



Original Article

New feature selection method for multi-channel EEG epileptic spike detection system

Nguyen Thi Anh Dao^{1,2}, Le Trung Thanh¹, Nguyen Viet Dung¹,
Nguyen Linh Trung^{1,*}, Le Vu Ha¹

¹ AVITECH, VNU University of Engineering and Technology, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam

² University of Technology and Logistics, Ho Town, Thuan Thanh, Bac Ninh, Vietnam

Received 22 March 2019

Revised 19 September 2019, Accepted 30 September 2019

Abstract: Epilepsy is one of the most common brain disorders. Electroencephalogram (EEG) is widely used in epilepsy diagnosis and treatment, with it the epileptic spikes can be observed. Tensor decomposition-based feature extraction has been proposed to facilitate automatic detection of EEG epileptic spikes. However, tensor decomposition may still result in a large number of features which are considered negligible in determining expected output performance. We proposed a new feature selection method that combines the Fisher score and p -value feature selection methods to rank the features by using the longest common sequences (LCS) to separate epileptic and non-epileptic spikes. The proposed method significantly outperformed several state-of-the-art feature selection methods.

Keywords: Electroencephalogram, EEG, epileptic spikes, tensor decomposition, feature extraction, feature selection.

1. Introduction

Epilepsy is a severe neurological disorder and is one of the most common brain disorders, accounting for 1% of all human diseases. According to a study in 2010 [1], there are about 50 millions people worldwide suffering from epilepsy, among them about 40 millions live in developing countries and 80 – 90% of

these people are not treated [2, 3]. Vietnam is one of those countries with a high incidence of epilepsy. According to [4], 0.44% of the Vietnam population are affected by epilepsy.

In epilepsy diagnosis and treatment, doctors often rely on observed seizure or epileptiform patterns (such as shape and density of spikes, sharp waves, and spike-wave complexes) in the electroencephalogram (EEG) of patients to determine the type of epilepsy and the affected area of the brain.

In recent years, there have been many studies on automatic detection of epileptic spikes [5–

* Corresponding author.

E-mail address: linhtrung@vnu.edu.vn

<https://doi.org/10.25073/2588-1086/vnucsce.230>

13]. These automatic epileptic spike detection methods mostly analyze EEG data on a single channel at a time. In reality, epileptic spikes on adjacent channels are likely to occur at the same time. Therefore, simultaneous multi-channel processing of EEG signals allows exploitation of the spatial correlation between epileptic spikes for improving the efficiency of epileptic spike detection.

While raw multi-channel EEG signals are two-dimensional, multi-channel EEG data can be represented by tensors of higher dimensions, with the dimensions correspond to such domains as time, frequency, scale, channel, object, group, etc. Tensor analysis has been utilized for automatic seizure detection [14–18]. An approach for automatic epileptic spike detection based on tensor decomposition was proposed in [19].

The purpose of tensor decomposition in multi-channel EEG signal processing is for feature extraction: the EEG data is reduced to a set of feature vectors. Another step, called feature selection, may be needed to further reduce the size of the feature vectors. A number of algorithms have been proposed for addressing the problem of feature selection so far. Recent surveys on feature selection are found in [20–25]. According to selection strategy perspective, feature selection algorithms can be categorized into three groups: filter, wrapper and embedded methods [20]. Filtering methods rank the features and then select the features that have high ranking scores before feeding them into learning algorithms. In the methods of the wrapper group, the features are scored using a learning algorithm, while in the embedded methods feature selection is incorporated with the training process. It is note that the filter methods are independent of any learning algorithms, while feature selection methods in the two latter groups rely highly on performance of learning algorithms for measuring the relevance of features. Feature

selection methods may be categorized into three groups: supervised, unsupervised, and semi-supervised methods. Supervised feature selections are generally for the problems of classification and regression. The main idea is to select a subset of extracted features that can maximize the relevance to the label information or regression targets [20, 21]. Unsupervised feature selections are generally for clustering problems. Different from supervised methods, they usually look for alternatives to evaluate feature relevance from unlabeled data such as the locality/variance preserving ability [26, 27]. Semi-supervised feature selections aim to utilize both labeled and unlabeled data [25]. The algorithms in this group often exploit the label information of labeled data and data distribution of unlabeled data to evaluate the important of features [28]. These methods are widely used in applications of machine learning [21, 23] and pattern recognition [29, 30], including EEG signal classification [31–34]. In [31], Garrett *et al.* proposed a feature selection method based on genetic algorithms and successfully applied it to EEG during finger movement. Maryann *et al.* used hybrid feature selection for seizure prediction focused on precursors [32]. Robert Jenke *et al.* used not only multivariate feature selection methods but also univariate selection methods for emotion recognition from EEG [33]. John Atkinson *et al.* combined a mutual information-based feature selection method and kernel classifiers in order to enhance the accuracy of the emotion classification [34]. Although these methods improve more or less the performance of EEG classifications, they do not fully consider the combination of different feature selection methods which may further improve the overall accuracy of the classifiers and detectors.

In [35], a multi-channel system for EEG epileptic spike detection base on tensor decomposition was proposed. The resulting set of features, however, is highly redundant in

determining the expected output (e.g., detected epileptic spikes). This motivates us to look for a feature selection model relevant to EEG epileptic datasets. We proposed a new method of feature selection that combines Fisher score and p -value to rank the features by using longest common sequences (LCS). The proposed method was compared with several well-known methods, including: Fisher score [36] and Laplacian score [37], Unsupervised Discriminative Feature Selection (UDFS) [38], Infinite Latent Feature Selection (ILFS) [39], and Local Learning-based Clustering Feature Selection (LLCFS) [40]. To the best of our knowledge, this study is the first work aiming to combine two widely used feature selection methods to enhance the effectiveness of dimensionality reduction in the problem of EEG classification.

