



**HAL**  
open science

## Dissimilarity-based time–frequency distributions as features for epileptic EEG signal classification

Youssef Ech-Choudany, Daniel Scida, Mustapha Assarar, J. Landré, Benaïssa Bellach, F. Morain-Nicolier

► **To cite this version:**

Youssef Ech-Choudany, Daniel Scida, Mustapha Assarar, J. Landré, Benaïssa Bellach, et al.. Dissimilarity-based time–frequency distributions as features for epileptic EEG signal classification. Biomedical Signal Processing and Control, 2021, 64, pp.102268. 10.1016/j.bspc.2020.102268 . hal-02973615

**HAL Id: hal-02973615**

**<https://hal.science/hal-02973615>**

Submitted on 24 Oct 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# Dissimilarity-based time-frequency distributions as features for epileptic EEG signal classification

Y. Ech-Choudany<sup>a,b,\*</sup>, D. Scida<sup>c</sup>, M. Assarar<sup>c</sup>, J. Landré<sup>b</sup>, B. Bellach<sup>a</sup>, F. Morain-Nicolier<sup>b</sup>

<sup>a</sup>*LSE2I Laboratory, National School of Applied Sciences, First Mohammed University, B.P. 669, Complexe Universitaire Hay elqods, 60000 Oujda, Morocco*

<sup>b</sup>*CReSTIC Laboratory, Université de Reims Champagne-Ardenne, IUT de Troyes, 9 rue de Québec, Troyes Cedex, France*

<sup>c</sup>*LISM Laboratory, Université de Reims Champagne-Ardenne, IUT de Troyes, 9 rue de Québec, Troyes Cedex, France*

---

## Abstract

This work aims at exploring a general framework embedding techniques from classifiers, Time Frequency Distributions (TFD) and dissimilarity measures for epileptic seizures detection. The proposed approach consists firstly in computing dissimilarities between TFD of electroencephalogram (EEG) signals and secondly in using them to define a decision rule. Compared to the existing approaches, the proposed one uses entire TFD of EEG signals and does not require arbitrary feature extraction. Several dissimilarity measures and TFDs have been compared to select the most appropriate for EEG signals. Classifiers, such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Linear Discriminate Analysis (LDA) and k-Nearest Neighbours (k-NN), have been combined with the proposed approach. In order to evaluate the proposed approach, 13 different classification problems (including 2, 3 and 5-class) pertaining to five types of EEG signals have been used. The comparison between results obtained with the proposed approach and results reported in the literature with the same database of epileptic EEG signals demonstrates the effectiveness of this approach for seizure detection. Experimental results show that this approach has achieved highest accuracy in the most studied classification problems. A high value of 98% is achieved for the 5-class problem. Further, in most classification problems with 2 and 3-class, it also yields a satisfactory accuracy of approximately 100%. The robustness of the proposed approach is evaluated with the addition of noise to the EEG signals at various signal-to-noise ratios (SNRs). The experimental results show that this approach has a good classification accuracy at low SNRs.

**Keywords:** Epileptic seizure, EEG signals, Dissimilarity, Classification, ANN, SVM, LDA, k-NN, TFA, TFD, SPWV, SNR

---

## 1. Introduction

Electroencephalography (EEG) is an important clinical tool that is used for studying the human brain functions and neurological disorders [1]. Epilepsy is one of the most common disorders of the nervous system and affects approximately 1% of the human population [2]. The detection of epileptic seizures within EEG signals is an effective way of diagnosing epilepsy [3, 4].

Numerous signal analysis and statistical methods of EEG signals are used to identify the epileptic seizures. Generally, the process of these methods is divided into two steps: feature extraction and classification. Feature extraction aims at extracting relevant information hidden in the EEG signals, which directly dominates the accuracy of the final classification. However, feasible features might be difficult to

---

\*Corresponding author

Email address: [y.echchoudany@gmail.com](mailto:y.echchoudany@gmail.com) (Y. Ech-Choudany)

obtain or inefficient for learning purposes, e.g., when experts cannot define features in a straightforward way or when data is high dimensional and consist of a large set of attributes. Another drawback of the use of features is that different objects may have the same representation as they differ from each other by properties that were not expressed in the chosen feature set. This results in class overlap: in some areas in the feature space, objects of different classes are represented by the same feature vectors. Consequently, they cannot be distinguished, which leads to misclassification errors.

An alternative to the use of features is the dissimilarity representation approach based on direct pairwise object comparisons. This approach is often called dissimilarity-based classification and does not require direct access to the features of the objects. If all the objects are considered in the comparison, then only identical objects will have a dissimilarity of zero. For such a representation class overlapping does not exist if the objects can be unambiguously labelled: there are no real-world objects in the application that belong to more than one class. Only identical objects have a zero-distance and they should have the same label as they are identical. Another advantage of the dissimilarity-based classification is that it uses the expert knowledge in a different way. Instead of features, a dissimilarity measure has to be supplied. The classification of objects using the dissimilarity-based approach has been an active research topic [5–9], but no attention has been paid to use this approach for epileptic seizure detection.

Several methods have been applied to epileptic EEG signals for feature extraction [3, 4, 10–28], such as representations based on the fast Fourier transform, discrete wavelet transform and time-frequency representations (TFR). Fast Fourier transform methods [12] are based on previous observations that epileptic seizures give rise to changes in certain frequency bands such as  $\delta(0.4–4Hz)$ ,  $\theta(4–8Hz)$ ,  $\alpha(8–12Hz)$  and  $\beta(12–30Hz)$ . Since the nature of epileptic EEG signals is non-stationary, these methods are not suitable for the frequency decomposition of these signals. Methods based on linear and quadratic TFRs have shown higher performance than conventional frequency analysis methods. From the variety of available approaches, the wavelet transform (WT) is the most widely used on feature extraction of epileptic EEG signals [3, 14, 19]. This can be explained by its excellent capabilities in multi-resolution representation. Despite the various resolutions it allows, this tool does not make it possible to adapt to all signals, especially non-stationary and non-linear signals such as EEG signals. Quadratic TFDs have been previously exploited in some works [11, 20, 22] to analyse epileptic seizures. In fact, these distributions have the advantage of characterizing the signals thanks to their energy distribution in the time-frequency plane. As shown in [29], quadratic TFDs provide a powerful framework for feature extraction and classification of non-stationary phenomena in signals.

