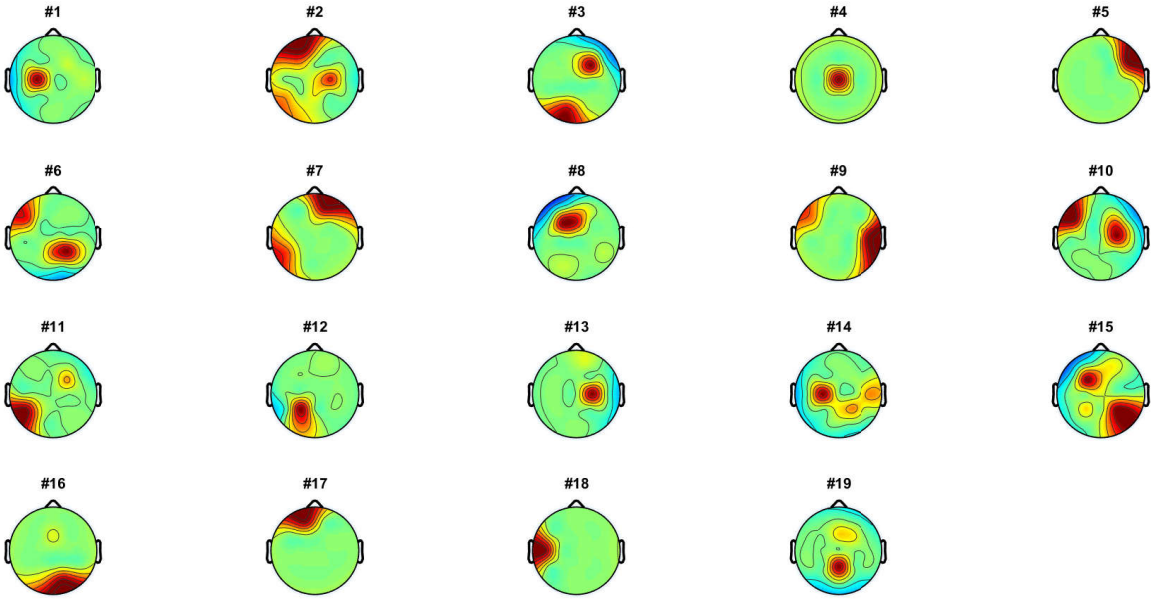
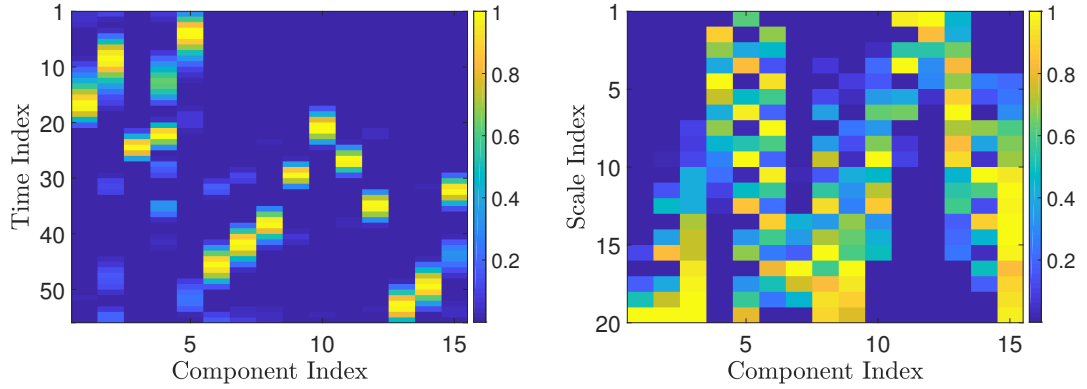


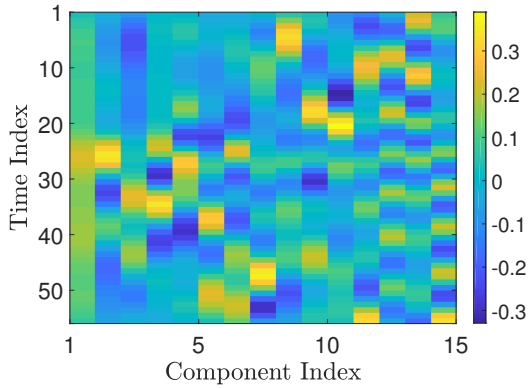
Fig. 6: Common channel factor $\mathbf{C} \in \mathbb{R}_{+}^{19\times 19}$ of the epileptic tensor derived from NTD.

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 $[r_1, r_2, r_3] = [15, 15, 15]$. It is due to the fact that CP decomposition requires the same number of rank-1 tensors (i.e. the number of components in the CP model), i.e. $r_1 = r_2 = r_3$. The results are shown in Figure 7. As we can see that, the two nonnegative types of tensor decomposition, including NTD and NCP, yielded sparse basis vectors in $\mathbf{A}$ and $\mathbf{B}$, but the original Tucker decomposition (TD) did not. However, the factor $\mathbf{B}$ obtained from NCP did not provide a good sparse representation for the wavelet scale mode.

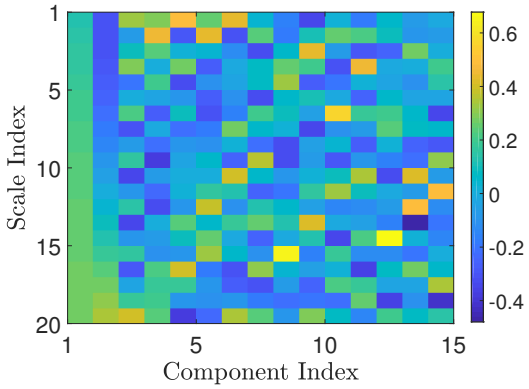


(a) Factor **A**: NTD

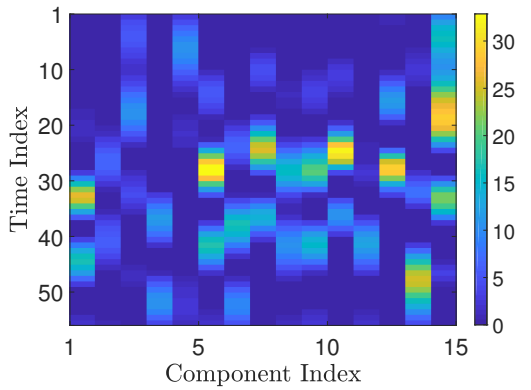
(b) Factor **B**: NTD



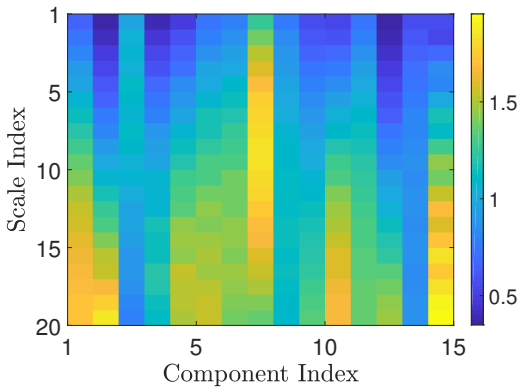
(c) Factor **A**: TD



(d) Factor **B**: TD



(e) Factor **A**: NCP



(f) Factor **B**: NCP

Fig. 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.

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 [64] 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

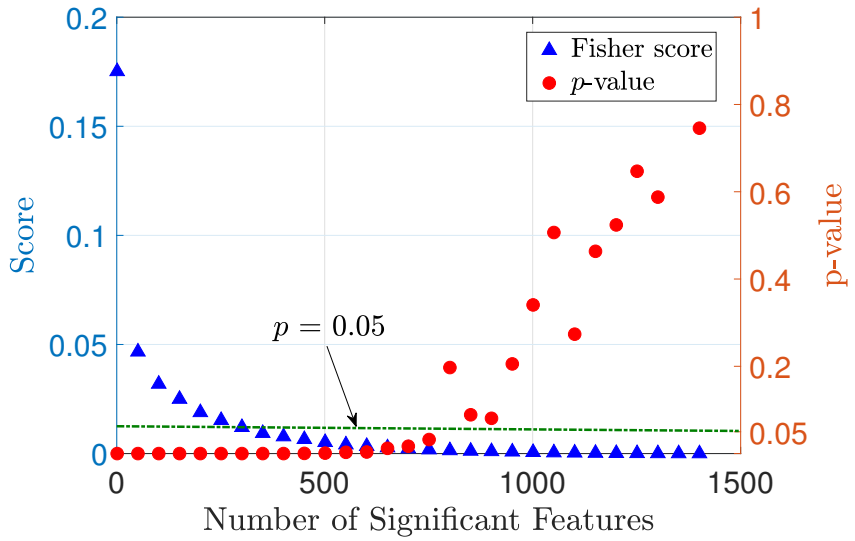


Fig. 8: Fisher scores (blue marks) and p-values (red mark) of features. Features are ordered based on their Fisher score. The significant level $p = 0.05$ is to make decision for rejecting H_0, removing features with $p > 0.05$.

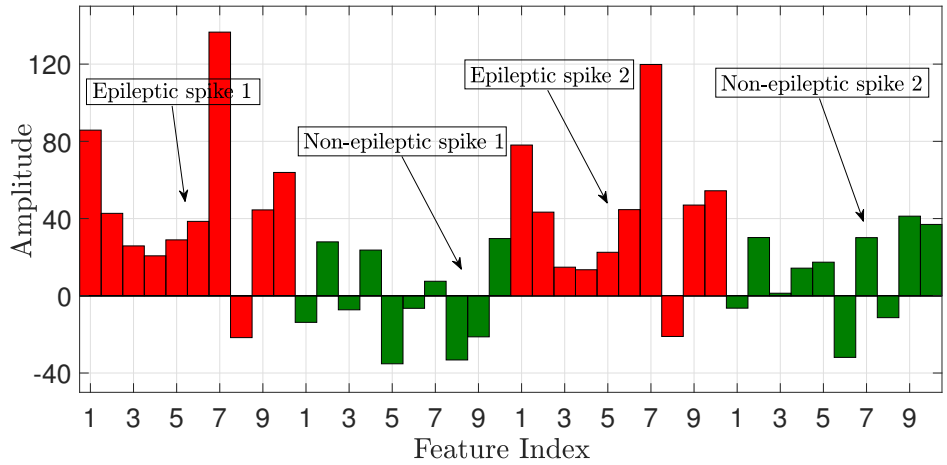


Fig. 9: Top 10 selected features of two typical epileptic spikes and two non-epileptic spikes. Behavior of features derived from epileptic spikes are similar, unlike non-epileptic spikes.

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 [64]. The experimental results are shown in Figures 8, 9, and 10.

It can be seen from Figure 8 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. As a result, these 500 features have stronger discrimination power than others. Furthermore, these features are significant different to that of non-epileptic class Ω_2, as illustrated in Figure 9. As a result, these selected features were efficient in detecting epileptic spikes. Performance of the SVM model using the

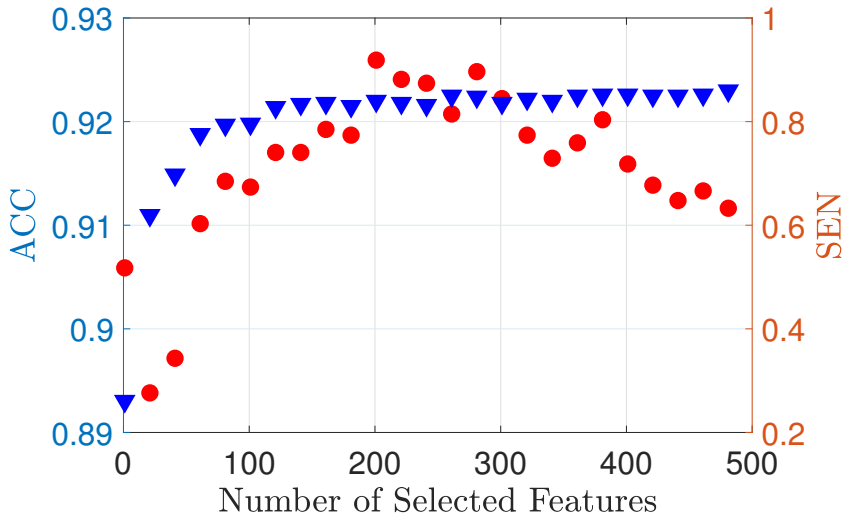


Fig. 10: Classification performance vs. number of selected features

first 500 significant features are shown in Figure 10. As we can see, the classifier achieved the top SEN at approximately 0.9 and overall ACC around 0.92 when using the first 200 features, while the corresponding values of the area under ROC curves (AUC) were always higher than 0.9, thus ranked as excellent result of classification[1].