The paper is organized as follows. Section 2 provides the background on tensor decomposition and our recently proposed multi-channel EEG epileptic spike detection. The proposed method is described in Section 3. Section 4 shows experimental results and discussions of the results. Finally, Section 5 concludes the paper.

2. Preliminaries

2.1. Notations and Tensor Decomposition

The notations of mathematical symbols used in this paper are listed in Table 1 [35]. A tensor is a generalization of vectors, matrices and can be seen as a multidimensional array [41]. Similar to matrix decomposition, tensor decomposition factorizes a tensor into a set of matrices called loading factors, and one small core tensor. Two well-known decomposition models are canonical decomposition (CP)¹ and Tucker. The main

¹Canonical decomposition is also called parallel factor analysis (PARAFAC).

Table 1: Mathematical Symbols

$a, \mathbf{a}, \mathbf{A}, \mathcal{A}$	scalar, vector, matrix and tensor
\mathbf{A}^T	the transpose of \mathbf{A}
\mathbf{A}^\dagger	the pseudo-inverse of \mathbf{A}
$\mathbf{A}_{(k)}$	the mode- k unfolding of \mathcal{A}
$\ \mathcal{A}\ _F$	the Frobenius norm of \mathcal{A}
\otimes	the Hadamand product
\oslash	the devision of two matrices
$\mathbf{A} \otimes \mathbf{B}$	the Kronecker product of \mathbf{A} and \mathbf{B}
$\mathcal{A} \times_k \mathbf{U}$	the k -mode product of \mathcal{A} with a matrix \mathbf{U}
$\mathcal{A} \boxplus \mathcal{B}$	the concatenation of \mathcal{A} and \mathcal{B}
$\langle \mathcal{A}, \mathcal{B} \rangle$	the inner product of \mathcal{A} and \mathcal{B}

difference is that the former yields a diagonal core tensor, while the latter does not require a diagonal core, but a set of orthogonal factors. Decomposition of an n -way tensor can be mathematically formulated as follows:

$$\mathcal{X} = \mathcal{G} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \cdots \times_n \mathbf{U}_n, \quad (1)$$

where $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \cdots \times I_n}$ is the decomposing tensor, $\mathcal{G} \in \mathbb{R}^{r_1 \times r_2 \cdots \times r_n}$ is the decomposed core tensor of \mathcal{X} , and $\{\mathbf{U}_i\}_{i=1}^n, \mathbf{U}_i \in \mathbb{R}^{I_i \times r_i}$ are the set of decomposed orthogonal factors.

In this work, we focus on nonnegative Tucker decomposition (NTD) in which both the core tensor \mathcal{G} and orthogonal factors \mathbf{U}_i are required to be nonnegative. In particular, NTD can be stated as the following minimization problem:

$$\begin{aligned} \min_{\mathcal{G}, \mathbf{U}_i} \|\mathcal{X} - \mathcal{G} \times_1 \mathbf{U}_1 \cdots \times_n \mathbf{U}_n\|_F^2 \\ \text{s.t. } \mathcal{G} \geq 0, \mathbf{U}_i \geq 0, \forall i = 1, 2, \dots, n. \end{aligned} \quad (2)$$

The solution of (2) can be obtained by using alternative minimization in which a variable (e.g., factor \mathbf{U}_1) is optimized while the others are kept fixed. We here re-introduce a standard NTD algorithm [42], which is used in our recently proposed multi-channel EEG epileptic spike detection system [35]. Particularly, the

objective function of (2) can be reformulated as

$$\begin{cases} \arg \min_{\mathbf{U}_{i \geq 0}} f_{\mathbf{U}} = \frac{1}{2} \sum_{j=1}^n \|\mathbf{X}_{(j)} - \mathbf{U}_j \mathbf{S}_j\|_F^2, \\ \arg \min_{\mathbf{G}_{\geq 0}} f_{\mathbf{G}} = \frac{1}{2} \|\text{vec}(\mathcal{X}) - \mathbf{F} \text{vec}(\mathcal{G})\|_2^2, \end{cases}$$

with $\mathbf{F} = \otimes \mathbf{U}_j$. The update rules for estimating the factors and the core tensor are given by

$$\begin{cases} \mathbf{U}_i = \mathbf{U}_i - \alpha \otimes \frac{\partial f_{\mathbf{U}}}{\partial \mathbf{U}_i}, \\ \mathcal{G} = \mathcal{G} - \alpha \otimes \frac{\partial f_{\mathbf{G}}}{\partial \mathcal{G}}, \end{cases}$$

where the step size α is computed by $\alpha = \mathbf{U}_i \oslash (\mathbf{U}_i \mathbf{X}_{(i)} \mathbf{G}_{(i)}^T)$.

2.2. A Multi-channel EEG Epileptic Spike Detection System

In this work, we inherit our recently proposed multi-channel system for EEG epileptic spike detection in [35]. Assume that we have the pre-processed multi-channel EEG recording at hand and input it to the system. The system then processes it in four main stages: data representation, feature extraction, feature selection, and classification.

Data representation

In this stage, each multi-channel EEG segment of K channels and I data samples around a spike, which is labeled as epileptic or non-epileptic, are analyzed by the continuous wavelet transform (CWT). We then obtain a K time-frequency representation matrices of sizes $I \times J$ for an EEG segment, with J being the number of wavelet scales. These matrices are concatenated into a three-way EEG tensor $\mathcal{X} \in \mathbb{R}_+^{I \times J \times K}$ (i.e., time \times scale \times channel). EEG tensors formed from epileptic spikes are called epileptic tensors, \mathcal{X}^{ep} , and those from non-epileptic spikes are called non-epileptic tensors, \mathcal{X}^{nep} .

Feature Extraction

In this second stage, we aim to find a feature space \mathcal{F}^{ep} that can span the set of training epileptic spikes. After that, both epileptic and non-epileptic spikes are projected onto \mathcal{F}^{ep} to produce the discriminant features.