Classifiers such as Support Vector Machines (SVM) [13, 14, 17, 18, 21, 28, 30],  $k$ -Nearest Neighbors ( $k$ -NN) [4], Quadratic Discriminant Analysis (QDA) [16], Artificial Neural Network (ANN) [10, 22–27] and Extreme Learning Machine (ELM) [3, 20] have been widely used to classify epileptic EEG signals. However, these methods are performed in the space generated by features.

The dissimilarity-based approach differs from typical pattern recognition approaches where objects to be classified are represented by feature vectors. In this approach, objects are described using pairwise dissimilarities [5, 6]. The motivation of this proposal is that the proximity information is more important for a class membership than features [7]. As a next step, the dissimilarity representation is transformed into a vector space in which traditional statistical classifiers can be used. The dissimilarity-based approach offers thereby an alternative to kernel methods based on similarities. The dissimilarity measures are more general than kernels. The later have to obey the Mercer condition so that the implicit construction of classifiers, such as SVM, can be possible in the related kernel spaces [9]. The dissimilarity approach has the advantage of using any measure as well as any classifier that works in vector spaces. The goal of this work is to investigate this advantageous approach to classify EEG signals and contribute to a better epileptic seizure detection. We propose an approach based on TFD and dissimilarity-based approach, called dissimilarity-based TFD approach.

Dissimilarity is a relative measure that quantifies the independency between two objects that have the same nature or characteristics. The more the two objects resemble one another, the lower the dissimilarity measure. In the case of classification problems, a dissimilarity measure is usually performed from a distance function. This distance can be defined as metric or semi-metric. The distance is called metric when conditions of reflectivity, positivity, symmetry and triangle inequality are verified. Moreover, the distance is called semi-metric when the triangle inequality is not fulfilled [31, 32].

The main contributions of this work are: (1) address the use of quadratic TFD of EEG signals in order to obtain the energy distribution in the time-frequency domain; (2) introduce the principle of the dissimilarity-based TFD approach; (3) use the proposed approach to classify epileptic EEG signals. Firstly, the proposed methods are presented in detail in section 2. Then, a description of the publicly available database of epileptic EEG signals is carried out. This database consists in five different sets of EEG signals. From these five sets, 13 different classification problems, including 2, 3 and 5-class, are defined. In order to evaluate the performance of the dissimilarity-based TFD approach, different statistical measures have been used. These measures include the accuracy, the sensitivity, the specificity, and the area under the Receiver Operating Characteristics (ROC) curve. The technique of 10-fold cross validation is also used. Different dissimilarity measures and TFDs have been compared to select the most relevant for epileptic EEG signals. Four classifiers, such as ANN, SVM, LDA and  $k$ -NN, are combined with the proposed approach. The computational complexity and the robustness of the proposed approach have been explored. The results obtained with the proposed approach under various classification problems are compared to results reported in literature with the same database of epileptic EEG signals. The comparison demonstrates the effectiveness of the proposed approach for seizure detection.

## 2. Dissimilarity-based TFD approach

The aim of this section is to present the different methods that we combine to analyse and classify the EEG signals. The main steps of the dissimilarity-based TFD approach are shown on Figure 1. Given three sets of signals: the representation set  $R$  with  $k$  prototype signals, the training set  $T$  with  $n$  training signals and the testing set  $S$  with  $m$  testing signals, the algorithm starts with the learning steps (first row of the figure) and the computation of the TFDs of the training and prototype signals. From  $n$  and  $k$  signals, a  $n \times k$  training dissimilarity matrix  $D(T, R)$  is constructed. A given training signal is thus described by  $k$  dissimilarities, coding its proximity to the prototype signals. Each of the  $k$  column of the training matrix is then injected in the learning step of the classifier. After this step, the classifier can work to test an unknown signal (second row of the figure). TFDs of the testing signals are computed. These TFDs are compared to the TFDs of the prototype signals by computing an  $m \times k$  testing dissimilarity matrix  $D(S, R)$ . This matrix is injected into the classifier to make decisions, producing labels for the  $m$  testing signals.

The success of this approach depends on three fundamental issues, namely the choice of the dissimilarity measure, the choice of the TFD and the type of classifier. More details about these issues are given in the next sections.

### 2.1. Dissimilarity-based classification

The dissimilarity representation has been discussed extensively [5–9], so here we will only focus on some aspects that are essential for this paper. These aspects are obtained from the paper of Duin et al. [6].

Unlike a feature-based approach, the idea of a dissimilarity-based approach is to represent an object via a vector of its dissimilarities with respect to a collection of objects. Figure 2 represents the difference