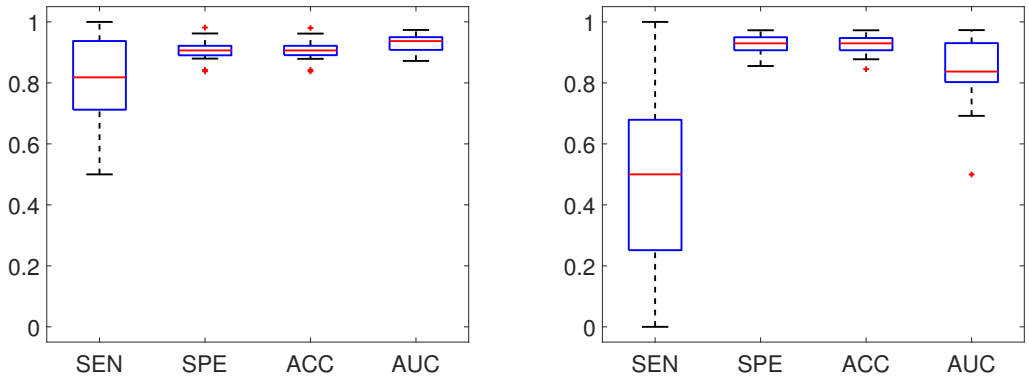
3) Classification:

To investigate how effective concatenation of input tensors is at detecting epileptic spikes, we compare the proposed method against Phan-Cichocki method in [44]. Second, we use other tensor-based approaches which were successfully applied to detect brain activities of interest in EEG signals as comparative methods, including Tucker-based [15], [18], CP-based [14], [16], NCP-based [65].

In order to evaluate the separability of the extracted features, we applied three well-known and widely used classification models in the platform WEKA [66] 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. Note that, we report here results of the SVM model trained with our features across 17 patients using the LOOCV method. For space reasons, the detailed results of other classifiers are reported in the supplementary materials.

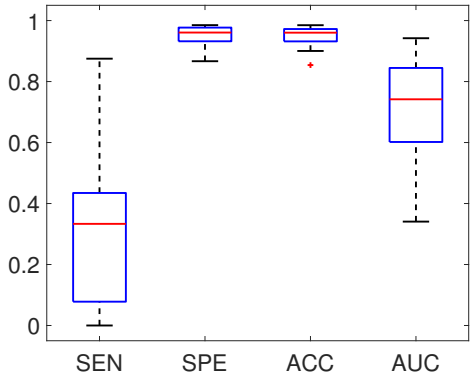
It can easily observe that feature extraction is key for EEG epileptic spike detection, and our approach outperforms the baselines on all evaluation metrics. First, the NTD-SVM model

[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.

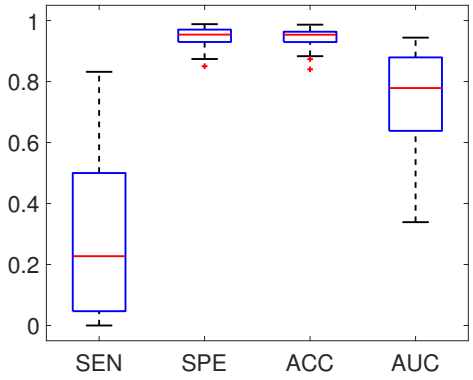


(a) NTD-SVM

(b) TD-SVM



(c) CP-SVM



(d) NCP-SVM

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

yielded strong results in term of all measurements (SEN, SPE, ACC and AUC) (see Tab II). Table II presents a quantitative statistic of EEG epileptic spike detection performance of our NTD-SVM model using 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 spire of the variation, the average metrics for SEN were still good, e.g. the arithmetic mean $\mathrm{SEN}_{\mathrm{AM}} = 0.8044$, $\mathrm{SEN}_{\mathrm{TA}} = 0.8516$, and $\overline{\mathrm{SEN}} = 0.8299 \pm 0.0297$. While the metrics with respect to the non-epileptic class, including SPE and ACC were all high with small standard deviations (i.e. $\bar{\rho} \geqslant 0.9$ and $\mathrm{S.D.} \approx 0.04$). 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 $\overline{\mathrm{AUC}} = 0.9323 \pm 0.0076$ on average. These results indicate that the features extracted by our method are able to use to detect epileptic spikes with good performance.

Second, our detection system outperforms the three other tensor-based approaches, including CP, NCP and unconstrained TD decompositions (see Fig 11 and 13(a)). Fig 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) plots the distribution of evaluation

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

performance across 17 patients in our EEG dataset. A box is based on the five number summary, 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. $\mathrm{IQR} = \mathrm{Q3} - \mathrm{Q1}$) 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 Fig 13(a) that shows ROC curves to illustrate overall performance of the four models. The ROC curve is drawn by ploting the true positive rate (TPR equivalent to SEN) and false positive rate that can be computed as $1 - \mathrm{SPE}$. As a result, the ROC curve allows us to derive a cost/benefit analysis for making decision. We can easily 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 base approach in this work. According to the Tab III, 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. $\mathrm{AUC} = 0.574 \pm 0.25$. The results of TD-based models were similar to that of CP-based models. Although TD-SVM might provide a good performance in term of

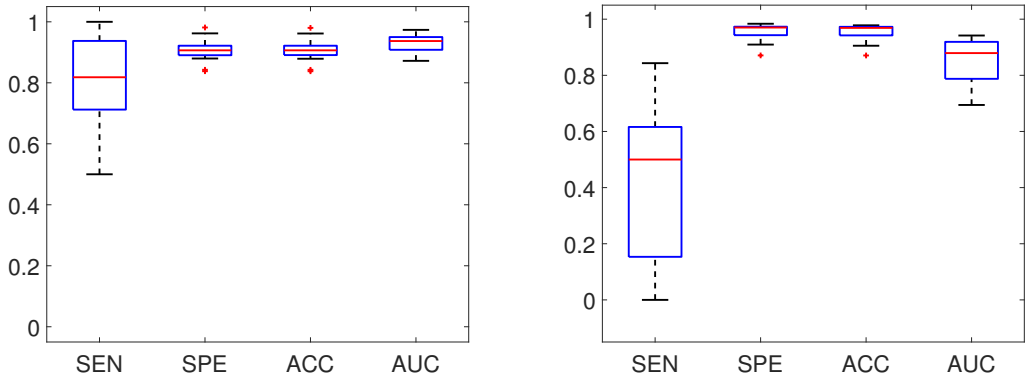
TABLE III: Detection performance comparison between using difference machine learning models

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 $\pm$ S.D.

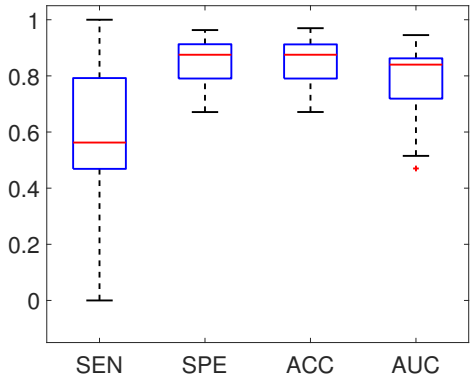
AUC (i.e. 0.836 ± 0.113), the resulting SEN was not good enough, around 0.5. Hence, a half of the total number of epileptic spikes were detected incorrectly and labeled as non-epileptic activities. Meanwhile, the NTD-based models yielded a 10% to 30% better performance than that of other tensor decompositions. The two best overall accuracy belonged to the NTD-based models, including NTD-SVM and NTD-DT (i.e. $\text{AUC} = 0.929 \pm 0.029$ and 0.914 ± 0.041 respectively, while $\text{SEN} \geqslant 0.8$ 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.

Third, 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 Tab III and Fig 12. The Tab III shows the overall performance of 16 models in term of all evaluation metrics. We can see that 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 $\text{AUC} = 0.929 \pm 0.029$ and $\overline{\text{AUC}} = 0.932 \pm 0.008$ in term 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 Fig 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 spire of that, the average SEN of the four models using

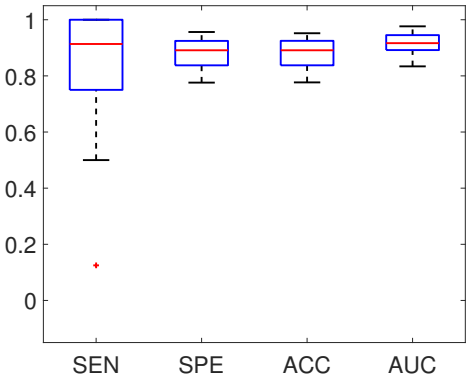


(a) NTD-SVM

(b) NTD-KNN



(c) NTD-NB



(d) NTD-DT

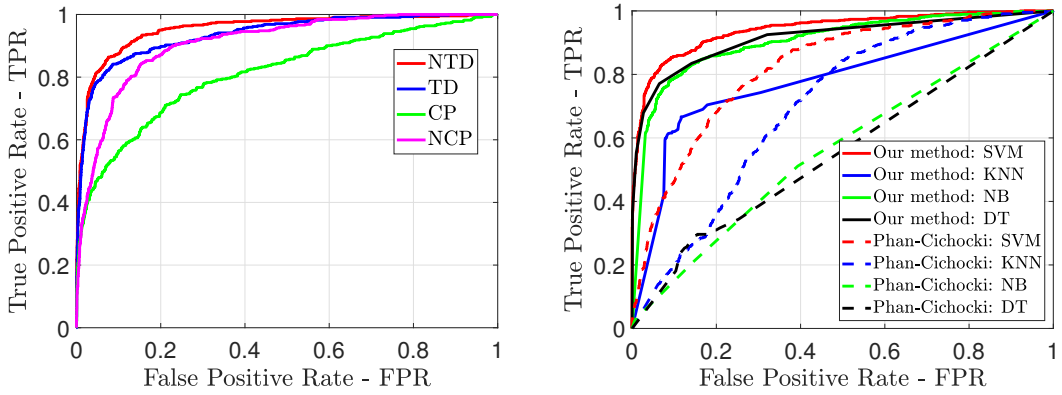
Fig. 12: Detection performance of four classifiers using our features

TABLE IV: 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

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. $\mathrm{SEN} < 0.5$, see the second column of Tab III). 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 which four evaluation metrics for AUC of NTD were all higher 0.9.

The concatenation of training input tensors is key for EEG epileptic spike detection. Accordingly, our method provided a better performance than Phan-Cichocki method in EEG epileptic spike detection problem. Tab IV and Fig 13(b) present a performance comparison of epileptic spike detection between using our method and Phan-Cichocki method. We note that, according to Phan-Cichocki method, the *complete* set of training tensors was used



(a) Tensor-based epileptic spike detection methods

(b) Concatenation of input tensors

Fig. 13: Performance comparison in terms of averaged ROC.

to concatenate a single four-way tensor whose the tensor decomposition was decomposed. 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 $\widetilde{\boldsymbol{\mathcal{X}}}_{\text{train}}$ 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 $\widetilde{\boldsymbol{\mathcal{X}}}_{\text{train}}$ 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 [54]. 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 three widely used classifiers KNN, NB and DT, as provided above. 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. While our NTD-SVM and NTD-DT models provided much better performance in term of SEN in which they detected more than 80% the number of epileptic spike correctly. The metrics with respect to non-epileptic class (i.e. SPE and ACC) of both four classifiers were also reasonable, 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 upplementary materials for further detailed results of the four classifiers using features extracted by Phan-Cichocki method.

VI. Connection to Related Works

In this section, we discuss our proposed method in relation to some previous techniques. In particular, we first examine its connection to simultaneous component analysis (SCA) [34], [35]. Next, an equivalence between concurrent subspace analysis (CSA) [37] and the proposed GSMLRAT are presented. Finally, two kinds of multilinear PCA, including MPCA [67] and nonnegative MPCA [40], are considered as further versions of GSMLRAT.

A. Connection to SCA

Simultaneous component analysis (SCA) can be considered as a special case of SMLRAT as well as SLRAM for two dimensional data [34], in which the objective function f_{SCA} requires the following optimization:

$$\arg\min f_{\text{SCA}} = \sum_{i=1}^{N} \|\mathbf{X}_i - \mathbf{T}_i\mathbf{P}_i^T\|_F^2,$$

where $\mathbf{X}_i \in \mathbb{R}^{I\times J}$ is the underlying input matrices, $\mathbf{T}_i \in \mathbb{R}^{I\times r_i}$ is component-score matrix and $\mathbf{P}_i \in \mathbb{R}^{J\times r_i}$ is the loading matrix. Note that, data used SCA often relates to a common mode that is the same for all data matrices. The common mode can be either the object mode (i.e. common row mode, $\mathbf{T}_1 = \mathbf{T}_2 = \ldots \mathbf{T}_N = \mathbf{T}$) or the variable mode (i.e. common column mode, $\mathbf{P}_1 = \mathbf{P}_2 = \ldots \mathbf{P}_N = \mathbf{P}$).

Clearly, the objective function f_{SCA} is equivalent to that of SLRAM in (5), where the two modes are with either the common object mode $\mathbf{T}_i = \mathbf{U}_1$ and $\mathbf{P}_i = \mathbf{U}_2\mathbf{F}_i$ or the common variable mode $\mathbf{T}_i = \mathbf{U}_1\mathbf{F}_i$ and $\mathbf{P}_i = \mathbf{U}_2$. The two representations resulted in the two well-known models, namely SUM-PCA [68] and SCA-P [69] respectively. Recently, Stegeman discussed the relationship between the SCA and Tucker-ALS in [35], as an evidence of Proposition 3. In particular, the function f_{SCA} can be reformulated as an extended version of SCA-P, that is,

$$\arg\min f_{\text{SCA-T3}} = \sum_{i=1}^{N} \Big\|\mathbf{X}_i - \mathbf{A}_i\Big(\sum_{r=1} c_{ir}\mathbf{G}_i\Big)\mathbf{B}^T\Big\|_F^2,$$

over three factors $\mathbf{A} \in \mathbb{R}^{I\times r_1}$, $\mathbf{B} \in \mathbb{R}^{J\times r_2}$, $\mathbf{C} \in \mathbb{R}^{N\times r_3}$ and a core tensor $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{r_1\times r_2\times r_3}$. The objective function $f_{\text{SCA-T3}}$ is identical to that of Tucker model for three-way tensors in [47]. Specifically, $\mathbf{A}_i$ contain component scores of r_1 components, $\mathbf{B}$ contains contains loadings of J variables on r_2 components, meanwhile $\mathbf{C}$ contains weights for N matrices $\mathbf{X}_i$ on r_3 components. As a result, the solution of minimizing $f_{\text{SCA-T3}}$ can be obtained by using alternative least square (ALS) method, which is analogous to ALS algorithm for Tucker decomposition.