In particular, the stage consists of the following four steps. Firstly, we concatenate all N_1 training epileptic tensors $\mathcal{X}_1^{\text{ep}}, \dots, \mathcal{X}_{N_1}^{\text{ep}}$ into a single 4-way epileptic tensor $\tilde{\mathcal{X}}^{\text{ep}} \in \mathbb{R}_+^{I \times J \times K \times N_1}$ as follows:

$$\tilde{\mathcal{X}}^{\text{ep}} = \mathcal{X}_1^{\text{ep}} \boxplus \mathcal{X}_2^{\text{ep}} \boxplus \dots \boxplus \mathcal{X}_{N_1}^{\text{ep}}.$$

Secondly, the multilinear rank $[r_1, r_2, r_3]$ of the EEG tensor $\tilde{\mathcal{X}}^{\text{ep}}$ can be determined by solving the following problems for $i = 1, 2, 3$:

$$r_i \stackrel{\Delta}{=} \underset{r}{\text{argmin}} \|\mathbf{X}_{(i)} - \mathbf{U}^{I \times r} \mathbf{\Lambda}^{r \times r} \mathbf{V}^{r \times JK}\|_2^2.$$

Thanks to the truncated HOSVD [43], the rank r_i can be selected as the number of r_i top eigenvalues of the corresponding covariance matrix of $\tilde{\mathcal{X}}^{\text{ep}}$.

Thirdly, we use NTD to decompose $\tilde{\mathcal{X}}^{\text{ep}}$ into loading factors $\mathbf{A} \in \mathbb{R}_+^{I_1 \times r_1}$ in the time domain, $\mathbf{B} \in \mathbb{R}_+^{I_2 \times r_2}$ in the wavelet scale domain, and $\mathbf{C} \in \mathbb{R}_+^{I_3 \times r_3}$ in the spatial/channel domain, as

$$\tilde{\mathcal{X}}^{\text{ep}} \stackrel{\text{NTD}}{=} \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \times_4 \mathbf{D}. \quad (3)$$

The epileptic feature space is then given by

$$\mathcal{F} = \mathcal{G} \times_4 \mathbf{D}.$$

Finally, we project all training EEG tensors $\mathcal{X}_i^{\text{train}}$ onto the resulting epileptic feature space \mathcal{F}^{ep} to produce the discriminant feature vector

$$\mathbf{f}_i = \text{vec}(\mathcal{X}_i^{\text{train}} \times_1 \mathbf{A}^\dagger \times_2 \mathbf{B}^\dagger \times_3 \mathbf{C}^\dagger).$$

Feature Selection

In this third stage, we use the Fisher score, which is one of the most widely used method for

feature selection [36], used to rank features. Let \mathbf{F} be the set of features obtained by NTD,

$$\mathbf{F} = \{\mathbf{f}_{(i)}\}_{i=1}^{r_1 \cdot r_2 \cdot r_3}.$$

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 following ratio of between-class variation and within-class variation:

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

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

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

In feature selection, each feature is selected independently depending on its Fisher score so that the higher the score the more significant the feature is. After ranking all features based on their Fisher scores, the top l features with highest Fisher scores are selected to form the set of selected features $\mathbf{F}_{\text{Fisher}} = \{\mathbf{f}_{(1)}, \mathbf{f}_{(2)}, \dots, \mathbf{f}_{(l)} | \mathbf{f}_{(i)} \in \mathbf{F}, i = 1, \dots, l\}$, for later use in classification.

Classification

In this final stage, selected features are fed into a classifier producing a binary class label as its output, deciding if the underlying spike is epileptic or non-epileptic. Well-known classifiers can be used for this tasks, including support vector machine (SVM), k-nearest neighbor (KNN) and naive bayes (NB) model.

3. Proposed method

In this paper, we improve the multi-channel system for EEG epileptic spike detection

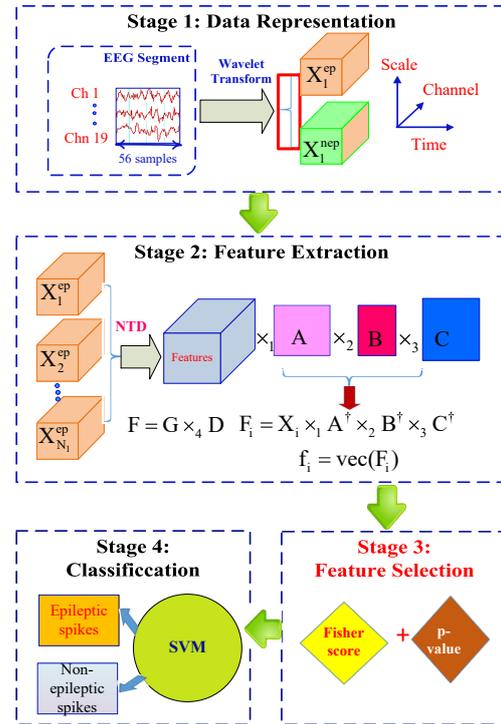


Fig. 1: Proposed combination of Fisher score and p -value for feature selection in the multi-channel EEG Epileptic Spike Detection System

proposed in [35] by replacing its feature selection algorithm (i.e., using the Fisher score) by a new method, which aims to combine two common feature selection methods– the Fisher score and the p -value–, to enhance the overall classification accuracy of the automatic spike detection system. The structure of the modified system is shown in Fig 1.

We exploit the fact that 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 $r = r_1 \cdot r_2 \cdot r_3$ 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. This motivates us to look for a model of feature selection relevant for EEG epileptic datasets.

In this stage, we apply the hypothesis testing [44] on each feature, and compare resulting p -values and Fisher scores [45] for each feature to assess the effectiveness of the classification. To select features, we propose to combine the Fisher scores and the p -values to rank the features by using the following selection rule: a more significant feature is one that has higher Fisher score and lower p -value. Since the Fisher score and p -value of each feature are calculated independently, it results in two separate sequences, of Fisher scores and of p -values. A solution to finding significant features is to first sort these sequences and then find the longest subsequence that is common to these two sorted sequences. The latter can be done by using the longest common subsequence (LCS) algorithm [46].