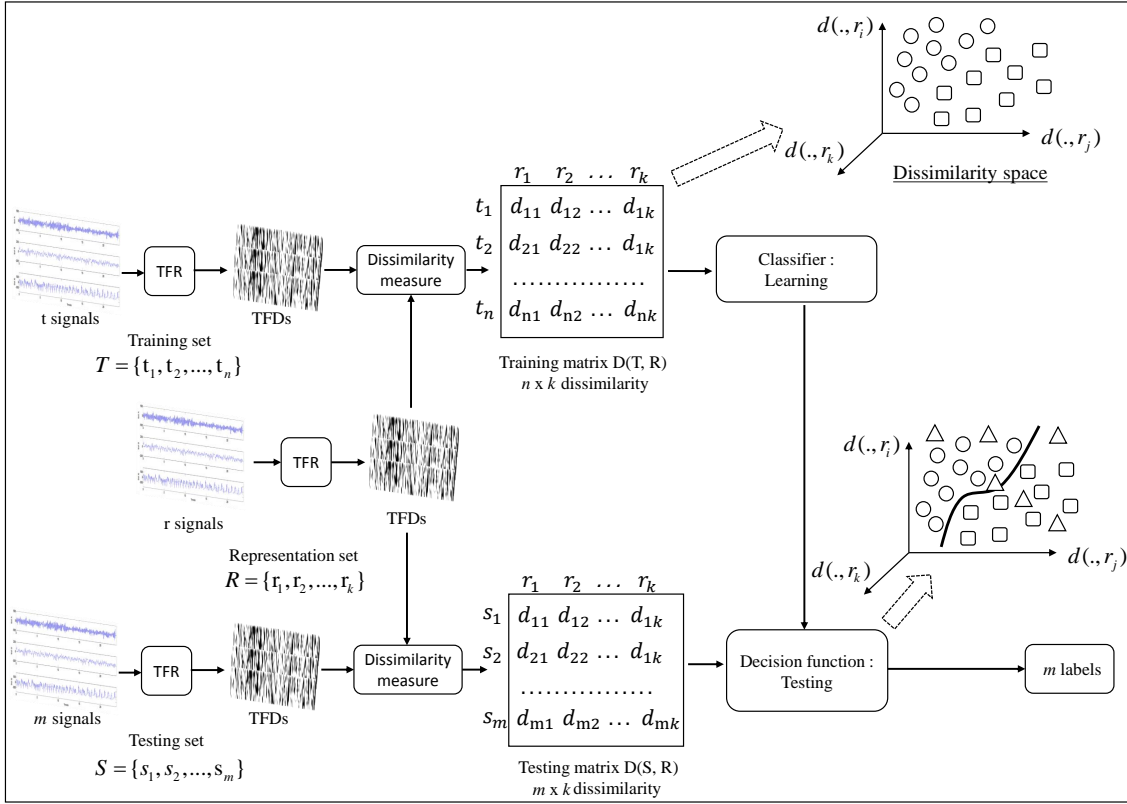


Figure 1: Flow diagram of dissimilarity-based TFD classification.

with respect to the representation between the traditional feature-based approach and the dissimilarity-based approach [7]. Three sets of objects may be distinguished for constructing a dissimilarity representation [6]. This first set is a representation set  $R = \{r_1, \dots, r_k\}$ . These are the objects we refer to. The dissimilarities to the representation set have to be computed for training objects as well as for test objects used for evaluation, or any objects to be classified later. The objects in the set  $R$  are called prototypes. The second one is a training set  $T$ . These are the objects that are used to train classifiers. In many applications we use  $T := R$ , but  $R$  may also be just a (small) subset of  $T$ , or be entirely different from  $T$ . The last set is a test set  $S$ . These are the objects that are used to evaluate classification procedures. After determining these three sets of objects the dissimilarity matrices  $D(T, R)$ ,  $D(T, T)$  and  $D(S, R)$  have to be computed. The next problem is how to use these three matrices for training and testing. In this work, two procedures have been considered :

- A direct approach to dissimilarities leads to the  $k$ -NN method. This rule is applied here to  $D(S, T)$ , such as the test objects in the set  $S$  are classified to the class on which most frequently occurs among the  $k$  nearest neighbours in  $T$ . The dissimilarity of two objects is determined using a distance function.

The  $k$ -NN classifier is expected to perform well if a good distance function is found and a training set  $T$  is representative. However, when a small representation set  $R$  is selected, the performance of the  $k$ -NN rule can significantly deteriorate. In addition, the decision rule of the  $k$ -NN classifier is based on local neighbourhoods only. It means that the  $k$  nearest neighbours found might not include the best representatives of a class to which an object should be assigned. Moreover, the  $k$ -NN does not work for an asymmetric distance measure. In such cases, a better generalization can be achieved by a classifier built in a dissimilarity space.

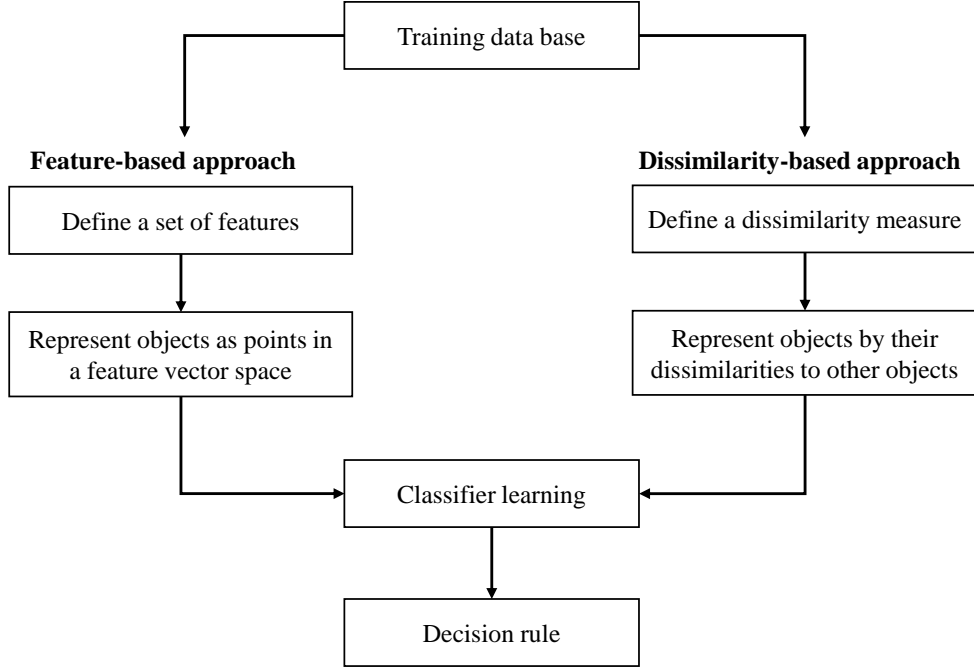


Figure 2: Difference between feature-based approach and dissimilarity-based approach.