B. Connection to CSA

Concurrent subspaces analysis (CSA) for higher way tensor objects was proposed by Xu *et al.* in [37]. The CSA model can be considered a further version of SMLRAT. In particular,

it was formulated by an optimization via the optimal reconstruction criterion as follows:

$$\begin{aligned} \underset{\{\mathbf{U}_k\}_{k=1}^n}{\arg\min} \quad & f_{\text{CSA}} = \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i - \boldsymbol{\mathcal{X}}_i^{\text{re}}\|_F^2, \\ \text{s.t.} \quad & \boldsymbol{\mathcal{X}}_i^{\text{re}} = \boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1\mathbf{U}_1^T \cdots \times_n \mathbf{U}_n\mathbf{U}_n^T, \end{aligned}$$

where $\boldsymbol{\mathcal{X}}_i \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_n}$ is the underlying n-way tensor, the reconstructed tensor $\boldsymbol{\mathcal{X}}_i^{\text{re}}$ of $\boldsymbol{\mathcal{X}}_i$ is derived via projection matrices $\{\mathbf{U}_k\}_{k=1}^n, \mathbf{U}_k \in \mathbb{R}^{r_k \times I_k}$.

As claimed in Proposition 3, for any estimation of common projection matrices $\{\mathbf{U}_k\}_{k=1}^n$, the set of optimal core tensors $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^N$ is given by

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

Replacing core tensors $\boldsymbol{\mathcal{G}}_i$ in the objective function f_{SMLRAT} in (6) by the optimal one, we obtain the objective function f_{CSA} for CSA, that is,

$$\arg\min_{\{\tilde{\boldsymbol{\mathcal{X}}}_i\}_{i=1}^N} f_{\text{SMLRAT}} = \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i - \tilde{\boldsymbol{\mathcal{X}}}_i\|_F^2$$

where the low multilinear rank tensor approximation $\tilde{\boldsymbol{\mathcal{X}}}_i$ is imposed by

$$\begin{aligned} \tilde{\boldsymbol{\mathcal{X}}}_i &= \left(\boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \cdots \times_n \mathbf{U}_n^T\right) \times_1 \mathbf{U}_1 \cdots \times_n \mathbf{U}_n \\ &= \boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1\mathbf{U}_1^T \cdots \times_n \mathbf{U}_n\mathbf{U}_n^T = \boldsymbol{\mathcal{X}}_i^{\text{re}}, \end{aligned}$$

because of the k-mode production's property as mentioned in Section II. In addition, Xu *et al.* produced an iterative algorithm for solving f_{CSA} in which each iterative step was inspired by the Tucker-ALS (HOOI) algorithm for $(n+1)$-way tensors. Hence, CSA can be viewed as a further version of SMLRAT.

C. Connection to MPCA

Multilinear principal component analysis (MPCA) is considered as a generalization of PCA for high way tensors [67]. In particular, MPCA minimizes the similar reconstruction error function to SMLRAT, as follows

$$\underset{\{\mathbf{U}_k\}_{k=1}^n}{\arg\min} \quad f_{\text{MPCA}} = \sum_{i=1}^{N} \|\bar{\boldsymbol{\mathcal{X}}}_i - \boldsymbol{\mathcal{G}}_i \times_1 \mathbf{U}_1 \cdots \times_n \mathbf{U}_n\|_F^2,$$

where the underlying tensors $\boldsymbol{\mathcal{X}}_i$ are centered by mean of the data samples as $\bar{\boldsymbol{\mathcal{X}}}_i = \boldsymbol{\mathcal{X}}_i - \bar{\boldsymbol{\mathcal{X}}}$, with the sample mean is defined by $\bar{\boldsymbol{\mathcal{X}}} = \frac{1}{N}\sum_{i=1}^N \boldsymbol{\mathcal{X}}_i$. In other words, SMLRAT on centered data is equivalent to MPCA. As a result, the unconstrained MPCA can be obtained from the GSMLRAT algorithm by adding above centering step while Tucker-ALS was used as the iteration step, as introduced by Lu *et al.* in [39].

Nonnegative MPCA (NMPCA) proposed by Panagakis *et al.* in [40] is also considered as a constrained version of SMLRAT. Specifically, nonnegative MPCA aims to define nonnegative projection matrices (i.e. $\mathbf{U}_k \in \mathbb{R}^{r_k \times I_k}$) in order to retain the nonnegativity property of the input

tensors. Clearly, the objective function f_{NMPCA} to be maximized in the nonnegative MPCA problem is similar to f_{MPCA} in MPCA as well as f_{SMLRAT} in SMLRAT, that is,

$$f_{\text{NMPCA}} = \frac{1}{2}\text{trace}\left(\mathbf{U}_k^T\big(\sum_{i=1}^{N}\boldsymbol{\mathcal{X}}_{i(k)}\widetilde{\mathbf{U}}_k\widetilde{\mathbf{U}}_k^T\boldsymbol{\mathcal{X}}_{i(k)}^T\big)\mathbf{U}_k\right)$$

where $\boldsymbol{\mathcal{X}}_i$ is underlying tensor, $\{\mathbf{U}_k\}_{k=1}^n$ are imposed to be nonnegative, $\widetilde{\mathbf{U}}_k$ is defined in Eq. (9). The reformulation can be directly obtained from the Theorem 1. The solution for maximizing f_{NMPCA} was developed by exploiting the structure of the Grassmann manifold and computed iteratively by a local optimization procedure in a similar manner to the ALS method. Furthermore, the algorithm is equivalent to the sequential NTD using low-rank approximation based NMF (lraSNTD) in [70]. Hence, nonnegative MPCA can be viewed a constrained version of nonnegative SMLRAT which the algorithm for nonnegative MPCA can be derived from the GSMLRAT algorithm by using lraSNTD in the decomposition stage.

VII. Conclusions

In this report, we have proposed a generalized algorithm, abbreviated as GSMLRAT, for solving the problem of simultaneous multilinear low-rank approximation of tensors. The equivalence between GSMLRAT and subspace analysis methods such as GSLRAM, CSA, SCA and MPCA have been presented in terms of mathematical models.

Inspired by the advantages of GSMLRAT and NTD, we 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 spike class and non-epileptic spike class with high accuracy. In order to reduce feature dimensionality as well as to achieve the best 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 classifiers such as KNN, NB, DT and SVM.

Appendix

A. Proof of Theorem 1

We can divide the proof into three steps, as follows.

a) Step 1: We show that minimizing the objective function f_{SMLRAT} corresponds to maximizing the following objective function:

$$f'_{\text{SMLRAT}}\big(\{\mathbf{U}_k\}_{k=1}^n\big) = \sum_{i=1}^{N}\|\boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \cdots \times_n \mathbf{U}_n^T\|_F^2.$$

By using the Frobenius norm property, the objective function f_{SMLRAT} can be written as

$$\begin{aligned} f_{\text{SMLRAT}} &\triangleq \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i - \widetilde{\boldsymbol{\mathcal{X}}}_i\|_F^2 = \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i\|_F^2 + \|\widetilde{\boldsymbol{\mathcal{X}}}_i\|_F^2 - 2\big\langle \boldsymbol{\mathcal{X}}_i, \widetilde{\boldsymbol{\mathcal{X}}}_i \big\rangle \\ &\overset{(a)}{=} \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i\|_F^2 + \|\boldsymbol{\mathcal{G}}_i\|_F^2 - 2\big\langle \boldsymbol{\mathcal{X}}_i, \boldsymbol{\mathcal{G}}_i \times_1 \mathbf{U}_1 \cdots \times_n \mathbf{U}_n \big\rangle \\ &= \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i\|_F^2 + \|\boldsymbol{\mathcal{G}}_i\|_F^2 - 2\big\langle \boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \cdots \times_n \mathbf{U}_n^T, \boldsymbol{\mathcal{G}}_i \big\rangle, \end{aligned}$$

where (a) follows due to the fact that $\mathbf{U}_i$, $i = 1, \cdots, n$, are orthogonal matrices. Assuming that $\boldsymbol{\mathcal{G}}_i$ and $\boldsymbol{\mathcal{G}}_j$ are independent for all $i \neq j$. The problem of minimizing f_{SMLRAT} is thus equivalently decomposed into N smaller least-square problems as

$$\arg\min_{\boldsymbol{\mathcal{G}}_i} \|\boldsymbol{\mathcal{G}}_i\|_F^2 - 2\big\langle \boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \cdots \times_n \mathbf{U}_n^T, \boldsymbol{\mathcal{G}}_i \big\rangle \tag{I1}$$

for $i = 1, 2, \ldots, N$. As a consequence, (I1) is equivalent to

$$\arg \max_{\{\mathbf{U}_k\}_{k=1}^n} f'_{\text{SMLRAT}}, \tag{I2}$$

where

$$f'_{\text{SMLRAT}} = \sum_{i=1}^{N} \|\boldsymbol{\mathcal{G}}_i\|_F^2 = \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \times \cdots \times_n \mathbf{U}_n^T\|_F^2.$$

b) Step 2: We show that, given n orthogonal matrices $\{\mathbf{U}_k\}_{k=1}^n, \mathbf{U}_k \in \mathbb{R}^{I_k \times r_k}$, if tensors $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^N$ are determined by

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

then $\{\boldsymbol{\mathcal{G}}_i\}_{i=1}^N$ are optimal tensors for f_{SMLRAT}. Due to (I1), the proof of this step can be found in [49, Theorem 4.1].

c) Step 3: Now we show how to obtain an optimal set of factor matrices $\{\mathbf{U}_k\}_{k=1}^n$. From (I2), we observe that

$$\begin{aligned} f'_{\text{SMLRAT}} &= \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_i \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \times \cdots \times_n \mathbf{U}_n^T\|_F^2 = \sum_{i=1}^{N} \|\mathbf{U}_k^T \mathbf{X}_{i(k)} \widetilde{\mathbf{U}}_k\|_F^2 \\ &= \sum_{i=1}^{N} \operatorname{trace}\{\mathbf{U}_j^T (\mathbf{X}_{i(k)} \widetilde{\mathbf{U}}_k \widetilde{\mathbf{U}}_k^T \mathbf{X}_{i(k)}^T) \mathbf{U}_k\} \\ &= \operatorname{trace}\{\mathbf{U}_k^T (\sum_{i=1}^{N} \mathbf{X}_{i(k)} \widetilde{\mathbf{U}}_k \widetilde{\mathbf{U}}_k^T \mathbf{X}_{i(k)}^T) \mathbf{U}_k\} = \operatorname{trace}\{\mathbf{U}_k^T \mathbf{R}_k \mathbf{U}_k\}, \end{aligned}$$

where

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

Therefore, maximizing f'_{SMLRAT} is equivalent to

$$\begin{aligned} \arg \max_{\mathbf{U}_j \in \mathbb{R}^{I_j \times r_j}} \quad & \operatorname{trace} \mathbf{U}_k^T \mathbf{R}_k \mathbf{U}_k \\ \text{s.t.} \quad & \mathbf{U}_k^T \mathbf{U}_k = \mathbf{I}_{r_k}. \end{aligned} \tag{I3}$$

As a result, the solution of maximizing f_{SMLRAT} is obtained from the r_k principal eigenvectors of $\mathbf{R}_k$ [71].

B. Proof of Proposition 3

The Tucker-ALS algorithm can provide an iterative procedure to yield optimal factors of tensor $\boldsymbol{\mathcal{X}}$ [49], as follows. For each iteration step, given for example a factor $\mathbf{U}_k$, one acquires the first r_k singular vectors of the k-mode unfolding matrix $\mathbf{Y}_{k(k)}$ of tensor $\boldsymbol{\mathcal{Y}}_k$ as

$$\mathbf{U}_k = \text{svd}(\mathbf{Y}_{k(k)}, r_k),$$

where $\boldsymbol{\mathcal{Y}}_k$ is given by

$$\boldsymbol{\mathcal{Y}}_k = \boldsymbol{\mathcal{X}} \times_1 \mathbf{U}_1^T \cdots \times_{k-1} \mathbf{U}_{k-1}^T \times_{k+1} \mathbf{U}_{k+1}^T \cdots \times_{n+1} \mathbf{U}_{n+1}^T.$$

Note that, in this analysis, the last factor $\mathbf{U}_{n+1}$ is full column-rank with $r_{n+1} = N$, so it should be an identity matrix $\mathbf{I}_N$. Therefore, the k-mode unfolding of $\boldsymbol{\mathcal{Y}}_k$ can be expressed by

$$\mathbf{Y}_{k(k)} = \mathbf{X}_{(k)}\big(\mathbf{I}_N \otimes \mathbf{U}_n \cdots \otimes \mathbf{U}_{k+1} \otimes \mathbf{U}_{k-1} \cdots \otimes \mathbf{U}_1\big).$$

Hence, another expression of the k-mode unfolding of $\boldsymbol{\mathcal{Y}}_k$ is given by

$$\mathbf{Y}_{k(k)} = [\mathbf{X}_{1(k)} \quad \mathbf{X}_{2(k)} \quad \dots \quad \mathbf{X}_{N(k)}](\mathbf{I}_N \otimes \widetilde{\mathbf{U}}_k),$$

where $\widetilde{\mathbf{U}}_k$ is defined as

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

As a result, the covariance matrix $\mathbf{R}_k$ is also determined as

$$\mathbf{R}_k = \mathbf{Y}_{k(k)}\mathbf{Y}_{k(k)}^T = \sum_{i=1}^{N} \mathbf{X}_{i(k)}\widetilde{\mathbf{U}}_k\widetilde{\mathbf{U}}_k^T\mathbf{X}_{i(k)}^T.$$

So factor $\mathbf{U}_k$ can be given by the top singular vectors of $\mathbf{Y}_{k(k)}$ as

$$\mathbf{U}_k = \text{svd}(\mathbf{Y}_{k(k)}, r_k) = \text{eig}(\mathbf{R}_k).$$

In other words, the factors $\{\mathbf{U}_k\}_{k=1}^n$ obtained from Tucker-ALS exactly form the solution of f_{SMLRAT} in (6) as proved in Theorem 1. This completes the proof of Proposition 3.