Assume that we have extracted n features from NTD, i.e., $\mathbf{F} = \{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_n\}$. Denote N_1 and N_2 the numbers of epileptic spikes and non-epileptic spikes, respectively. Denote Ω_1 and Ω_2 are the classes consisting these 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 . Then, the proposed feature selection method is composed of three main tasks. The first task is to rank the features by using their Fisher scores, as described in Section 2.2. The second task is to compute p -value for each feature \mathbf{f}_i . The third task is to combine Fisher scores and the p -values. Next, we will describe the second and the third tasks.

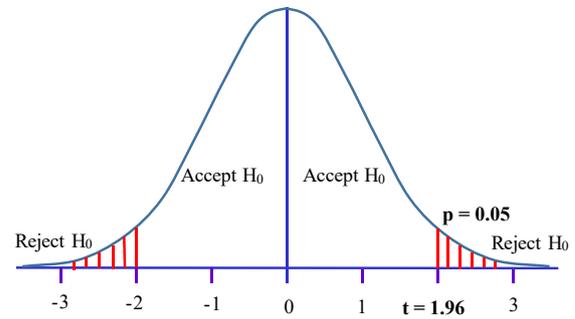


Fig. 2: A p -value is the probability of an observed result assuming that the null hypothesis H_0 is true.

Feature selection using p -values

In hypothesis testing, p -value (probability value) is the probability of observing a value as unlikely or more unlikely than the value of the test statistic when the null hypothesis is true [47], as shown in Fig. 2. The higher value of p , the lower the reliability of the result. A statistical significance level α is generally used to evaluate the results of hypothesis testing. When p is smaller than the significance level, we can have sufficient evidence to reject the hypothesis. In medical applications, α is often chosen at 0.05, 0.01, or 0.001 [44]. In this work, the null hypothesis H_0 states that there is no difference between the means of two groups (i.e., epileptic spikes and non-epileptic spikes). For each feature \mathbf{f}_i , the smaller the p -value of the feature the more significant the feature is. Given a value α , if $\alpha > p$ the test rejects the null hypothesis, and vice versa. The t -test value for each feature $\mathbf{f}_{(i)}$ can be computed as follows:

$$t(\mathbf{f}_{(i)}) = \frac{|\mu_{i,1} - \mu_{i,2}|}{\sqrt{\sigma_{i,1}^2/N_1 + \sigma_{i,2}^2/N_2}}. \quad (4)$$

The higher the t -test value, the higher the difference between the two means is. From the t -test value, the corresponding p -value is obtained

by using the T-tables [44]. Therefore, by sorting features according to their p -values, we obtain a set of significant features $\mathbf{F}_{p\text{-val}}$.

Feature selection using both Fisher scores and p -values

To find the longest common subsequence (LCS) of the two ranked feature sequences $\mathbf{F}_{\text{Fisher}}$ and $\mathbf{F}_{p\text{-val}}$ obtained from the above steps respectively based on Fisher score and p -value, we use a dynamic programming algorithm, as follows:

Let \mathbf{L} be a table such that each entry $\mathbf{L}(i, j)$ is the largest length of the common subsequence between $\mathbf{F}_{\text{Fisher}}^{(i)} \subset \mathbf{F}_{\text{Fisher}}$ and $\mathbf{F}_{p\text{-val}}^{(j)} \subset \mathbf{F}_{p\text{-val}}$, $i \leq l_1, j \leq l_2$, where l_1 and l_2 are the lengths of $\mathbf{F}_{\text{Fisher}}$ and $\mathbf{F}_{p\text{-val}}$, respectively. Since the solution for each subproblem $\mathbf{L}(i, j)$ depends on the preceding subproblems $\mathbf{L}(i - 1, j)$, $\mathbf{L}(i, j - 1)$, and $\mathbf{L}(i - 1, j - 1)$, the solution to finding the LCS corresponds is found by recursively solving the subproblems starting from $\mathbf{L}(0, 0)$, as follows

$$\mathbf{L}(i, j) = \begin{cases} \mathbf{L}(i - 1, j) + 1, & \mathbf{F}_{\text{Fisher}}^{(i)} = \mathbf{F}_{p\text{-val}}^{(j)}, \\ \max(\mathbf{L}(i - 1, j), \mathbf{L}(i, j - 1)), & \mathbf{F}_{\text{Fisher}}^{(i)} \neq \mathbf{F}_{p\text{-val}}^{(j)}. \end{cases}$$

with $\mathbf{L}(0, j) = \mathbf{L}(i, 0) = 0$.

As a result, $\mathbf{L}(l_1, l_2)$ is the largest length of the common sequence between $\mathbf{F}_{\text{Fisher}}$ and $\mathbf{F}_{p\text{-val}}$. After that, The LCS is established by tracking elements of the common sequence using table \mathbf{L} and the following rules:

- (i) if the neighbors of $\mathbf{L}(i, j)$ are identical, then they are appended to the LCS;
- (ii) otherwise, compare the values of $\mathbf{L}(i, j - 1)$ and $\mathbf{L}(i - 1, j)$ and follow the direction of the greater value.

4. Experimental results

4.1. EEG dataset

EEG data used in this study were recorded from 17 epilepsy patients of the National Pediatric Hospital using the 10 – 20 international standard with 19 EEG data channels, the sampling rate was 256Hz. Among these patients, there are 11 females and 6 males, with the youngest being 4-year-old and the oldest being 72-year-old. The total number of recorded epileptic spikes in the whole dataset is 1442 and the number of randomly selected non-epileptic spikes is 6114. Table 2 represents the details of the dataset.