- The dissimilarity space. This space is postulated as a Euclidean vector space defined by the dissimilarity vectors  $d(\cdot, R) = [d(\cdot, r_1), \dots, d(\cdot, r_k)]$  computed to the representation set  $R$  as dimensions. The dissimilarity vectors serve as features for the objects in the training set, which can be used as an input to any classifier [34].

## 2.2. Time-frequency representations

TFR have been developed to characterize the evolution of the frequency content of a non-stationary signal as a function of time. The TFR map a one-dimensional signal of time  $s(t)$ , into a two-dimensional function of time and frequency  $\text{TFR}_{s(t)}(t, f)$ . From the variety of available approaches, the linear TFRs, such as WT, are the most widely used on feature extraction of epileptic EEG signals. However, the time resolution and frequency resolutions of the linear TFRs cannot be made arbitrarily well simultaneously. The time and frequency resolution of the linear TFRs are determined by the time-frequency localization of the used atoms. For example, WT analyses higher frequencies with good time resolution but poor frequency resolution and vice versa. In this section, we present several TFRs. Their performances in the problem covered in this paper are estimated and discussed in section 4.2.

Although the linearity of a TFR is a desirable property, the quadratic structure of a TFR is an intuitively reasonable assumption when we want to interpret a TFR as a time-frequency energy distribution. Ideally, it is desirable to achieve energy distribution in the time-frequency plane without loss of resolution. Among the TFDs, the most famous one is that of Gabor known as the Short-Time Fourier Transform (STFT). Given a signal  $s(t)$  and a short time window  $h(t)$ , the STFT of  $s(t)$  is defined by:

$$\text{STFT}_s(t, f) = \int_{-\infty}^{+\infty} s(u)h(u - t)e^{-2j\pi fu} du \quad (1)$$

Where,  $t$  is the time,  $f$  is the frequency.

STFT compromises window length and frequency resolution (Heisenberg's uncertainty principle). A shorter window makes it possible to obtain a very good time resolution but a poor frequency resolution and vice versa. To overcome this limitation, many TFDs have been proposed [35, 36]. The most of these distributions are known as the Cohen's class [37]. In this work, only a non-exhaustive list of TFDs belonging to the Cohen's class are presented and used. Thus, the general form of the Cohen's class of distributions is defined by:

$$C_s(t, f) = \int \int \int_{\mathbb{R}^3} s(u + \frac{\tau}{2}) s^*(u - \frac{\tau}{2}) \phi(v, \tau) e^{-2j\pi f\tau - j2\pi tv + j2\pi uv} dud\tau dv \quad (2)$$

Where  $s(t)$  is the signal,  $s^*(t)$  is its complex conjugate, and  $\phi(v, \tau)$  is a two-dimensional kernel function. Each TFD of the Cohen's class has its own kernel.

The most studied element of Cohen's class is certainly the Wigner-Ville Distribution (WVD). However, The main problem in applying the WVD is the cross-terms (interference), which are the result of the quadratic nature of this transformation. Cross-terms may mask low-energy components, especially in the case of large dispersion of energy levels of individual components, and make the interpretation of the WVD difficult. Time and frequency smoothing operations of the WVD reduce the influence of the unwanted cross-terms. The obtained distribution from these operations is named the Smoothed Pseudo Wigner-Ville Distribution (SPWVD) [38–40].

The WVD and the SPWVD are defined as :

$$\text{WVD}_s(t, f) = \int s(t + \frac{\tau}{2}) s^*(t - \frac{\tau}{2}) e^{-j2\pi f\tau} d\tau. \quad (3)$$

$$\text{SPWVD}_s^{h,g}(t, f) = \int \int h(t - u) \widehat{g}(f - v) \text{WVD}_s(t, f) dudv. \quad (4)$$

Thanks to the smoothing operations, the SPWVD is certainly one of the elements of Cohen's class which provides a precise representation of the energy of the signal over the time-frequency plane. To this end, the SPWVD has been compared to many TFDs of the Cohen's class and then has been used throughout this manuscript. Table 1 presents some TFDs of the Cohen's class, which are used in our study along with the corresponding kernels.

### 2.3. Dissimilarity measure

The dissimilarity is a useful tool to measure the internal relationship of any two objects [31, 32]. They are often defined as a distance function  $d$ . The lower the distance value, the larger the dissimilarity, and vice versa. As mentioned in section 2.2, the performance of dissimilarity-based classifiers is directly linked to a good choice of a distance function. Numerous distance functions are available, including the  $p$ -norm distance, the correlation, the f-divergence and the spectral Log-deviation. In this section, a non-exhaustive list of distances is presented. The performances of these dissimilarities, in the addressed problem, are estimated and discussed in section 4.1.

To introduce the definition of each distance, let's take two signals ( $S_1, S_2$ ) and their TFDs ( $\text{TFD}_1, \text{TFD}_2$ ).

The  $p$ -norm distance (noted by  $d_p$ ) is defined by:

$$d_p(\text{TFD}_{s_1}, \text{TFD}_{s_2}) = \left[ \int \int |\text{TFD}_{s_1}(t, f) - \text{TFD}_{s_2}(t, f)|^p dt df \right]^{\frac{1}{p}}. \quad (5)$$

We note that, for given values of  $p$ , there are particular cases of distances:

Table 1: TFDs of Cohen's class and their kernel functions.