C. Proof of Proposition 5

The nonnegative factors $\{\mathbf{U}_k\}_{k=1}^n$ can be computed by the following steps [36, Section III.D]. Given the minimization problem (11), we define the Lagrangian function $\mathcal{L}$ by using Lagrange multipliers (i.e, $\{\boldsymbol{\Lambda}_k\}_{k=1}^n$ and $\{\boldsymbol{\Theta}_j\}_{j=1}^N$) as follows:

$$\mathcal{L} = \sum_{i=1}^{N} \|\boldsymbol{\mathcal{X}}_{i(k)} - \mathbf{U}_k\mathbf{G}_{i(k)}\widetilde{\mathbf{U}}_k^T\|_F^2 - \sum_{j=1}^{N} \text{trace}\{\boldsymbol{\Theta}_j\mathbf{G}_{j(k)}^T\} - \sum_{l=1}^{n} \text{trace}\{\boldsymbol{\Lambda}_l\mathbf{U}_l^T\}$$

Since the Karush-Kuhn-Tucker (KKT) condition states that complementary slackness (i.e. $[\boldsymbol{\Lambda}_k]_{a,b}[\mathbf{U}_k]_{a,b} = 0$) must hold, we can take the partial derivative of the Lagrangian $\mathcal{L}$ with respect to $\mathbf{U}_k$ and equate it to zero to obtain an updating rule. In particular, the derivative $\partial\mathcal{L}/\partial\mathbf{U}_k$ is determined by

$$\frac{\partial\mathcal{L}}{\partial\mathbf{U}_k} = \sum_{i=1}^{N}(\mathbf{X}_{i(k)} - \mathbf{U}_k\mathbf{G}_{i(k)}\widetilde{\mathbf{U}}_k^T)\widetilde{\mathbf{U}}_k\mathbf{G}_{i(k)}^T - \boldsymbol{\Lambda}_k.$$

Hence, the relationship between each elements of $\mathbf{U}_k^{\text{old}}$ and $\mathbf{U}_k^{\text{new}}$ can be expressed as

$$[\mathbf{U}_k]^{\text{old}}_{a,b}[\sum_{i=1}^{N}\mathbf{X}_{i(k)}\widetilde{\mathbf{U}}_k\mathbf{G}_{i(k)}^T]_{a,b} = [\mathbf{U}_k]^{\text{new}}_{a,b}[\mathbf{U}_k^{\text{old}}\sum_{i=1}^{N}\mathbf{G}_{i(k)}\widetilde{\mathbf{U}}_k^T\widetilde{\mathbf{U}}_k\mathbf{G}_{i(k)}^T]_{a,b}$$

Thus,

$$[\mathbf{U}_k]^{\text{new}}_{a,b} = \frac{[\mathbf{U}_k]^{\text{old}}_{a,b}[\sum_{i=1}^{N}\mathbf{X}_{i(k)}\widetilde{\mathbf{U}}_k\mathbf{G}_{i(k)}^T]_{a,b}}{[\mathbf{U}_k^{\text{old}}\sum_{i=1}^{N}\mathbf{G}_{i(k)}\widetilde{\mathbf{U}}_k^T\widetilde{\mathbf{U}}_k\mathbf{G}_{i(k)}^T]_{a,b}}.$$

In other words, by denoting $\mathbf{S}_k^i = \mathbf{G}_{i(k)}\widetilde{\mathbf{U}}_k^T$, we have the updating rule as

$$\mathbf{U}_k^{\text{new}} = \mathbf{U}_k^{\text{old}} \circledast \frac{\sum_{i=1}^{N}\mathbf{X}_{i(k)}\mathbf{S}_k^{i\,T}}{\mathbf{U}_k^{\text{old}}\sum_{i=1}^{N}\mathbf{S}_k^i\mathbf{S}_k^{i\,T}}, \tag{I4}$$

where $\circledast$ represents the element-wise matrix product.

When we apply NTD (e.g. an NTD algorithm proposed by Kim *et al.* in [72, Table 5]) on $\boldsymbol{\mathcal{X}}$, where $\boldsymbol{\mathcal{X}} = \boldsymbol{\mathcal{X}}_1 \boxplus \boldsymbol{\mathcal{X}}_2 \cdots \boxplus \boldsymbol{\mathcal{X}}_N$, we also obtain the similar update rule for computing nonnegative factors $\mathbf{U}_k$. In particular, in each iteration, $\mathbf{U}_k$ is updated as

$$\mathbf{U}_k^{\text{new}} = \mathbf{U}_k^{\text{old}} \circledast \frac{[\boldsymbol{\mathcal{X}}_{(k)}\mathbf{S}_k^T]}{\mathbf{U}_k^{\text{old}}\mathbf{S}_k\mathbf{S}_k^T}$$

where the k-mode unfolding matrices of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\mathcal{G}}$ can be expressed by

$$\mathbf{X}_{(k)} = [\mathbf{X}_{1(k)} \quad \mathbf{X}_{2(k)} \quad \dots \quad \mathbf{X}_{N(k)}],$$
$$\mathbf{G}_{(k)} = [\mathbf{G}_{1(k)} \quad \mathbf{G}_{2(k)} \quad \dots \quad \mathbf{G}_{N(k)}].$$

Meanwhile, the matrix $\mathbf{S}_k$ is defined by

$$\mathbf{S}_k = \mathbf{G}_{(k)}(\mathbf{I}_N \otimes \widetilde{\mathbf{U}}_k),$$

because the last factor $\mathbf{U}_{n+1}$ is full column-rank (i.e. $r_{n+1} = N$). Hence, these formulations can be expressed in term of sub-tensors as

$$\mathbf{X}_{(k)}\mathbf{S}_k^T = \sum_{i=1}^{N}\mathbf{X}_{i(k)}\mathbf{S}_k^{i\,T},$$
$$\mathbf{S}_k\mathbf{S}_k^T = \sum_{i=1}^{N}\mathbf{S}_k^i\mathbf{S}_k^{i\,T}.$$

In other words, we obtain the rule in (I4). This completes the proof of Proposition (5).

D. Performance of Other Classifiers

We present here the results of using KNN and NB and Decision Tree as the classifier in all the three experiments.

TABLE V: Detection performance of the CP-KNN model using Leave-One-Out Cross-Validation (LOOCV)

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	0	8	14748	397	0.0000	0.9738	0.9733	0.8550
2	635	20484	163	472	20392	92	0.2567	0.9955	0.9733	0.8569
3	6	14975	0	6	14731	244	0.0000	0.9837	0.9833	0.7009
4	16	30751	3	13	28421	2330	0.1875	0.9242	0.9238	0.6202
5	351	25916	121	230	25748	168	0.3447	0.9935	0.9848	0.8210
6	22	44387	2	20	43713	674	0.0909	0.9848	0.9844	0.7182
7	2	2036	0	2	1917	119	0.0000	0.9416	0.9406	0.6636
8	11	29351	3	8	29127	224	0.2727	0.9924	0.9921	0.8145
9	1	3742	0	1	3686	56	0.0000	0.985	0.9848	0.6740
10	8	2371	0	8	2328	43	0.0000	0.9819	0.9786	0.7756
11	2	1565	0	2	1503	62	0.0000	0.9604	0.9592	0.5163
12	3	53302	0	3	51156	2146	0.0000	0.9597	0.9597	0.6038
13	5	69583	0	5	68806	777	0.0000	0.9888	0.9888	0.7653
14	8	6217	2	6	5879	338	0.2500	0.9456	0.9447	0.5804
15	324	11219	102	222	10360	859	0.3148	0.9234	0.9064	0.5453
16	28	35495	0	28	21211	2004	0.0000	0.9137	0.9126	0.4419
17	12	21170	1	11	20672	498	0.0833	0.9765	0.9760	0.7559

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.1059 $\pm$ 0.1324	0.1059	0.2753	0.0100	0.0209	0.1255 $\pm$ 0.1071
SPE :	0.9662 $\pm$ 0.0270	0.9662	0.9662	0.9706	0.9637	0.9667 $\pm$ 0.0028
ACC :	0.9627 $\pm$ 0.0275	0.9627	0.9680	0.9628	0.9695	0.9658 $\pm$ 0.0035
AUC :	0.6888 $\pm$ 0.1235	0.6888	0.6888	0.6873	0.6634	0.6820 $\pm$ 0.0124

TABLE VI: Detection performance of the NCP-KNN model using Leave-One-Out Cross-Validation (LOOCV)

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	1	7	14882	263	0.1250	0.9826	0.9822	0.8957
2	635	20484	167	468	20433	51	0.2630	0.9975	0.9754	0.7048
3	6	14975	0	6	14719	256	0.0000	0.9829	0.9825	0.8455
4	16	30751	1	15	27662	3089	0.0625	0.8995	0.8991	0.7479
5	351	25916	240	111	25408	508	0.6838	0.9804	0.9764	0.9724
6	22	44387	5	17	43004	1383	0.2273	0.9688	0.9685	0.8718
7	2	2036	0	2	1908	128	0.0000	0.9371	0.9362	0.5624
8	11	29351	0	11	27697	1654	0.0000	0.9436	0.9433	0.7774
9	1	3742	0	1	3604	138	0.0000	0.9631	0.9629	0.7803
10	8	2371	2	6	2357	14	0.2500	0.9941	0.9916	0.9025
11	2	1565	1	1	1514	51	0.5000	0.9674	0.9668	0.6463
12	3	53302	0	3	48159	5143	0.0000	0.9035	0.9035	0.4584
13	5	69583	0	5	67924	1659	0.0000	0.9762	0.9761	0.8241
14	8	6217	1	7	5819	398	0.1250	0.936	0.9349	0.6792
15	324	11219	92	232	10514	705	0.2840	0.9372	0.9188	0.5659
16	28	35495	4	24	21294	1921	0.1429	0.9173	0.9163	0.4932
17	12	21170	1	11	19878	1292	0.0833	0.9309	0.9385	0.7316

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.1616 $\pm$ 0.1937	0.1616	0.3571	0.1555	0.0801	0.1886 $\pm$ 0.1183
SPE :	0.9545 $\pm$ 0.0305	0.9545	0.9503	0.9557	0.9616	0.9555 $\pm$ 0.0046
ACC :	0.9514 $\pm$ 0.0296	0.9514	0.9480	0.9525	0.9543	0.9515 $\pm$ 0.0026
AUC :	0.7329 $\pm$ 0.1496	0.7329	0.7329	0.7273	0.7103	0.7258 $\pm$ 0.0107

TABLE VII: Detection performance of the TD-KNN model using Leave-One-Out Cross-Validation (LOOCV)