The dataset is divided into two sets, including the training set and the testing set, using either the 10-fold cross-validation method or the leave-one-out cross-validation (LOOCV) method. In the 10-fold cross-validation method, the whole dataset is divided into 10 parts, one part is used for testing when the remaining 9 parts are for training. This partitioning process is repeated until all parts in dataset are tested. In the LOOCV method, in each testing case, the classifier model is fitted by using a training data composed of 16 patients and then tested by the remaining patient. The process is repeated until every patient in the dataset has been placed in the testing set once.

4.2. Evaluation metrics

To evaluate performance of a classifier, we use three widely used statistical evaluation metrics [48], including accuracy (ACC), sensitivity (SEN) and specificity (SPE).

True positive (TP) and false positive (FP) are the number of spikes that the doctor labels as epileptic spikes and non-epileptic spikes, respectively, while the system classifies both as epileptic spikes. True negative (TN) and false negative (FN) are the number of spikes that the doctor labels as epileptic spikes and non-

Table 2: EEG Dataset

Patient	Gender	Age	Duration	EPs/Non-EPs	Patient	Gender	Age	Duration	EPs/Non-EPs
1	Male	4	19m21s	8/393	10	Male	21	23m57s	8/274
2	Male	6	22m25s	635/193	11	Male	72	15m26s	2/117
3	Male	9	11m24s	6/188	12	Female	10	17m7s	3/582
4	Male	9	11m24s	16/453	13	Female	13	18m53s	5/514
5	Male	11	16m16s	351/816	14	Female	16	20m14s	8/76
6	Male	12	17m49s	22/602	15	Female	20	14m32s	324/202
7	Male	15	22m0s	2/50	16a	Female	22	17m 56s	19/156
8	Male	16	22m58s	11/589	16b	Female	22	9m 41s	9/216
9	Male	20	27m13s	1/75	17	Female	28	5m31s	12/618

EPs = Number of epileptic spikes; Non-EPs = Number of non-epileptic spikes.

epileptic spikes, while the system classifies as non-epileptic spikes.

ACC presents the proportion of the (epileptic and non-epileptic) spikes correctly classified over the total number of (epileptic and non-epileptic) spikes:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}}.$$

SEN measures the proportion of actual epileptic spikes that are correctly classified, as given by

$$\text{SEN} = \frac{\text{TP}}{\text{TP} + \text{FN}}.$$

SPE provides similar information as SEN but for non-epileptic spikes, as given by

$$\text{SPE} = \frac{\text{TN}}{\text{TN} + \text{FP}}.$$

In addition, the receiver operating characteristic (ROC) curve is also used to illustrate the performance of the system [49]. The curve is drawn by plotting the TP rate (equivalent to SEN defined above) and the FP rate ($1 - \text{SPE}$). As a result, the ROC curve allows us to derive a cost/benefit analysis for making decision. An key metric of ROC is the area under the ROC curve (AUC). AUC is used to compare

the performance of classifiers. Classifiers may have different ROC curves but if these curves have the same AUC values, then these classifiers are considered to have the same performance. Performance ranking based on AUC includes: [0.9–1] as excellent, [0.8–0.9] as good, [0.7–0.8] as fair, [0.6–0.7] as poor, and [0.5–0.6] as failed.

4.3. Results and discussions

The feature extraction method proposed in [19] is applied on this dataset, resulting in 1442 three-way epileptic tensors $\mathcal{X}^{ep} \in \mathbb{R}_+^{56 \times 20 \times 19}$ and 6114 three-way non-epileptic tensors $\mathcal{X}^{nep} \in \mathbb{R}_+^{56 \times 20 \times 19}$. Similar to [19], the rank components corresponding to time, frequency, and channel are determined as $r_1 = 15$, $r_2 = 10$, and $r_3 = 19$, respectively. The four-way epileptic tensor $\tilde{\mathcal{X}}^{ep} \in \mathbb{R}_+^{56 \times 20 \times 19 \times k}$ is constructed by concatenating these k three-way epileptic tensors. NTD is performed to obtain the common factors $\mathbf{A} \in \mathbb{R}_+^{56 \times 15}$, $\mathbf{B} \in \mathbb{R}_+^{20 \times 10}$, and $\mathbf{C} \in \mathbb{R}_+^{19 \times 19}$ of the training epileptic tensor $\tilde{\mathcal{X}}^{ep}$. The common factors of the training non-epileptic tensor are also obtained in a similar way.

The proposed feature selection method is compared with other state-of-the-art models mentioned in Section 1, including G-Fisher score, Laplacian score, UDFS, ILFS, and LLCFS, in terms of number of selected features and

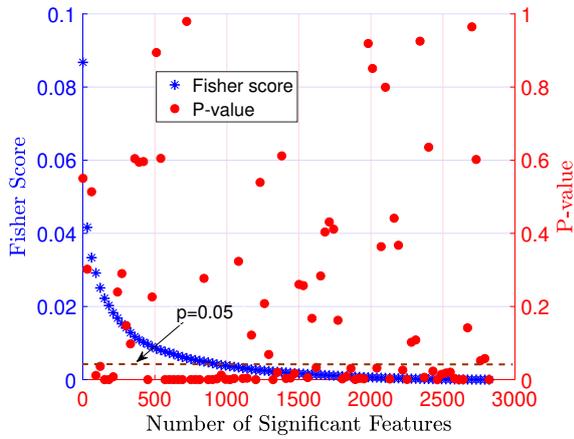


Fig. 3: Fisher scores and p -values of 2850 features, sorted by Fisher score. Features with p -value $p > 0.05$ will be removed.

classification performance. For implementing the reference feature selection methods, we use a feature selection toolbox, introduced in [39].

Figure 3 helps explain how the proposed method selects features. By choosing $\alpha = 0.05$ for hypothesis testing, more than 600 features with the highest Fisher scores and having their p -value lower than 0.05 are selected out of the original 2850 features. It should be noted that all the top 500 features ranked by Fisher score have their p -value very close to zero, meaning they are able to completely reject the null hypothesis H_0 , giving them strong discriminative power. Another interesting result is that the selected features for the epileptic class are significantly different from those of the non-epileptic class, as shown in Figure 4.