Distribution	Kernel ( $\phi(v, \tau)$ )
WV	1
PWV	$h(\tau)$ $h(\tau)$ : window function
SPWV	$\hat{g}(v)h(\tau)$ $g(v), h(\tau)$ : window functions
Rihaczek (RIH)	$e^{j\pi v\tau}$
Margenau-Hill (MH)	$\cos(\pi v\tau)$
Pseudo Margenau-Hill (PMH)	$h(\tau)e^{j\pi v\tau}$ $h(\tau)$ : window function
Choi-Williams (CW)	$e^{-\frac{1}{2}(\frac{\pi v\tau}{\sigma})^2}$ $\sigma$ : scaling factor
Born-Jordan (BJ)	$\frac{\sin(\pi v\tau)}{\pi v\tau}$
Zhao-Atlas-Marks (ZAM)	$h(\tau)\frac{\sin(\pi v\tau)}{\pi v\tau}$ $h(\tau)$ : window function
Generalized Rectangular (GRECT)	$\frac{\sin(\frac{2\pi v\sigma}{ \tau ^\alpha})}{\pi v}$ $\sigma$ : scaling factor, $\alpha$ : dissymmetry ratio
Reduced Interference (RI)	$\int_{-\infty}^{+\infty} h(t)e^{-2j\pi v\tau t} dt$ $h(t)$ : hamming window

- $p=1$ : Manhattan distance.
- $p=2$ : Euclidean distance.
- $p=\infty$  : infinity norm distance also called Chebyshev distance.

The correlation distance (noted  $d_{corr}$ ) is defined by:

$$d_{corr}(\text{TFD}_{s_1}, \text{TFD}_{s_2}) = \frac{[d_2(\text{TFD}_{s_1}, \text{TFD}_{s_2})]^2}{\int \int |\text{TFD}_{s_1}(t, f)|^2 dt df + \int \int |\text{TFD}_{s_2}(t, f)|^2 dt df}. \quad (6)$$

One weakness of the basic  $p$ -norm distance functions is that if one of the input attributes of an object has a relatively large range, then it can overpower the other attributes. Therefore, the TFDs are often normalized by:

$$\text{NTFD}(t, f) = \frac{|\text{TFD}(t, f)|}{\int \int |\text{TFD}(t, f)| dt df}. \quad (7)$$

The  $p$ -norm distance between two normalized TFDs gives a new family of distance (noted  $d_{N_p}$ ):

$$d_{N_p}(\text{TFD}_{s_1}, \text{TFD}_{s_2}) = d_p(\text{NTFD}_{S_1}, \text{NTFD}_{S_2}). \quad (8)$$

We also note that, for a particular case of  $p=1$  in the  $d_{N_p}$  distance, the obtained distance is the Kolmogorov distance.

The  $f$ -divergence [41] makes it possible to have a family of distances parametrized by two functions.  $f$  is a continuous convex function on  $\mathbb{R}^+$  and  $g$  an increasing function on  $\mathbb{R}$ . The  $f$ -divergence between two probability density  $p_1(x)$  and  $p_2(x)$  is defined by:



$$d_{f-div}(p_1, p_2) = g \left\{ E_1 \left[ f \left( \frac{p_2(x)}{p_1(x)} \right) \right] \right\} \quad (9)$$

Where  $E_1$  is the expectation operator with respect to  $p_1$ .

The quadratic time-frequency analysis presents the energy distribution of a signal in the time-frequency domain, therefore it is considered like a density of probability. The normalized TFD is positive, therefore the f-divergence can be applied as:

$$d_{f-div}(\text{TFD}_{S_1}, \text{TFD}_{S_2}) = g \left[ \int \int f \left( \frac{\text{TFD}_{S_2}(x)}{\text{TFD}_{S_1}(x)} \right) \text{TFD}_{S_1} dt df \right]. \quad (10)$$

The table 2 presents a list of some f-divergence commonly used in signal processing for TFD classification [42].

The Kullback-Leibler divergence measure is not symmetric. This measure can be adapted to the TFDs like a distance as follows:

$$d(\text{TFD}_{S_1}, \text{TFD}_{S_2}) = d_{f-div}(\text{TFD}_{S_1}, \text{TFD}_{S_2}) + d_{f-div}(\text{TFD}_{S_2}, \text{TFD}_{S_1}). \quad (11)$$

The spectral Log-deviation [43] (noted  $d_{\text{SLD}_p}$ ) is defined by the  $p$ -norm distance between the logarithm of two normalized TFDs. The logarithm is used here to lower the influence of the high peak values of the TFD.

$$d_{\text{SLD}_p}(\text{TFD}_{S_1}, \text{TFD}_{S_2}) = d_p(\log(|\text{NTFD}_{S_1}|), \log(|\text{NTFD}_{S_2}|)). \quad (12)$$

### 3. Database of epileptic EEG signals

The EEG signals used in this study are from the publicly available database provided by Andrzejak et al. [47]. This database consists in five different sets of EEG signals. The EEG signals of the first set (denoted A) have been recorded from surface EEG recordings carried out on five healthy volunteers. The EEG signals of the second set (denoted B) have been collected from surface EEG recordings of the same five healthy volunteers with eyes closed. The EEG signals of the third set (denoted C) and the fourth set (denoted D) were obtained during seizure-free intervals (interictal periods) of five epileptic patients. EEG signals of the set C were recorded from hippocampal formation of opposite hemispheres of the brain. EEG signals in the set D were recorded from the epileptogenic zone. The EEG signals of the last set (denoted E) were recorded during seizure activity (ictal periods). Each set consists in 100 single channel EEG segments of 23.6 s duration, which were digitized at a sampling rate of 173.61 Hz and 12 bits A/D resolution. Figure 3 shows an example of an EEG signal from each set.

The pre-processing step (noise reduction, signal alignment, etc.) of the EEG signals of this database is not studied in this paper. In this paper, we follow the same validation procedure in [4-25], where only raw signals are processed, without any pre-processing.