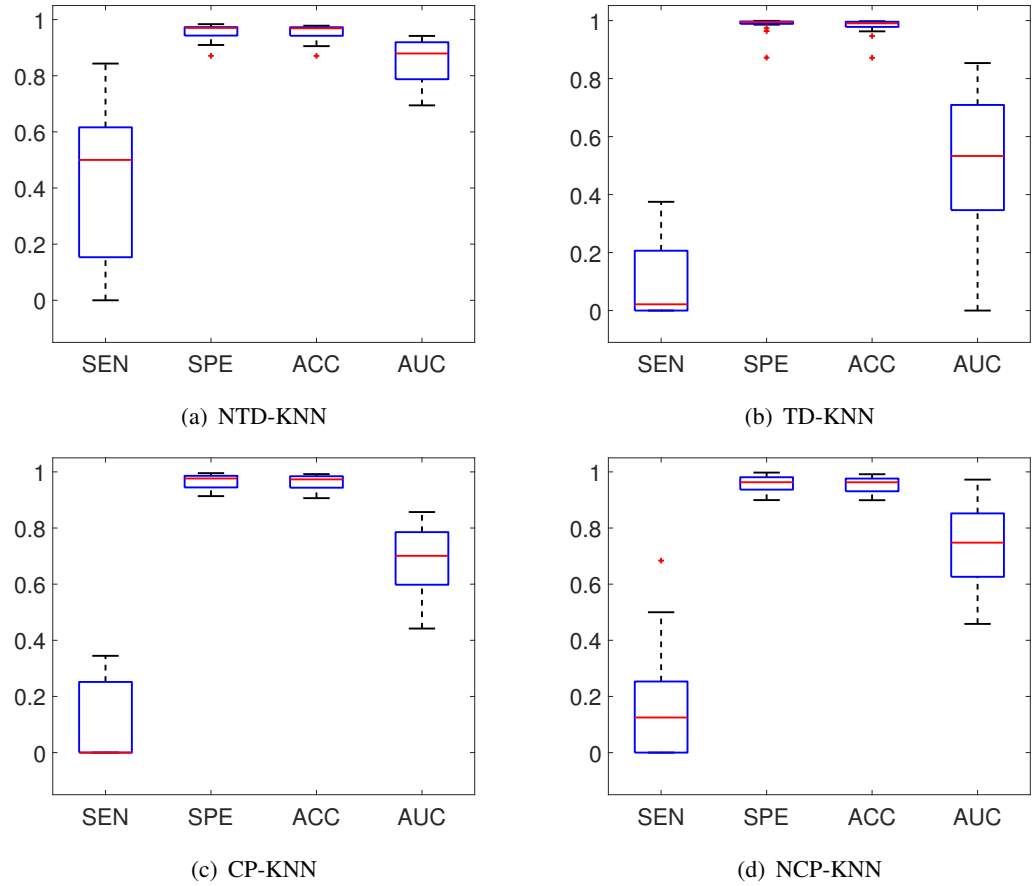
Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	3	5	15116	29	0.3750	0.9981	0.9978	0.8535
2	635	20484	14	621	20466	18	0.0220	0.9991	0.9697	0.4599
3	6	14975	0	6	14927	48	0.0000	0.9968	0.9964	0.3217
4	16	30751	0	16	30478	273	0.0000	0.9911	0.9906	0.5332
5	351	25916	70	281	25701	215	0.1994	0.9917	0.9811	0.4932
6	22	44387	5	17	43766	621	0.2273	0.9860	0.9856	0.7975
7	2	2036	0	2	2026	10	0.0000	0.9951	0.9941	0.4481
8	11	29351	1	10	29226	125	0.0909	0.9957	0.9954	0.7765
9	1	3742	0	1	3733	9	0.0000	0.9976	0.9973	0.0000
10	8	2371	2	6	2361	10	0.2500	0.9958	0.9933	0.6114
11	2	1565	0	2	1549	16	0.0000	0.9898	0.9885	0.6834
12	3	53302	0	3	52757	545	0.0000	0.9898	0.9897	0.2712
13	5	69583	0	5	69412	171	0.0000	0.9975	0.9975	0.7713
14	8	6217	3	5	5424	793	0.3750	0.8724	0.8718	0.6884
15	324	11219	7	317	10921	298	0.0216	0.9734	0.9467	0.2968
16	28	35495	3	25	22378	837	0.1071	0.9639	0.9629	0.3546
17	12	21170	0	12	21094	76	0.0000	0.9964	0.9958	0.5533

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.0981 $\pm$ 0.1351	0.0981	0.0749	0.1080	0.0436	0.0812 $\pm$ 0.0286
SPE :	0.9841 $\pm$ 0.0303	0.9841	0.9891	0.9830	0.9853	0.9854 $\pm$ 0.0026
ACC :	0.9797 $\pm$ 0.0312	0.9797	0.9856	0.9786	0.9897	0.9834 $\pm$ 0.0052
AUC :	0.5244 $\pm$ 0.2291	0.5244	0.5244	0.5144	0.3424	0.4764 $\pm$ 0.0895

TABLE VIII: Detection performance of the NTD-KNN model using Leave-One-Out Cross-Validation (LOOCV)

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	5	3	14740	405	0.6250	0.9733	0.9731	0.9261
2	635	20484	366	269	20150	334	0.5764	0.9837	0.9714	0.9358
3	6	14975	4	2	14569	406	0.6667	0.9728	0.9728	0.8665
4	16	30751	8	8	26780	3971	0.5000	0.8709	0.8707	0.7667
5	351	25916	296	55	25220	696	0.8433	0.9731	0.9714	0.9418
6	22	44387	10	12	42591	1796	0.4545	0.9595	0.9593	0.8919
7	2	2036	0	2	1909	127	0.0000	0.9376	0.9367	0.8320
8	11	29351	2	9	28558	793	0.1818	0.9730	0.9727	0.8982
9	1	3742	0	1	3361	81	0.0000	0.9784	0.9781	0.9049
10	8	2371	7	1	2301	70	0.1250	0.9705	0.9676	0.9123
11	2	1565	1	1	1518	47	0.5000	0.9700	0.9694	0.7201
12	3	53302	0	3	51625	1677	0.0000	0.9685	0.9685	0.8083
13	5	69583	1	4	67927	1656	0.2000	0.9762	0.9761	0.9261
14	8	6217	4	4	5896	321	0.500	0.9484	0.9478	0.8379
15	324	11219	247	77	10204	1015	0.7623	0.9095	0.9054	0.7592
16	28	35495	17	11	21568	1647	0.6071	0.9291	0.9287	0.6944
17	12	21170	4	8	20671	499	0.3333	0.9764	0.9761	0.9147

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.4045 $\pm$ 0.2743	0.4045	0.6741	0.3765	0.1546	0.4024 $\pm$ 0.2128
SPE :	0.9571 $\pm$ 0.0299	0.9571	0.9578	0.9586	0.9601	0.9584 $\pm$ 0.0013
ACC :	0.9556 $\pm$ 0.0297	0.9556	0.95567	0.9559	0.9653	0.9584 $\pm$ 0.0046
AUC :	0.8551 $\pm$ 0.0794	0.8551	0.8551	0.8555	0.8594	0.8563 $\pm$ 0.0021



(a) NTD-KNN (b) TD-KNN

(c) CP-KNN (d) NCP-KNN

Fig. 14: Detection performance of the KNN model when applying our feature extraction against using different tensor-based approaches (CP, NCP, TD and NTD).

TABLE IX: Detection performance of the CP-NB model using Leave-One-Out Cross-Validation (LOOCV)

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	1	7	12995	2150	0.1250	0.8580	0.8577	0.7262
2	635	20484	377	258	18757	1727	0.5937	0.9157	0.9060	0.8165
3	6	14975	1	5	10712	4263	0.1667	0.7153	0.7151	0.3110
4	16	30751	8	8	19439	11312	0.5000	0.6321	0.6321	0.5727
5	351	25916	295	56	22103	3813	0.8405	0.8529	0.8527	0.9217
6	22	44387	9	13	32628	11759	0.4091	0.7351	0.7349	0.6218
7	2	2036	1	1	1376	660	0.5000	0.6758	0.6757	0.5421
8	11	29351	8	3	24033	5318	0.7273	0.8188	0.8188	0.8745
9	1	3742	1	0	2713	1029	1.0000	0.7250	0.7251	0.9281
10	8	2371	5	3	1870	501	0.6250	0.7887	0.7881	0.7072
11	2	1565	2	0	825	740	1.0000	0.5272	0.5278	0.9387
12	3	53302	1	2	32770	20532	0.3333	0.6148	0.6148	0.5130
13	5	69583	4	1	61514	8069	0.8000	0.8840	0.8840	0.9249
14	8	6217	4	4	3391	2826	0.5000	0.5454	0.5454	0.5234
15	324	11219	217	107	6332	4887	0.6698	0.5644	0.5674	0.6400
16	28	35495	9	19	10385	12830	0.3214	0.4473	0.4472	0.3414
17	12	21170	2	10	16138	5032	0.1667	0.7623	0.7620	0.3523

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.5458 $\pm$ 0.2745	0.5458	0.6553	0.5729	0.7343	0.6271 $\pm$ 0.0853
SPE :	0.7096 $\pm$ 0.1377	0.7096	0.7404	0.7092	0.6758	0.7088 $\pm$ 0.0264
ACC :	0.7091 $\pm$ 0.1366	0.7091	0.7401	0.7086	0.6939	0.7129 $\pm$ 0.0075
AUC :	0.6621 $\pm$ 0.2161	0.6621	0.6621	0.6832	0.7659	0.6933 $\pm$ 0.0494

TABLE X: Detection performance of the NCP-NB model using Leave-One-Out Cross-Validation (LOOCV)

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	1	7	13617	1528	0.1250	0.8991	0.8987	0.6159
2	635	20484	402	233	15901	4583	0.6331	0.7763	0.772	0.7374
3	6	14975	0	6	12779	2196	0.0000	0.8534	0.853	0.3297
4	16	30751	6	10	23095	7656	0.3750	0.7510	0.7508	0.6114
5	351	25916	321	30	25337	579	0.9145	0.9777	0.9768	0.9813
6	22	44387	11	11	38381	6006	0.5000	0.8647	0.8645	0.8340
7	2	2036	0	2	1169	867	0.0000	0.5742	0.5736	0.4163
8	11	29351	6	5	22922	6429	0.5455	0.7810	0.7809	0.7049
9	1	3742	1	0	2954	788	1.0000	0.7894	0.7895	0.7902
10	8	2371	4	4	2137	234	0.5000	0.9013	0.9000	0.7241
11	2	1565	2	0	1000	565	1.0000	0.639	0.6394	0.9677
12	3	53302	0	3	24824	28478	0.0000	0.4657	0.4657	0.1129
13	5	69583	0	5	57649	11934	0.0000	0.8285	0.8284	0.3434
14	8	6217	0	8	4272	1945	0.0000	0.6871	0.6863	0.3500
15	324	11219	78	246	6589	4630	0.2407	0.5873	0.5776	0.4762
16	28	35495	6	22	11542	11673	0.2143	0.4972	0.4968	0.3080
17	12	21170	2	10	15586	5584	0.1667	0.7362	0.7359	0.4528

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.3656 $\pm$ 0.3580	0.3656	0.5825	0.3882	0.5509	0.4718 $\pm$ 0.1107
SPE :	0.7417 $\pm$ 0.1474	0.7417	0.7452	0.7369	0.7203	0.7360 $\pm$ 0.0110
ACC :	0.7406 $\pm$ 0.1478	0.7406	0.7445	0.7358	0.7163	0.7343 $\pm$ 0.012s5
AUC :	0.5739 $\pm$ 0.2500	0.5739	0.5739	0.5800	0.6209	0.5872 $\pm$ 0.0226

TABLE XI: Detection performance of the TD-NB model using Leave-One-Out Cross-Validation (LOOCV)

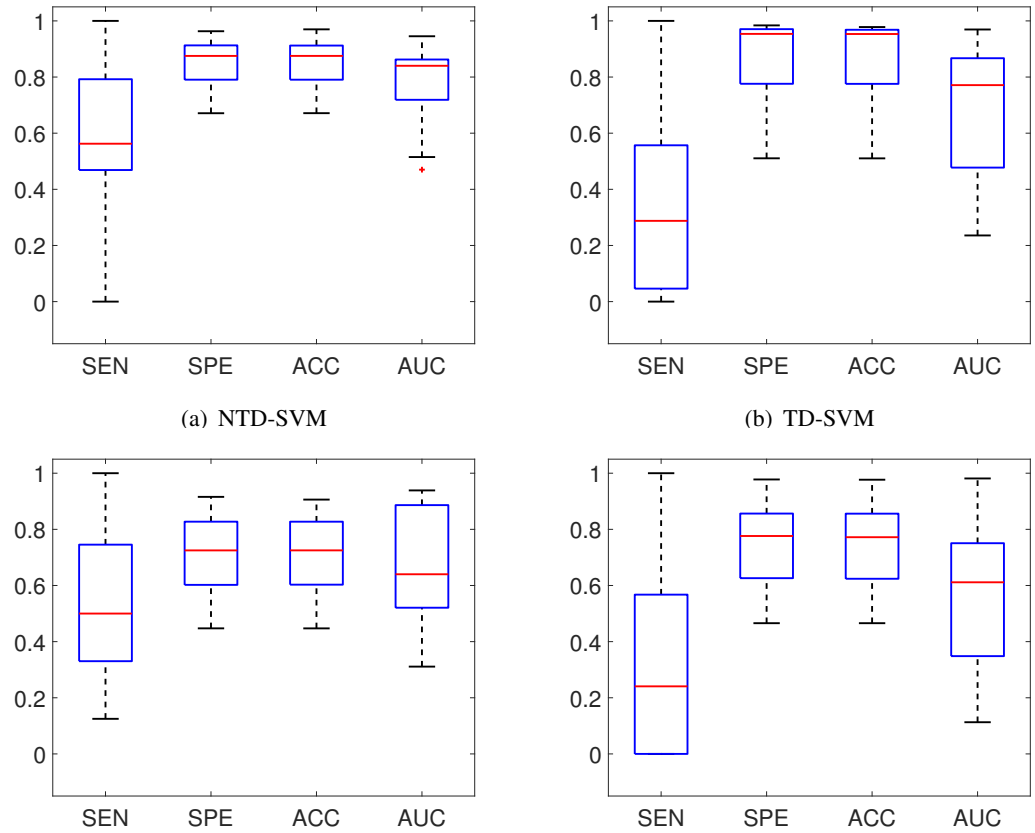
Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	5	3	14662	483	0.6250	0.9681	0.9679	0.9315
2	635	20484	166	469	20156	328	0.2614	0.9840	0.9623	0.8412
3	6	14975	3	3	14527	448	0.5	0.9701	0.9699	0.7712
4	16	30751	3	13	29328	1423	0.1875	0.9537	0.9533	0.7635
5	351	25916	101	250	14541	11375	0.2877	0.5611	0.5574	0.3448
6	22	44387	13	9	43148	1239	0.5909	0.9721	0.9719	0.833
7	2	2036	0	2	1851	185	0.0000	0.9091	0.9082	0.7665
8	11	29351	6	5	27959	1392	0.5455	0.9526	0.9524	0.9164
9	1	3742	0	1	3661	81	0	0.9784	0.9781	0.2357
10	8	2371	5	3	2322	49	0.6250	0.9793	0.9781	0.9694
11	2	1565	2	0	1072	493	1.0000	0.6850	0.6854	0.9447
12	3	53302	0	3	43366	9936	0.0000	0.8136	0.8135	0.4956
13	5	69583	0	5	67254	2329	0.0000	0.9665	0.9665	0.8506
14	8	6217	3	5	5013	1204	0.375	0.8063	0.8058	0.7058
15	324	11219	20	304	7332	3887	0.0617	0.6535	0.6369	0.2996
16	28	35495	10	18	11851	11364	0.3571	0.5105	0.5103	0.4228
17	12	21170	3	9	20194	976	0.2500	0.9539	0.9535	0.8393