To compare the influence of feature selection methods on classification performance, we choose a linear-kernel support vector machine (SVM) as the classifier. Four performance metrics are evaluated for each method, including ACC, SEN, SPE, and AUC [48].

Figure 5 shows the performances of the system using SVM with different feature selection methods. Given a same number of

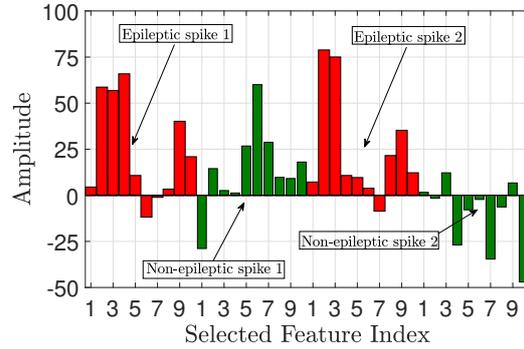


Fig. 4: Vectors of top 10 features selected for each of the two classes of epileptic spikes and non-epileptic spikes. While the feature vectors of two epileptic spikes are similar to each other, the non-epileptic feature vectors are not.

selected features, the system always performs better with the proposed method than with other methods, usually achieving an improvement of between 5% and 10% in terms of SEN, ACC, and AUC. AUC of the system with the proposed method is always higher than 0.9 when the number of selected features is higher than 50, that means excellent overall performance can be achieved with only about 50 features out of 2850. It is also shown that the performance reaches its best and remains stable when the number of features is greater than 70, with SEN, ACC, and AUC of around 80%, 92%, and 0.95, respectively. On the contrary, to achieve a similar performance, other methods need to select at least 250 features. The proposed method has outperformed the existing state-of-the-art methods in this analysis.

Tables 3 and 4 provide the system performance measures from our experiments using leave-one-out cross validation and 10-fold cross validation, respectively. In these experiments SVM is used with the first 100 features selected by the proposed method implemented in the feature selection stage of the system. It can be seen from Table 4 that the average performance of the proposed system is excellent, while in Table 3 the performance

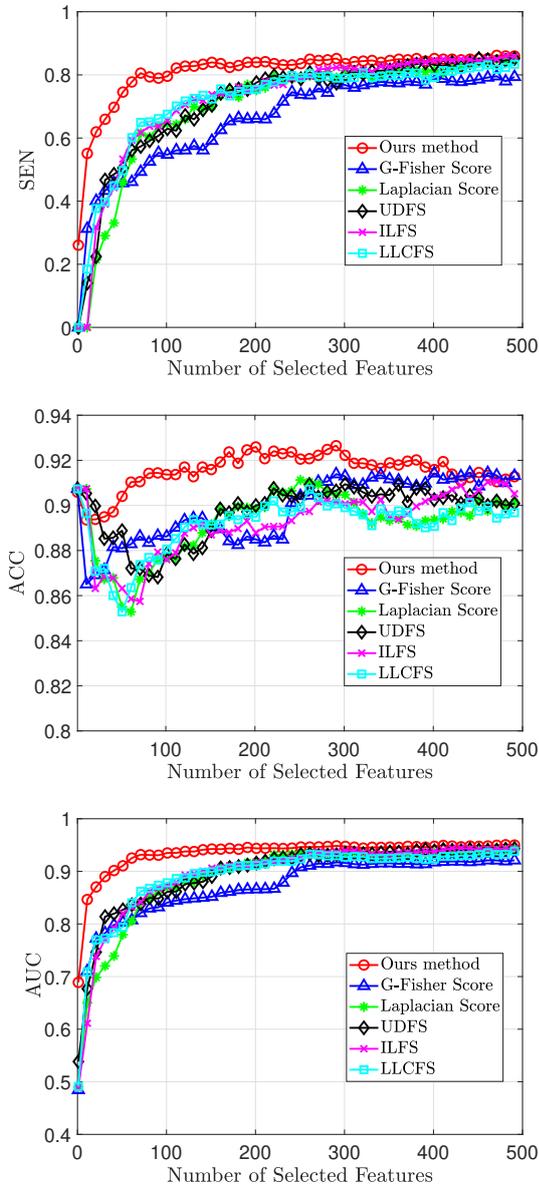


Fig. 5: Performances of the system using SVM with different feature selection methods.

may vary from patient to patient. The worst performances often happen only to patients whose EEG contains very few epileptic spikes. For example, the system fails to detect any epileptic spike of patient #9 (SEN is 0%), whose EEG has only one epileptic spike over 75

Table 3: Performance measures of the proposed SVM-employed system, using leave-one-out cross validation with the first 100 significant features.

Pat.	EPs/Non-EPs	SEN	SPE	ACC	AUC
1	8/393	75%	97.71%	97.26%	0.9066
2	635/193	78.90%	95.34%	82.73%	0.9511
3	6/188	100%	96.28%	96.39%	0.9885
4	16/453	100%	96.03%	96.16%	0.9970
5	351/816	85.75%	96.69%	93.40%	0.9655
6	22/602	77.27%	97.01%	96.31%	0.9723
7	2/50	100.0%	98.00%	98.08%	0.9900
8	11/589	81.82%	96.77%	96.50%	0.9750
9	1/75	0.00%	100%	98.68%	0.9920
10	8/274	75.00%	96.72%	96.10%	0.9658
11	2/117	50.00%	95.73%	94.96%	0.9573
12	3/582	33.33%	95.70%	95.38%	0.9364
13	5/514	80.00%	95.72%	95.57%	0.9712
14	8/76	87.50%	97.37%	96.43%	0.9655
15	324/202	80.25%	97.52%	86.88%	0.9655
16	38/372	84.21%	97.85%	96.59%	0.9417
17	12/618	100.0%	94.81%	94.83%	0.9919

Table 4: Performance measures of the proposed SVM-employed system, using 10-fold cross validation with the first 100 significant features.