Table 2: f-divergence commonly used in signal processing for TFD classification.

Divergence	$f(x)$	$g(x)$	$d_{f-div}(\text{NTFD}_{S_1}, \text{NTFD}_{S_2})$
Kolmogorov	$x$	$ 1 - x $	$\int \int  \text{NTFD}_{S_1}(t, f) - \text{NTFD}_{S_2}(t, f)  dt df$
Matusita	$x^{1/2}$	$ 1 - x^{1/2} ^2$	$[\int \int  (\text{NTFD}_{S_1}(t, f))^{1/2} - (\text{NTFD}_{S_2}(t, f))^{1/2} ^2 dt df]^{1/2}$
Kullback	$x$	$(x - 1)\log(x)$	$\int \int  \text{NTFD}_{S_1}(t, f) - \text{NTFD}_{S_2}(t, f)  \log \frac{\text{NTFD}_{S_1}(t, f)}{\text{NTFD}_{S_2}(t, f)} dt df$
Kullback-Leibler	$x$	$-\log(x)$	$\int \int \text{NTFD}_{S_1}(t, f) \log \frac{\text{NTFD}_{S_1}(t, f)}{\text{NTFD}_{S_2}(t, f)} dt df$

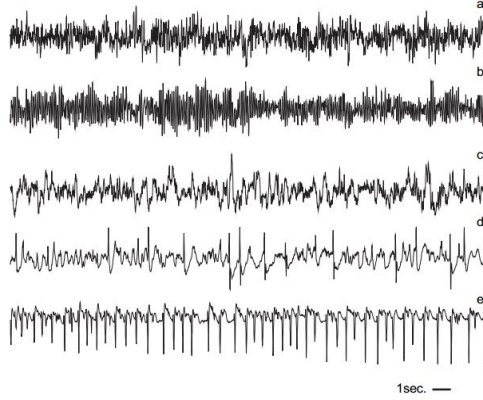


Figure 3: Examples of EEG signals from the used epileptic database [47].

#### 4. Results and discussion

The dataset presented in section 3 is used to evaluate the performance of our approach. 13 classification problems, including 2, 3, and 5-class, are considered to test the used approach. The classification instances are shown in Table 3. Because of the small size of this database, the training set  $T$  and the representation set  $R$  are considered as the same ( $T := R$ ) throughout this paper.

In order to evaluate the classification performance of the dissimilarity-based TFD approach, different statistical measures have been used in this work. These measures include the accuracy, the sensitivity, the specificity, and the Area Under the ROC Curve (AUC) [44]. The technique of 10-fold cross validation [45] is also used.

The total classification accuracy is defined as:  $\frac{TP+TN}{TP+TN+FP+FN}$ .

The specificity is defined as:  $\frac{TN}{FP+TN}$ .

The sensitivity is defined as:  $\frac{TP}{FN+TP}$ .

where TP and TN represent the total number of correctly detected true positive signals and true negative signals. The FP and FN represent the total number of false positive signals and false negative signals.

The ROC graph plots sensitivity and specificity values. The AUC is a scalar value, pertaining to the interval  $[0, 1]$ , obtained by measuring the area under the ROC curve. The AUC of a classifier is equivalent to the probability that this classifier gives a better rank to a positive element compared to a

Table 3: Classification instances of the three classification problems of EEG signals.

Number of class	Classification instance	class 1	class 2	class 3	class 4	class 5
2	A, E			-	-	-
	B, E	Normal	ictal	-	-	-
	AB, E			-	-	-
2	C, E			-	-	-
	D, E	interictal	ictal	-	-	-
	CD, E			-	-	-
2	ACD, E			-	-	-
	BCD, E	Normal + interictal	ictal	-	-	-
	ABCD, E			-	-	-
3	A, C, E				-	-
	A, D, E	Normal	interictal	ictal	-	-
	AB, CD, E				-	-
5	A, B, C, D, E	Normal (eyes open)	Normal (eyes closed)	interictal (epileptogenic)	interictal (hippocampal)	ictal

negative element, both randomly selected from the learning database. A perfect classifier possesses an AUC equal to 1. More details about the ROC graph and the AUC can be found in [46].

This section includes seven subsections. The goal of the first and the second one is to compare different dissimilarity measures and TFDs in order to choose the most appropriate one. The aim of the third is to evaluate the proposed approach using various classifiers. The primary focus of the fourth and the fifth ones is to provide an idea about the computational complexity and the robustness of the proposed approaches, respectively. The main goal of the sixth subsection is to compare the results obtained by the proposed approach with the literature results. In the last one, we present and discuss the obtained results.

#### 4.1. Dissimilarity measures evaluation

In section 2.2, we have seen that the dissimilarity-based approach requires a dissimilarity measure (distance). The performance of the classifiers based on this approach is directly linked to the best choice of this distance. The distance should be defined for any pair of samples and should be a metric or a semi-metric one. Different types of distances, including the p-norm distance, the correlation, the f-divergence, and the spectral Log-deviation, can be used to compare TFDs.

To choose the appropriate distance for the proposed approach, the SPWV distribution with a Gaussian-smoothing window, the  $k$ -NN classifier-based dissimilarity as distance, ANN, SVM and LDA classifier based dissimilarity as features, the 5-class classification problem (instances: A, B, C, D, E), the accuracy and the 10-fold validation were used. Eight distances have been tested, such as Euclidean, correlation, Kolmogorov, Kullback, Matusita, Kullback-Leiber,  $d_{SLD_1}$  and  $d_{SLD_2}$  distance. The obtained results are reported in Table 4.