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.3333 $\pm$ 0.2861	0.3333	0.2357	0.3287	0.2145	0.2781 $\pm$ 0.0618
SPE :	0.8599 $\pm$ 0.1600	0.8599	0.8743	0.8562	0.8675	0.8645 $\pm$ 0.0081
ACC :	0.8571 $\pm$ 0.1607	0.8571	0.8718	0.8534	0.9059	0.8721 $\pm$ 0.0239
AUC :	0.7019 $\pm$ 0.2433	0.7019	0.7019	0.6879	0.5748	0.6666 $\pm$0.0616

TABLE XII: Detection performance of the NTD-NB model using Leave-One-Out Cross-Validation (LOOCV)

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	4	4	13580	1565	0.5000	0.8967	0.8965	0.8466
2	635	20484	453	182	19731	753	0.7134	0.9632	0.9699	0.9221
3	6	14975	6	0	12969	2006	1.0000	0.8660	0.8661	0.9389
4	16	30751	7	9	23615	7136	0.4375	0.7679	0.7678	0.6463
5	351	25916	323	28	23274	2642	0.9202	0.8981	0.8984	0.9453
6	22	44387	16	6	39262	5125	0.7273	0.8845	0.8845	0.8512
7	2	2036	1	1	1698	338	0.5000	0.8340	0.8337	0.8530
8	11	29351	7	4	27548	1803	0.6364	0.9386	0.9385	0.8501
9	1	3742	0	1	3408	334	0.0000	0.9107	0.9105	0.4701
10	8	2371	5	3	2169	202	0.6250	0.9148	0.9138	0.6768
11	2	1565	1	1	1132	433	0.5000	0.7233	0.723	0.515
12	3	53302	0	3	43353	9949	0.0000	0.8133	0.8133	0.761
13	5	69583	0	5	64485	5098	0.0000	0.9267	0.9267	0.8719
14	8	6217	4	4	5121	1096	0.5000	0.8237	0.8233	0.7923
15	324	11219	278	46	7939	3280	0.8580	0.7076	0.7119	0.8336
16	28	35495	24	4	15579	7636	0.8571	0.6711	0.6713	0.8064
17	12	21170	9	3	19348	1822	0.7500	0.9139	0.9138	0.9206

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.5603 $\pm$ 0.3131	0.5603	0.7892	0.5353	0.2700	0.5387 $\pm$ 0.2145
SPE :	0.8502 $\pm$ 0.0873	0.8502	0.8636	0.8519	0.8399	0.8514 $\pm$ 0.0097
ACC :	0.8508 $\pm$ 0.0873	0.8508	0.8633	0.8524	0.8618	0.8571 $\pm$ 0.0064
AUC :	0.7942 $\pm$ 0.1406	0.7942	0.7942	0.7826	0.6517	0.7557 $\pm$ 0.0695



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

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

Fig. 15: Detection performance of the NB model when applying our feature extraction against using different tensor-based approaches (CP, NCP, TD and NTD).

TABLE XIII: Detection performance of the CP-DT 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	12902	2243	0.8750	0.8519	0.8519	0.9063
2	635	20484	424	211	18829	1655	0.6677	0.9192	0.9116	0.8297
3	6	14975	4	2	12250	2725	0.6667	0.8180	0.8180	0.8869
4	16	30751	13	3	25093	5658	0.8125	0.8160	0.8160	0.8721
5	351	25916	300	51	22939	2977	0.8547	0.8851	0.8847	0.9039
6	22	44387	21	1	38397	5990	0.9545	0.8651	0.8651	0.9266
7	2	2036	2	0	1717	319	1.0000	0.8433	0.8435	0.896
8	11	29351	11	0	25072	4279	1.0000	0.8542	0.8543	0.9242
9	1	3742	1	0	3169	573	1.0000	0.8469	0.8469	0.862
10	8	2371	8	0	2100	271	1.0000	0.8857	0.8861	0.9493
11	2	1565	2	0	1246	319	1.0000	0.7962	0.7964	0.8839
12	3	53302	2	1	40766	12536	0.6667	0.7648	0.7648	0.8458
13	5	69583	0	5	53968	15615	0.000 0	0.7756	0.7755	0.4446
14	8	6217	8	0	4850	1367	1.0000	0.7801	0.7804	0.8867
15	324	11219	261	63	9147	2072	0.8056	0.8153	0.8150	0.8381
16	28	35495	23	5	17586	5629	0.8214	0.7575	0.7576	0.8171
17	12	21170	11	1	17275	3895	0.9167	0.8160	0.8161	0.8989

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.8260 $\pm$ 0.2460	0.8260	0.7614	0.8352	0.9002	0.8307 $\pm$ 0.0568
SPE :	0.8289 $\pm$ 0.0459	0.8289	0.8185	0.8309	0.8373	0.8289 $\pm$ 0.0078
ACC :	0.8285 $\pm$ 0.0450	0.8285	0.8183	0.8304	0.8276	0.8262 $\pm$ 0.0054
AUC :	0.8572 $\pm$ 0.1122	0.8572	0.8572	0.8555	0.8553	0.8563 $\pm$ 0.0010

TABLE XIV: Detection performance of the NCP-DT model using Leave-One-Out Cross-Validation (LOOCV)

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	6	2	13056	2089	0.7500	0.8621	0.8620	0.8967
2	635	20484	430	205	18851	1633	0.6772	0.9203	0.9130	0.8067
3	6	14975	6	0	12890	2085	1.0000	0.8608	0.8608	0.9355
4	16	30751	16	0	25254	5497	1.0000	0.8212	0.8213	0.9085
5	351	25916	326	25	23836	2080	0.9288	0.9197	0.9199	0.9452
6	22	44387	20	2	37682	6705	0.9091	0.8489	0.8490	0.9168
7	2	2036	1	1	1599	437	0.5000	0.7854	0.7851	0.8390
8	11	29351	10	1	24405	4946	0.9091	0.8315	0.8315	0.8761
9	1	3742	1	0	3197	545	1.0000	0.8544	0.8544	0.9526
10	8	2371	7	1	2114	257	0.8750	0.8916	0.8916	0.9505
11	2	1565	2	0	1238	327	1.0000	0.7911	0.7913	0.8717
12	3	53302	2	1	40255	13047	0.6667	0.7552	0.7552	0.7975
13	5	69583	5	0	61616	7967	1.0000	0.8855	0.8855	0.9321
14	8	6217	8	0	5045	1172	1.0000	0.8115	0.8117	0.8905
15	324	11219	257	67	8979	2240	0.7932	0.8003	0.8001	0.8307
16	28	35495	27	1	17620	5595	0.9643	0.7590	0.7592	0.8830
17	12	21170	10	2	16798	4372	0.8333	0.7935	0.7935	0.8659

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.8710 $\pm$ 0.1487	0.8710	0.7864	0.8670	0.8743	0.8497 $\pm$ 0.0423
SPE :	0.8348 $\pm$ 0.0515	0.8348	0.8375	0.8364	0.8290	0.8344 $\pm$ 0.0038
ACC :	0.8344 $\pm$ 0.0505	0.8344	0.8373	0.8359	0.8277	0.8339 $\pm$ 0.0043
AUC :	0.8882 $\pm$ 0.0489	0.8882	0.8882	0.8892	0.9025	0.8920 $\pm$ 0.0070

TABLE XV: Detection performance of the TD-DT model using Leave-One-Out Cross-Validation (LOOCV)

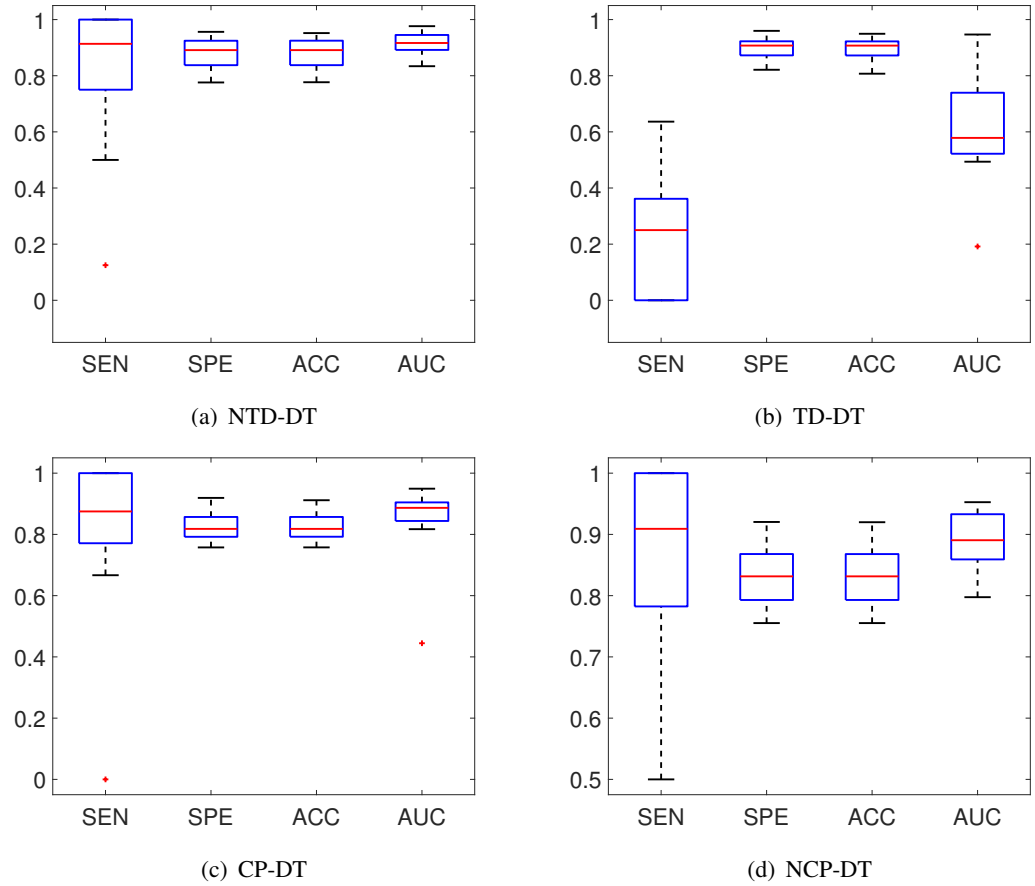
Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	3	5	13686	1459	0.3750	0.9037	0.9034	0.5924
2	635	20484	133	502	19660	824	0.2094	0.9598	0.9372	0.5106
3	6	14975	0	6	13791	1184	0.0000	0.9209	0.9206	0.8288
4	16	30751	2	14	27938	2813	0.1250	0.9085	0.9081	0.6403
5	351	25916	105	246	21762	4154	0.2991	0.8397	0.8325	0.5257
6	22	44387	4	18	40453	3934	0.1818	0.9114	0.9110	0.5243
7	2	2036	0	2	1837	199	0.0000	0.9023	0.9014	0.5153
8	11	29351	7	4	26632	2719	0.6364	0.9074	0.9073	0.7275
9	1	3742	0	1	3504	238	0.0000	0.9364	0.9361	0.8055
10	8	2371	2	6	2143	228	0.2500	0.9038	0.9016	0.5785
11	2	1565	0	2	1321	244	0.0000	0.8441	0.843	0.1917
12	3	53302	0	3	47012	6290	0.0000	0.8820	0.8819	0.4937
13	5	69583	2	3	66063	3520	0.4000	0.9494	0.9494	0.775
14	8	6217	5	3	5768	449	0.6250	0.9278	0.9274	0.9468
15	324	11219	95	229	9223	1996	0.2932	0.8221	0.8072	0.5729
16	28	35495	10	18	19065	4150	0.3571	0.8212	0.8207	0.5722
17	12	21170	4	8	19330	1840	0.3333	0.9131	0.9128	0.6100

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.2403 $\pm$ 0.2067	0.2403	0.2580	0.2481	0.0963	0.2107 $\pm$ 0.0766
SPE :	0.8973 $\pm$ 0.0420	0.8973	0.9035	0.8979	0.8950	0.8984 $\pm$ 0.0036
ACC :	0.8942 $\pm$ 0.0427	0.8942	0.9010	0.8948	0.9101	0.9000 $\pm$ 0.0074
AUC :	0.6124 $\pm$ 0.1709	0.6124	0.6124	0.6176	0.6423	0.8338 $\pm$ 0.0043

TABLE XVI: Detection performance of the NTD-DT model using Leave-One-Out Cross-Validation (LOOCV)

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	8	0	13497	1648	1.0000	0.8912	0.8912	0.9569
2	635	20484	494	141	19588	896	0.7780	0.9563	0.9509	0.8955
3	6	14975	3	3	13989	986	0.5000	0.9342	0.9340	0.8585
4	16	30751	2	14	26069	4682	0.1250	0.8477	0.8478	0.9370
5	351	25916	333	18	23879	2037	0.9487	0.9214	0.9218	0.9286
6	22	44387	19	3	39000	5387	0.8636	0.8786	0.8786	0.9331
7	2	2036	1	1	1939	97	0.5000	0.9524	0.9519	0.8507
8	11	29351	10	1	26641	2710	0.9091	0.9077	0.9077	0.9450
9	1	3742	1	0	3298	444	1.0000	0.8813	0.8814	0.8991
10	8	2371	8	0	2221	150	1.0000	0.9367	0.9369	0.9765
11	2	1565	2	0	127	348	1.0000	0.7776	0.7779	0.9128
12	3	53302	2	1	43019	10283	0.6667	0.8071	0.8071	0.8812
13	5	69583	5	0	63460	6123	1.0000	0.9120	0.9120	0.9677
14	8	6217	7	1	5312	905	0.8750	0.8544	0.8545	0.9166
15	324	11219	296	28	8760	2513	0.9136	0.7760	0.7799	0.8339
16	28	35495	27	1	18029	5186	0.9643	0.7766	0.7768	0.8963
17	12	21170	12	0	18917	2253	1.0000	0.8936	0.8936	0.9456

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.8261 $\pm$ 0.2457	0.8261	0.8530	0.8482	0.8622	0.8474 $\pm$ 0.0153
SPE :	0.8768 $\pm$ 0.0612	0.8768	0.8730	0.8780	0.8836	0.8778 $\pm$ 0.0044
ACC :	0.8767 $\pm$ 0.0603	0.8767	0.8729	0.8778	0.8787	0.8765 $\pm$ 0.0026
AUC :	0.9138 $\pm$ 0.0412	0.9138	0.9138	0.9139	0.9022	0.9110 $\pm$ 0.0058



(a) NTD-DT (b) TD-DT

(c) CP-DT (d) NCP-DT

Fig. 16: Detection performance of the DT model when applying our feature extraction against using different tensor-based approaches (CP, NCP, TD and NTD).