Case	EPs/Non-EPs	SEN	SPE	ACC	AUC
1	144/611	81.25%	96.73%	93.77%	0.9579
2	144/611	81.94%	97.55%	94.57%	0.9664
3	144/611	88.89%	93.84%	92.98%	0.9594
4	144/611	80.56%	95.74%	92.85%	0.9583
5	144/611	77.08%	97.22%	93.38%	0.9588
6	144/611	81.25%	96.56%	93.64%	0.9671
7	144/611	81.25%	96.73%	93.77%	0.9657
8	144/611	83.33%	95.91%	93.51%	0.9673
9	144/611	86.11%	96.73%	94.70%	0.9707
10	146/616	86.30%	97.40%	95.27%	0.9720
Average:		82.80%	96.45%	93.84%	0.9643

non-epileptic spikes.

We also experiment with different classifiers on the proposed system, namely SVM, KNN (K-Nearest Neighbors), and NB (Naive Bayes). Performance of the system with different classifiers are presented in Table 5. In general, SVM performs slightly better than the other two classifiers.

Table 5: Performances of the system using SVM, KNN, and NB with first 100 significant features selected by the proposed methods

Metric	SVM	KNN	NB
SEN	82.80%	82.80%	82.80%
SPE	96.45%	97.96%	84.66%
ACC	93.84%	90.30%	84.01%
AUC	0.9643	0.8806	0.9024

5. Conclusions

In this paper, we introduced a new feature selection method which combined Fisher score and p -value methods in the stage of feature selection of the multi-channel EEG epileptic spike detection system recently proposed in [35], in order to improve its performance for classifying epileptic and non-epileptic spikes. Effectively, the proposed feature selection method reduced the dimension of the feature space and achieved good separability between epileptic spikes and non-epileptic spikes. The numerical experiments have indicated that the proposed method outperforms several state-of-the-art methods, including the generalized Fisher score, Laplacian score, UDFS, ILFS and LLCFS.

Acknowledgments

This work has been supported by VNU University of Engineering and Technology under project number CN18.15.