The higher classification accuracies of the all classifiers are obtained by the distances based on the normalized TFDs (Table 4). However, the best classification accuracy of each classifiers was obtained with the same dissimilarity measure (spectral Log-deviation  $d_{LSD_1}$ ). For that, in the following section only the  $k$ -NN classifier based  $d_{LSD_1}$  as distance is used to determine an appropriate TFD for the proposed approach.

#### 4.2. TFDs evaluation

To choose an appropriate TFD for the proposed approach, the 5-class classification problem (instances: A, B, C, D, E), the  $k$ -NN classifier based dissimilarity as distances, the  $d_{LSD_1}$  distance, the accuracy and 10-fold cross-validation were used.

11 TFDs were tested and divided into three categories, the first includes distributions that do not use smoothing windows (MH, RIH, WV), the second includes distributions that use frequency-smoothing windows (PMH, PWV) and the third category presents distributions that use both frequency and time smoothing windows (SPWV, CW, GRECT, RI, ZAM, BJ). For categories two and three, a Gaussian-smoothing window is used. The  $k$ -NN classifier was ran for each TFD and the obtained accuracies were noted. The results obtained are shown in Table 5.

For the first category of distributions, the higher accuracy (86.60%) is obtained by the WV distribution. For the second one, the higher accuracy (89.60%) is obtained by the PWV distribution. Finally, for

Table 4: Classification accuracies of the  $k$ -NN, SVM, ANN, and LDA classifier using different distances. Performance evaluation using the 5-class classification problem (instances: A, B, C, D, E) with 10-fold cross validation.

Dissimilarity measures		Euclidean	Correlation	Kolmogorov	Kullback	Matusita	Kullback-Leiber	$d_{SLD_1}$	$d_{SLD_2}$
Approach	Classifier								
dissimilarity as feature	ANN	80.20	94.60	94.80	92.20	91.20	94.40	96.60	95.60
	SVM	34.40	90.00	93.40	90.40	78.80	90.00	94.00	94.00
	LDA	82.60	94.80	96.20	94.80	92.00	72.80	97.60	97.00
dissimilarity as distance	$k$ -NN	22.40	90.20	88.00	90.20	89.60	90.20	94.40	93.60

the third category of distributions, the higher accuracy (94.40%) is obtained by the SPWV distribution. The TFDs which use both frequency and time windows present excellent classification accuracy results. It means that the elimination of cross-terms significantly increases the quality of the classification. In all categories, the higher accuracy values were obtained by the WV distribution and its extensions (PWV, SPWV). However, the best accuracy of all categories of TFDs was obtained by the SPWV distribution (94.40%).

To eliminate cross-terms, many time and frequency smoothing windows can be used. To choose the best one, six of them (Gaussian, Hamming, Hanning, Rectangular, Parzen and Kaiser) with a window length of 128-points were tested. The SPWV distribution,  $k$ -NN classifier, accuracy, and the 5-class classification (instances: A, B, C, D, E) with 10-fold cross-validation were used. The obtained results are presented in Table 6. According to the results, the best accuracy classifications (95.40%) were obtained by Hanning and Parzen smoothing windows. For this reason, the SPWV distribution with a Hanning smoothing window was selected in the following to evaluate the proposed approach.

### 4.3. Evaluation of classifiers

The goal of this section is to evaluate the proposed approach using different classifiers. Firstly, the performances of these methods are detailed and compared using the 5-class classification problem. Secondly, the performance of the same methods is checked under the 2 and 3-class classification problems. To evaluate the performance of these methods, the SPWV with Hanning smoothing window is used with the  $d_{LSD_1}$  distance. The classification performances were checked using the 10-fold cross-validation, AUC, specificity, sensitivity and accuracy.

Four classifiers have been compared,  $k$ -NN based dissimilarity as (features, distances), ANN, SVM and LDA based dissimilarity as features.  $k$ -NN number of neighbours, ANN architecture, SVM and LDA parameters are derived by trial and error. For the ANN classifier, a three-layer feedforward neural network architecture is used. The complexity of this architecture is characterized by the number of neurons in the hidden layer. A linear kernel is used to perform the SVM classifier. All these classifiers were implemented by using the MATLAB software with neural-networks toolbox, statistics and machine learning toolbox. The default parameters of the pre-implemented classifiers have been used. For each classifier, the new parameters that give the highest accuracies are presented in Table 7. And the obtained results are summarized in Table 8.

From table 8, it can be observed that the highest performances of all classifiers are obtained for classes A, B and E. The specificity and the sensitivity obtained in these classes are about 100% and the AUC is approximately 1. However, low performances of these classifiers are obtained in class C and

Table 5: Classification accuracies of the  $k$ -NN based dissimilarity as distance classifier using the first, second and third category of TFDs. The  $d_{LSD_1}$  distance is used. Performance evaluation using the 5-class classification problem (instances: A, B, C, D, E) with 10-fold cross validation.

Category	1			2		3					
Distribution	WV	RIH	MH	PWV	PMH	SPWV	CW	GRECT	RI	ZAM	BJ
Accuracy (%)	86.60	66.80	65.80	89.60	74.20	94.40	86.60	86.40	85.00	83.80	88.00

Table 6: Classification accuracies of the  $k$ -NN based dissimilarity as distances classifier using the SPWV distribution with different smoothing windows. The  $d_{LSD_1}$  distance is used. Performance evaluation using the 5-class classification problem (instances: A, B, C, D, E) and the 10-fold cross validation.