TABLE XVII: Detection performance of the NTD-SVM model based on Phan-Cichocki Method

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	5	3	14269	876	0.6250	0.9422	0.9420	0.8417
2	635	20484	165	470	20045	439	0.2598	0.9786	0.9570	0.7528
3	6	14975	4	2	802	802	0.6667	0.9464	0.9463	0.8373
4	16	30751	0	16	28848	1903	0.0000	0.9381	0.9376	0.5292
5	351	25916	244	107	24185	1731	0.6952	0.9332	0.9300	0.8507
6	22	44387	3	19	43512	875	0.1364	0.9803	0.9399	0.7548
7	2	2036	1	1	1933	103	0.5000	0.9494	0.9490	0.7846
8	11	29351	2	9	27569	1782	0.1818	0.9393	0.9390	0.6607
9	1	3742	0	1	3463	279	0.0000	0.9254	0.9252	0.8399
10	8	2371	4	4	2316	55	0.5000	0.9768	0.9752	0.8404
11	2	1565	1	1	1424	141	0.5000	0.9099	0.9094	0.7613
12	3	53302	2	1	50222	3080	0.6667	0.9422	0.9422	0.9487
13	5	69583	2	3	68465	1118	0.4000	0.9839	0.9839	0.9846
14	8	6217	5	3	5724	493	0.6250	0.9207	0.9203	0.9360
15	324	11219	112	212	9634	1585	0.3457	0.8587	0.8443	0.8064
16	28	35495	14	14	20336	2879	0.5000	0.8760	0.8755	0.8004
17	12	21170	4	8	19608	1562	0.3333	0.9262	0.9259	0.8492

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.4080 $\pm$ 0.2283	0.4080	0.3939	0.4033	0.3091	0.3786 $\pm$ 0.0467
SPE :	0.9369 $\pm$ 0.0343	0.9369	0.9119	0.9375	0.9379	0.9311 $\pm$ 0.0128
ACC :	0.9319 $\pm$ 0.0332	0.9319	0.9099	0.9326	0.9348	0.9273 $\pm$ 0.0116
AUC :	0.8105 $\pm$ 0.1069	0.8105	0.8105	0.8137	0.8331	0.8170 $\pm$ 0.0109

TABLE XVIII: Detection performance of the NTD-NB model based on Phan-Cichocki Method

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	5	3	13657	1488	0.6250	0.9017	0.9016	0.8546
2	635	20484	238	397	19441	1043	0.3748	0.9491	0.9318	0.7958
3	6	14975	4	2	12914	2061	0.6667	0.8624	0.8623	0.7522
4	16	30751	4	12	24562	6189	0.2500	0.7987	0.7985	0.6128
5	351	25916	140	211	11746	14170	0.3989	0.4532	0.4525	0.4311
6	22	44387	11	11	35750	8637	0.5000	0.8054	0.8053	0.7281
7	2	2036	1	1	1271	765	0.5000	0.6243	0.6241	0.6221
8	11	29351	7	4	27092	2259	0.6364	0.9230	0.9229	0.8374
9	1	3742	0	1	3231	511	0.0000	0.8634	0.8632	0.1937
10	8	2371	2	6	2212	159	0.2500	0.9329	0.9306	0.6327
11	2	1565	1	1	996	569	0.5000	0.6364	0.6362	0.5267
12	3	53302	0	3	41139	12163	0.0000	0.7718	0.7718	0.5317
13	5	69583	4	1	51251	18332	0.8000	0.7365	0.7365	0.8449
14	8	6217	0	8	4902	1315	0.0000	0.7885	0.7875	0.4538
15	324	11219	52	272	7264	3955	0.1605	0.6475	0.6338	0.4787
16	28	35495	15	13	16035	7180	0.5357	0.6907	0.6905	0.6222
17	12	21170	5	7	18790	2380	0.4167	0.8876	0.8873	0.8475

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.3891 $\pm$ 0.2456	0.3891	0.3391	0.3921	0.2642	0.3461 $\pm$ 0.0598
SPE :	0.7808 $\pm$ 0.1353	0.7808	0.7785	0.7914	0.7635	0.7785 $\pm$ 0.0115
ACC :	0.7786 $\pm$ 0.1348	0.7786	0.7768	0.7890	0.7873	0.7829 $\pm$ 0.0061
AUC :	0.6333 $\pm$ 0.1843	0.6333	0.6333	0.6251	0.4663	0.5895 $\pm$ 0.0822

TABLE XIX: Detection performance of the NTD-KNN model based on Phan-Cichocki Method

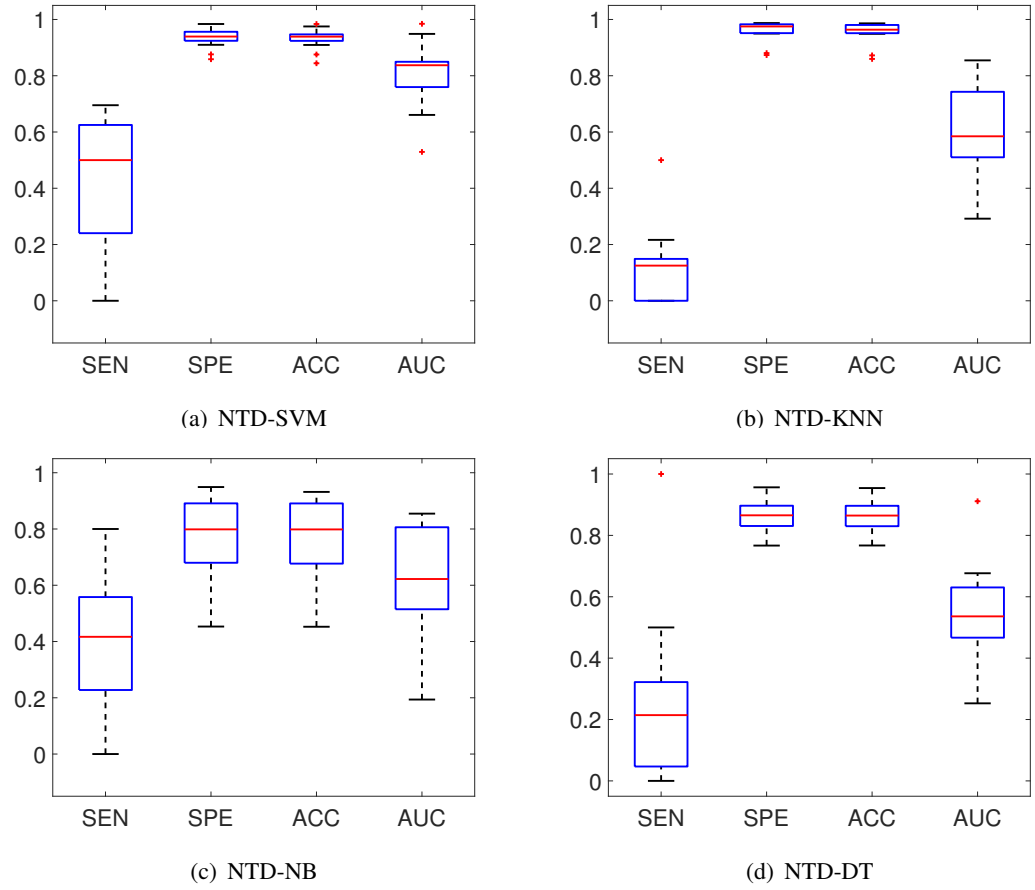
Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	4	4	14774	371	0.5000	0.9755	0.9753	0.8168
2	635	20484	24	611	20224	260	0.0378	0.9873	0.9588	0.6726
3	6	14975	0	6	14774	201	0.0000	0.9866	0.9862	0.5693
4	16	30751	1	15	29298	1453	0.0625	0.9527	0.9523	0.3919
5	351	25916	76	275	25438	478	0.2165	0.9816	0.9713	0.4157
6	22	44387	3	19	43512	875	0.1364	0.9803	0.9799	0.5830
7	2	2036	0	2	1964	72	0.0000	0.9646	0.9637	0.2919
8	11	29351	2	9	28835	516	0.1818	0.9824	0.9821	0.7541
9	1	3742	0	1	3561	181	0.0000	0.9516	0.9514	0.8546
10	8	2371	1	7	2339	32	0.1250	0.9865	0.9836	0.8147
11	2	1565	0	2	1367	198	0.0000	0.8735	0.8724	0.6294
12	3	53302	0	3	50682	2620	0.0000	0.9508	0.9508	0.5142
13	5	69583	0	5	68476	1107	0.0000	0.9841	0.9840	0.5847
14	8	6217	1	7	6063	154	0.1250	0.9752	0.9741	0.5880
15	324	11219	46	278	9878	1341	0.1420	0.8805	0.8597	0.4977
16	28	35495	14	24	22052	1163	0.1429	0.9499	0.9489	0.5603
17	12	21170	2	10	20401	769	0.1667	0.9637	0.9632	0.7391

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.1080 $\pm$ 0.1263	0.1080	0.1137	0.1073	0.0376	0.0917 $\pm$ 0.0362
SPE :	0.9604 $\pm$ 0.0342	0.9604	0.9459	0.9623	0.9490	0.9544 $\pm$ 0.0082
ACC :	0.9563 $\pm$ 0.0354	0.9563	0.9427	0.9543	0.9501	0.9509 $\pm$ 0.0060
AUC :	0.6046 $\pm$ 0.1581	0.6046	0.6046	0.6046	0.6629	0.6221 $\pm$ 0.0278

TABLE XX: Detection performance of the NTD-DT model based on Phan-Cichocki Method

Pat.	Spikes	Non-Spikes	TP	FP	TN	FN	SEN	SPE	ACC	AUC
1	8	15145	0	8	13095	2050	0.0000	0.8646	0.8642	0.3471
2	635	20484	113	522	19557	927	0.1780	0.9547	0.9314	0.6468
3	6	14975	3	3	12962	2013	0.5000	0.8656	0.8654	0.6766
4	16	30751	1	15	25835	4916	0.0625	0.8401	0.8397	0.4694
5	351	25916	141	210	22717	3199	0.4017	0.8766	0.8702	0.5361
6	22	44387	7	15	38052	6335	0.3182	0.8573	0.8570	0.6301
7	2	2036	0	2	1667	369	0.0000	0.8188	0.8180	0.2527
8	11	29351	3	8	26466	2885	0.2727	0.9017	0.9015	0.6313
9	1	3742	0	1	3349	393	0.0000	0.8950	0.8947	0.3799
10	8	2371	2	6	2268	103	0.2500	0.9566	0.9542	0.6296
11	2	1565	2	0	1200	365	1.0000	0.7668	0.7671	0.9112
12	3	53302	1	2	43466	9836	0.3333	0.8155	0.8154	0.4931
13	5	69583	0	5	63971	5612	0.0000	0.9193	0.9193	0.4586
14	8	6217	2	6	5380	837	0.2500	0.8654	0.8646	0.5709
15	324	11219	37	287	9069	2150	0.1142	0.8084	0.7889	0.5120
16	28	35495	6	22	19376	3839	0.2143	0.8346	0.8339	0.5517
17	12	21170	2	10	18437	2733	0.1667	0.8709	0.8705	0.4987

Average Performance:	Mean $\pm$ S.D.	ρ_{AM}	ρ_{TA}	ρ_{TWA}	ρ_{TEW}	$\bar{\rho} \pm$ S.D.
SEN :	0.2389 $\pm$ 0.2469	0.2389	0.2219	0.2214	0.1862	0.2171 $\pm$ 0.0222
SPE :	0.8654 $\pm$ 0.0505	0.8654	0.8707	0.8701	0.8586	0.8662 $\pm$ 0.0056
ACC :	0.8621 $\pm$ 0.0495	0.8621	0.8682	0.8667	0.8612	0.8646 $\pm$ 0.0034
AUC :	0.5409 $\pm$ 0.1490	0.5409	0.5409	0.5339	0.4701	0.5214 $\pm$ 0.0344



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

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

Fig. 17: Detection performance of the four classifiers using Phan-Cichocki's features.