References

- [1] A. T. Berg, C. P. Panayiotopoulos, Atlas of Epilepsies, 1st Edition, Springer-Verlag London, 2010.
- [2] N. Senanayake, G. C. Román, Epidemiology of epilepsy in developing countries., Bulletin of the World Health Organization 71 (2) (1993) 247.
- [3] A. Carpio, W. A. Hauser, Epilepsy in the developing world, Current neurology and neuroscience reports 9 (4) (2009) 319–326.
- [4] N. A. Tuan, L. Q. Cuong, P. Allebeck, N. T. K. Chuc, H. E. Persson, T. Tomson, The prevalence of epilepsy in a rural district of vietnam: A population-based study from the epibavi project, Epilepsia 49 (9) (2008) 1634–1637.
- [5] J. Gotman, Automatic recognition of epileptic seizures in the EEG, Electroencephalography and Clinical Neurophysiology 54 (5) (1982) 530–540.
- [6] Ö. Özdamar, T. Kalayci, Detection of spikes with artificial neural networks using raw EEG, Computers and Biomedical Research 31 (2) (1998) 122–142.
- [7] C. C. Pang, A. R. Upton, G. Shine, M. V. Kamath, A comparison of algorithms for detection of spikes in the electroencephalogram, IEEE Transactions on Biomedical Engineering 50 (4) (2003) 521–526.
- [8] A. Ossadtchi, S. Baillet, J. Mosher, D. Thyerlei, W. Sutherling, R. Leahy, Automated interictal spike detection and source localization in magnetoencephalography using independent components analysis and spatio-temporal clustering, Clinical Neurophysiology 115 (3) (2004) 508–522.
- [9] H. S. Liu, T. Zhang, F. S. Yang, A multistage, multimethod approach for automatic detection and classification of epileptiform EEG, IEEE Transactions on Biomedical Engineering 49 (12) (2002) 1557–1566.
- [10] N. Acır, C. Güzeliş, Automatic spike detection in EEG by a two-stage procedure based on support vector machines, Computers in Biology and Medicine 34 (7) (2004) 561–575.
- [11] G. Xu, J. Wang, Q. Zhang, S. Zhang, J. Zhu, A spike detection method in EEG based on improved morphological filter, Computers in biology and medicine 37 (11) (2007) 1647–1652.
- [12] T.-W. Shen, X. Kuo, Y.-L. Hsin, Ant K-Means Clustering Method on Epileptic Spike Detection, in: Natural Computation, 2009. ICNC'09. Fifth International Conference on, Vol. 6, IEEE, 2009, pp. 334–338.
- [13] H. Hamid, B. Boashash, A time-frequency approach for EEG spike detection, Iranica Journal of Energy & Environment 2 (4) (2011) 390–395.
- [14] E. Acar, C. Aykut-Bingol, H. Bingol, R. Bro, B. Yener, Multiway analysis of epilepsy tensors, Bioinformatics 23 (13) (2007) i10–i18.
- [15] M. De Vos, A. Vergult, L. De Lathauwer, W. De Clercq, S. Van Huffel, P. Dupont, A. Palmi, W. Van Paesschen, Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone, NeuroImage 37 (3) (2007) 844–854.
- [16] M. Ontivero-Ortega, Y. Garcia-Puente, 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.
- [17] Y. R. Aldana, B. Hunyadi, E. J. M. Reyes, V. R. Rodriguez, S. V. Huffel, Nonconvulsive epileptic seizure detection in scalp EEG using multiway data analysis, *IEEE Journal of Biomedical and Health Informatics* (2018) 1–12.
- [18] E. Pippa, V. G. Kanas, E. I. Zacharaki, V. Tsirka, M. Koutroumanidis, V. Megalooikonomou, EEG-based classification of epileptic and non-epileptic events using multi-array decomposition, *International Journal of Monitoring and Surveillance Technologies Research* 4 (2) (2016) 1–15.
- [19] N. T. Anh-Dao, T. Le Thanh, N. Linh-Trung, H. V. Le, Nonnegative tensor decomposition for eeg epileptic spike detection, in: 2018 5th NAFOSTED Conference on Information and Computer Science (NICS), 2018, pp. 194–199.
- [20] J. Tang, S. Alelyani, H. Liu, Feature selection for classification: A review, *Data classification: algorithms and applications* (2014) 37.
- [21] G. Chandrashekar, F. Sahin, A survey on feature selection methods, *Computers & Electrical Engineering* 40 (1) (2014) 16–28.
- [22] J. R. Vergara, P. A. Estévez, A review of feature selection methods based on mutual information, *Neural computing and applications* 24 (1) (2014) 175–186.
- [23] S. Khalid, T. Khalil, S. Nasreen, A survey of feature selection and feature extraction techniques in machine learning, in: 2014 Science and Information Conference, IEEE, 2014, pp. 372–378.
- [24] B. Xue, M. Zhang, W. N. Browne, X. Yao, A survey on evolutionary computation approaches to feature selection, *IEEE Transactions on Evolutionary Computation* 20 (4) (2016) 606–626.
- [25] R. Sheikhpour, M. A. Sarram, S. Gharaghani, M. A. Z. Chahooki, A survey on semi-supervised feature selection methods, *Pattern Recognition* 64 (2017) 141–158.
- [26] Y. Kim, W. N. Street, F. Menczer, Feature selection in unsupervised learning via evolutionary search, in: Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '00, ACM, New York, NY, USA, 2000, pp. 365–369.
- [27] M. Qian, C. Zhai, Robust unsupervised feature selection, in: Twenty-Third International Joint Conference on Artificial Intelligence, California, USA, 2013, pp. 1621–1627.
- [28] M. Kalakech, P. Biela, L. Macaire, D. Hamad, Constraint scores for semi-supervised feature selection: A comparative study, *Pattern Recognition Letters* 32 (5) (2011) 656–665.
- [29] P. Pudil, J. Novovičová, J. Kittler, Floating search methods in feature selection, *Pattern recognition letters* 15 (11) (1994) 1119–1125.
- [30] A. Jain, D. Zongker, Feature selection: Evaluation, application, and small sample performance, *IEEE transactions on pattern analysis and machine intelligence* 19 (2) (1997) 153–158.
- [31] D. Garrett, D. A. Peterson, C. W. Anderson, M. H. Thaut, Comparison of linear, nonlinear, and feature selection methods for eeg signal classification, *IEEE Transactions on neural systems and rehabilitation engineering* 11 (2) (2003) 141–144.
- [32] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echaz, B. Litt, Epileptic seizure prediction using hybrid feature selection over multiple intracranial eeg electrode contacts: a report of four patients, *IEEE transactions on biomedical engineering* 50 (5) (2003) 603–615.
- [33] R. Jenke, A. Peer, M. Buss, Feature extraction and selection for emotion recognition from eeg, *IEEE Transactions on Affective Computing* 5 (3) (2014) 327–339.
- [34] J. Atkinson, D. Campos, Improving bci-based emotion recognition by combining eeg feature selection and kernel classifiers, *Expert Systems with Applications* 47 (2016) 35–41.
- [35] L. T. Thanh, N. T. Anh-Dao, V.-D. Nguyen, N. L. Trung, K. Abed-Meraim, Multi-channel eeg epileptic spike detection by a new method of tensor decomposition, *Journal of Neural Engineering* (2019 (provisionally accepted)).
- [36] Q. Gu, Z. Li, J. Han, Generalized Fisher score for feature selection, in: 27th Conference on Uncertainty in Artificial Intelligence, AUAI Press, 2011, pp. 266–273.
- [37] X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, in: Advances in Neural Information Processing Systems, 2006, pp. 507–514.
- [38] Y. Yang, H. T. Shen, Z. Ma, Z. Huang, X. Zhou, L_{2,1}-norm Regularized Discriminative Feature Selection for Unsupervised Learning, in: 22nd International Joint Conference on Artificial Intelligence, AAAI Press, 2011, pp. 1589–1594.
- [39] G. Roffo, S. Melzi, U. Castellani, A. Vinciarelli, Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach, in: 2017 IEEE International Conference on Computer Vision, 2017.
- [40] H. Zeng, Y.-M. Cheung, Feature selection and kernel learning for local learning-based clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33 (8) (2011) 1532–1547.
- [41] T. G. Kolda, B. W. Bader, Tensor decompositions and applications, *SIAM review* 51 (3) (2009) 455–500.

- [42] Y.-D. Kim, S. Choi, Nonnegative tucker decomposition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1–8.
- [43] L. De Lathauwer, B. De Moor, J. Vandewalle, On the best rank-1 and rank-(r_1, r_2, \dots, r_n) approximation of higher-order tensors, *SIAM journal on Matrix Analysis and Applications* 21 (4) (2000) 1324–1342.
- [44] G. Van Belle, L. D. Fisher, P. J. Heagerty, T. Lumley, *Biostatistics: A Methodology for the Health Sciences*, Vol. 519, John Wiley & Sons, 2004.
- [45] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classification*, John Wiley & Sons, 2012.
- [46] T. H. Cormen, C. E. Leiserson, R. L. Rivest, C. Stein, *Introduction to algorithms*, 2009.
- [47] R. L. Wasserstein, N. A. Lazar, et al., The asa’s statement on p values: context, process, and purpose, *The American Statistician* 70 (2) (2016) 129–133.
- [48] D. M. Powers, Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation, *International Journal of Machine Learning Technology* 2 (1) (2011) 37–63.
- [49] T. Fawcett, An introduction to roc analysis, *Pattern recognition letters* 27 (8) (2006) 861–874.