References

[1] A. T. Berg and C. P. Panayiotopoulos, *Atlas of Epilepsies*, 1st ed. Springer-Verlag London, 2010.

[2] J. Gotman, "Automatic recognition of epileptic seizures in the EEG," *Electroencephalography and Clinical Neurophysiology*, vol. 54, no. 5, pp. 530–540, 1982.

[3] ——, "Automatic detection of seizures and spikes," *Journal of Clinical Neurophysiology*, vol. 16, no. 2, pp. 130–140, 1999.

[4] 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., 2012, ch. 4, pp. 75–98.

[5] L. Orosco, "A survey of performance and techniques for automatic epilepsy detection," *Journal of Medical and Biological Engineering*, vol. 33, no. 6, pp. 526–537, 2013.

[6] U. R. Acharya, S. V. Sree, G. Swapna, R. J. Martis, and J. S. Suri, "Automated EEG analysis of epilepsy: A review," *Knowledge-Based Systems*, vol. 45, pp. 147–165, 2013.

[7] H. S. Liu, T. Zhang, and F. S. Yang, "A multistage, multimethod approach for automatic detection and classification of epileptiform EEG," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 12, pp. 1557–1566, 2002.

[8] N. Acır, İ. Öztura, B. Baklan, and C. Güzeliş, "Automatic detection of epileptiform events in EEG by a three-stage procedure based on artificial neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 1, pp. 30–40, 2005.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] M. Ontivero-Ortega, Y. Garcia-Puente, and E. Martínez-Montes, "Comparison of classifiers to detect epileptic seizures via PARAFAC decomposition," in *VI Latin American Congress on Biomedical Engineering, 29-31 October, 2014, Parana, Argentina. IFMBE Proceedings.*, vol. 49. Springer, 2015, pp. 500–503.

[15] L. Spyrou, S. Kouchaki, and S. Sanei, "Multiview classification of brain data through tensor factorisation," in *International Workshop on Machine Learning for Signal Processing*. IEEE, 2015, pp. 1–6.

[16] Y. R. Aldana, B. Hunyadi, E. J. M. Reyes, V. R. Rodriguez, and S. V. Huffel, "Nonconvulsive epileptic seizure detection in scalp EEG using multiway data analysis," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–12, 2018.

[17] 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.

[18] 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*, vol. 4, no. 2, pp. 1–15, 2016.

[19] B. Hunyadi, P. Dupont, W. Van Paesschen, and S. Van Huffel, "Tensor decompositions and data fusion in epileptic electroencephalography and functional magnetic resonance imaging data," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 7, no. 1, p. e1197, 2017.

[20] I. Markovsky, *Low Rank Approximation*. Springer, 2012.

[21] A. Cichocki, N. Lee, I. Oseledets, A.-H. Phan, Q. Zhao, D. P. Mandic *et al.*, "Tensor networks for dimensionality reduction and large-scale optimization: Part 1 low-rank tensor decompositions," *Foundations and Trends® in Machine Learning*, vol. 9, no. 4-5, pp. 249–429, 2016.

[22] N. Kishore Kumar and J. Schneider, "Literature survey on low rank approximation of matrices," *Linear and Multilinear Algebra*, vol. 65, no. 11, pp. 2212–2244, 2017.

[23] C. Ding and J. Ye, "2-dimensional singular value decomposition for 2D maps and images," in *SIAM International Conference on Data Mining*, 2005, pp. 32–43.

[24] K. Inoue and K. Urahama, "DSVD: A tensor-based image compression and recognition method," in *IEEE International Symposium on Circuits and Systems*. IEEE, 2005, pp. 6308–6311.

[25] J. Yang, D. Zhang, A. F. Frangi, and J. yu Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, Jan 2004.

[26] D. Zhang and Z.-H. Zhou, "(2D) 2PCA: Two-directional two-dimensional PCA for efficient face representation and recognition," *Neurocomputing*, vol. 69, no. 1-3, pp. 224–231, 2005.

[27] C. M. Crainiceanu, B. S. Caffo, S. Luo, V. M. Zipunnikov, and N. M. Punjabi, "Population Value Decomposition, a Framework for the Analysis of Image Populations," *Journal of the American Statistical Association*, vol. 106, no. 495, pp. 775–790, 2011.
[28] J. Ye, "Generalized low rank approximations of matrices," *Machine Learning*, vol. 61, no. 1-3, pp. 167–191, 2005.
[29] J. Liu and S. Chen, "Non-iterative generalized low rank approximation of matrices," *Pattern Recognition Letters*, vol. 27, no. 9, pp. 1002–1008, 2006.
[30] J. Liu, S. Chen, Z. Zhou, and X. Tan, "Generalized Low-Rank Approximations of Matrices Revisited," *IEEE Transactions on Neural Networks*, vol. 21, no. 4, pp. 621–632, April 2010.
[31] K. Inoue, K. Hara, and K. Urahama, "Symmetric generalized low rank approximations of matrices," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2012, pp. 949–952.
[32] Jiarong, Y. Wei, and Z. X. Shi, "Robust Generalized Low Rank Approximations of Matrices," *PLOS ONE*, vol. 10, no. 9, pp. 1–23, 09 2015.
[33] J. Ye, R. Janardan, and Q. Li, "Two-dimensional linear discriminant analysis," in *Advances in Neural Information Processing Systems*, 2005, pp. 1569–1576.
[34] K. Van Deun, A. K. Smilde, M. J. van der Werf, H. A. Kiers, and I. Van Mechelen, "A structured overview of simultaneous component based data integration," *Bmc Bioinformatics*, vol. 10, no. 1, p. 246, 2009.
[35] A. Stegeman, "Simultaneous Component Analysis by Means of Tucker3," *Psychometrika*, vol. 83, no. 1, pp. 21–47, Mar 2018.
[36] X. Li, M. K. Ng, G. Cong, Y. Ye, and Q. Wu, "MR-NTD: Manifold Regularization Nonnegative Tucker Decomposition for Tensor Data Dimension Reduction and Representation," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 8, pp. 1787–1800, Aug 2017.
[37] D. Xu, S. Yan, L. Zhang, S. Lin, H.-J. Zhang, and T. S. Huang, "Reconstruction and recognition of tensor-based objects with concurrent subspaces analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 1, pp. 36–47, 2008.
[38] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, and H. Zhang, "Multilinear Discriminant Analysis for Face Recognition," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 212–220, Jan 2007.
[39] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "MPCA: Multilinear Principal Component Analysis of Tensor Objects," *IEEE Transactions on Neural Networks*, vol. 19, no. 1, pp. 18–39, Jan 2008.
[40] Y. Panagakis, C. Kotropoulos, and G. R. Arce, "Non-negative multilinear principal component analysis of auditory temporal modulations for music genre classification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 3, pp. 576–588, 2010.
[41] Z. Lai, Y. Xu, Q. Chen, J. Yang, and D. Zhang, "Multilinear Sparse Principal Component Analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 25, no. 10, pp. 1942–1950, Oct 2014.
[42] N. T. Anh-Dao, L. T. Thanh, N. Linh-Trung, and H. V. Le, "Nonnegative tensor decomposition for EEG epileptic spike detection," in *NAFOSTED Conference on Information and Computer Science (NICS)*. IEEE, 2018.
[43] A. Cichocki, R. Zdunek, A. H. Phan, and S. ichi Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.
[44] A. H. Phan and A. Cichocki, "Tensor decompositions for feature extraction and classification of high dimensional datasets," *Nonlinear theory and its applications, IEICE*, vol. 1, no. 1, pp. 37–68, 2010.
[45] M. Boussé, N. Vervliet, I. Domanov, O. Debals, and L. De Lathauwer, "Linear systems with a canonical polyadic decomposition constrained solution: Algorithms and applications," *Numerical Linear Algebra with Applications*, vol. 25, no. 6, p. e2190, e2190 nla.2190.
[46] M. Boussé, G. Goovaerts, N. Vervliet, O. Debals, S. V. Huffel, and L. D. Lathauwer, "Irregular heartbeat classification using kronecker product equations," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2017, pp. 438–441.
[47] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
[48] R. A. Harshman, "Foundations of the PARAFAC procedure: Models and conditions for an" explanatory" multimodal factor analysis," *UCLA Working Papers in Phonetics*, vol. 16, no. 1-84, 1970. [Online]. Available: http://publish.uwo.ca
[49] L. De Lathauwer, B. De Moor, and J. Vandewalle, "On the best rank-1 and rank-(r1, r2,..., rn) approximation of higher-order tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 21, no. 4, pp. 1324–1342, 2000.
[50] B. N. Sheehan and Y. Saad, "Higher order orthogonal iteration of tensors (HOOI) and its relation to PCA and GLRAM," in *SIAM International Conference on Data Mining*. SIAM, 2007, pp. 355–365.
[51] I. Jeon, E. E. Papalexakis, U. Kang, and C. Faloutsos, "HaTen2: Billion-scale tensor decompositions," in *2015 IEEE 31st International Conference on Data Engineering*, April 2015, pp. 1047–1058.
[52] W. Austin, G. Ballard, and T. G. Kolda, "Parallel Tensor Compression for Large-Scale Scientific Data," in *2016 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, May 2016, pp. 912–922.
[53] J. Oh, K. Shin, E. E. Papalexakis, C. Faloutsos, and H. Yu, "S-HOT: Scalable High-Order Tucker Decomposition," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*. ACM, 2017, pp. 761–770.
[54] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, "A review on ensembles for the class imbalance problem: Bagging-, boosting-, and hybrid-based approaches," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 42, no. 4, pp. 463–484, July 2012.
[55] D. Tax, "One-class classification," *PhD thesis, Delft University of Technology*, 2001.
[56] S. S. Khan and M. G. Madden, "One-class classification: taxonomy of study and review of techniques," *The Knowledge Engineering Review*, vol. 29, no. 3, pp. 345–374, 2014.
[57] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. John Wiley & Sons, 2012.

[58] A. Cichocki, D. Mandic, L. D. 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, March 2015.
[59] H. S. Liu, T. Zhang, and F. S. Yang, "A multistage, multimethod approach for automatic detection and classification of epileptiform eeg," *IEEE Transactions on biomedical engineering*, vol. 49, no. 12, pp. 1557–1566, 2002.
[60] N. Acir, I. Oztura, M. Kuntalp, B. Baklan, and C. Guzelis, "Automatic detection of epileptiform events in eeg by a three-stage procedure based on artificial neural networks," *IEEE Transactions on Biomedical Engineering*, vol. 52, no. 1, pp. 30–40, 2005.
[61] Y.-C. Liu, C.-C. Lin, J.-J. Tsai, and Y.-N. Sun, "Model-based spike detection of epileptic eeg data," *Sensors*, vol. 13, no. 9, pp. 12 536–12 547, 2013.
[62] L. Xuyen, L. Thanh, D. Viet, T. Long, N. Trung, and N. Thuan, "Deep learning for epileptic spike detection," *VNU Journal of Science: Computer Science and Communication Engineering*, vol. 33, no. 2, 2018.
[63] A. J. Casson, E. Luna, and E. Rodriguez-Villegas, "Performance metrics for the accurate characterisation of interictal spike detection algorithms," *Journal of neuroscience methods*, vol. 177, no. 2, pp. 479–487, 2009.
[64] G. Van Belle, L. D. Fisher, P. J. Heagerty, and T. Lumley, *Biostatistics: A Methodology for the Health Sciences*. John Wiley & Sons, 2004, vol. 519.
[65] D. Wang, Y. Zhu, T. Ristaniemi, and F. Cong, "Extracting multi-mode ERP features using fifth-order nonnegative tensor decomposition," *Journal of Neuroscience Methods*, vol. 308, pp. 240–247, 2018.
[66] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newsletter*, vol. 11, no. 1, pp. 10–18, 2009.
[67] A. Zare, A. Ozdemir, M. A. Iwen, and S. Aviyente, "Extension of PCA to Higher Order Data Structures: An Introduction to Tensors, Tensor Decompositions, and Tensor PCA," *Proceedings of the IEEE*, vol. 106, no. 8, pp. 1341–1358, Aug 2018.
[68] A. K. Smilde, J. A. Westerhuis, and S. de Jong, "A framework for sequential multiblock component methods," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 17, no. 6, pp. 323–337, 2003.
[69] H. A. Kiers and J. M. ten Berge, "Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple simultaneous structure," *British Journal of mathematical and statistical psychology*, vol. 47, no. 1, pp. 109–126, 1994.
[70] G. Zhou, A. Cichocki, and S. Xie, "Fast Nonnegative Matrix/Tensor Factorization Based on Low-Rank Approximation," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2928–2940, June 2012.
[71] E. Kokiopoulou, J. Chen, and Y. Saad, "Trace optimization and eigenproblems in dimension reduction methods," *Numerical Linear Algebra with Applications*, vol. 18, no. 3, pp. 565–602, 2011.
[72] Y.-D. Kim and S. Choi, "Nonnegative tucker decomposition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1–8.