Smoothing window :	Gaussian	Hamming	Hanning	Rectangular	Parzen	Kaiser
Accuracy (%) :	94.40	88.80	95.40	79.40	95.40	95.00

D. According to accuracy and the average of the AUC, It is obvious that the LDA based dissimilarity as features is first ranked in terms of its classification accuracy of the EEG signals (98%), while the ANN base dissimilarity as features came second (97.60%). The SVM based dissimilarity as feature and  $k$ -NN based dissimilarity as distances had approximately the same accuracy ( $\sim 95\%$ ). The  $k$ -NN based dissimilarity as features ranked last in terms of its classification accuracy (88.80%). In addition, when the AUC is examined, the LDA and ANN based dissimilarity as features achieved high classification success with the values of 0.9875 and 0.9850 respectively. Thus, it can be seen that the performance of the SVM based dissimilarity as feature was approximately the same as that of the  $k$ -NN based dissimilarity as distances ( $\sim 0.9720$ ). Finally, the  $k$ -NN based dissimilarity as distances achieved a low AUC of 0.9300 compared to other classifiers.

**Summary:** The highest classification accuracy and AUC are obtained by the LDA (98%, 0.9875) and ANN (97.60%, 0.9850) based dissimilarity as features. The second-highest classification accuracy and AUC are obtained by the  $k$ -NN (95.40%, 0.9738) based dissimilarity as distances and SVM (94.20%, 0.9698) based dissimilarity as features. Thus, for all used classifiers, most errors are due to the misclassification between classes C and D. It should be mentioned that this does not have a major impact on the study, since both these sets are obtained during seizure-free instances of the same epileptic patients. The only difference between the EEG signals of the sets C and D is the zone of the brain from which they

Table 7: Parameter values of the proposed classifiers.

Classifier	parameter	Value
ANN	Number of neurons in the hidden layer	100
SVM	The parameter C that controls the trade-off between the complexity of decision rule and the training-error rate [48]	100
$k$ -NN	The number of neighbours $k$	1
LDA	No change	

Table 8: Classification performance of EEG epileptic signals using ANN, SVM,  $k$ -NN, ANN, based dissimilarity as features and  $k$ -NN based dissimilarity as distances. The SPWV distribution with Hanning smoothing window and the  $d_{LSD_1}$  distance were used. 5-class classification problem (instances: A, B, C, D, E). Performance evaluation using the 10-fold cross validation.

Approach	Classifier	Class	Specificity (%)	Sensitivity (%)	AUC	AUC (average)	Accuracy (%)
Dissimilarity as feature	ANN	A	100	100	1	0.9850	97.60
		B	99.00	100	0.9988		
		C	96.80	92.00	0.9562		
		D	92.40	97.00	0.9750		
		E	100	99.00	0.9950		
	SVM	A	84.70	100	1	0.9698	94.20
		B	99.00	99.00	0.9938		
		C	94.50	86.00	0.9238		
		D	96.70	87.00	0.9388		
		E	98.00	99.00	0.9925		
	$k$ -NN	A	99.00	99.00	0.9938	0.9300	88.80
		B	97.00	96.00	0.9763		
		C	78.40	80.00	0.8725		
		D	76.30	74.00	0.8413		
		E	93.10	95.00	0.9663		
	LDA	A	100	100	1	0.9875	98.00
		B	100	100	1		
		C	95.00	96.00	0.9738		
		D	95.00	95.00	0.9688		
		E	100	99.00	0.9950		
Dissimilarity as distance	$k$ -NN	A	91.50	97.00	0.9738	0.9738	95.40
		B	96.80	90.00	0.9463		
		C	93.30	98.00	0.9813		
		D	97.90	95.00	0.9725		
		E	100	99.00	0.9950		

were recorded. Which means that the EEG signals of these two sets are more similar compared to the signals of the other sets (A, B and E). As shown below, when sets C and D are considered as the same class, all classifiers achieved highest accuracies in most classification problems.

Tables 9 and 10 show the classification performances obtained by ANN, LDA, SVM based dissimilarity as features and  $k$ -NN based dissimilarity as distances over the classification instances of the 2 and 3-class classification problems. We can clearly see that all the classifiers had high performances in all studied instances. For all problems, all the classifiers had an AUC greater than or equal to 0.9900 and an accuracy greater than or equal to 99.30

#### 4.4. Approach complexity

Complexity can be viewed as two parameters i.e.: 1) time complexity, which is a measure or estimate of the running time of an algorithm, and 2) memory complexity, which is the storage space required by an algorithm to complete its execution. In the following, Complexity of the proposed approach is evaluated experimentally and approximated using the Big-O notation. Firstly, the running time, and memory space of the algorithm are measured according to the size of the learning database. Secondly, the results obtained are plotted and fitted using various functions. Finally, functions that give the best fit are considered.

##### -Time complexity:

The algorithm of the dissimilarity-based TFD approach is divided into three steps. The first one aims at extracting the TFDs from the EEG signals. The learning matrix (dissimilarity matrix) is assessed from the TFDs in the second step. The third one includes the learning and the testing algorithms of the classifier. To assess the complexity of each step, the running time is measured according to the number of signals in the learning database. Figures 4 and 5 plot curves of the running times for the three executed steps. The Table 11 summarizes the Big-O notation and the function name of the time complexities of each step.

In this experiment, the algorithm of the proposed approach is programmed using MATLAB. This software is installed on a computer with the following specifications: Intel(R) Xeon(R) CPU E5-1620 v2-3.70GHz, RAM-8Go, Windows 8 (64bits) operating system.

Table 9: Classification of EEG signals using ANN, LDA, and SVM (with linear kernel) based dissimilarity as features and  $k$ -NN based dissimilarity as distances. The SPWV distribution with Hanning smoothing window and the  $d_{LSD_1}$  distance were used. 2-class classification problem. Performance evaluation using the 10-fold cross validation.

classifier	ANN		LDA		$k$ -NN		SVM	
	AUC	Accuracy (%)	AUC	Accuracy (%)	AUC	Accuracy (%)	AUC	Accuracy (%)
A, E	<b>1</b>	<b>100</b>	<b>1</b>	<b>100</b>	<b>1</b>	<b>100</b>	<b>1</b>	<b>100</b>
B, E	<b>1</b>	<b>100</b>	<b>1</b>	<b>100</b>	0.9950	99.50	<b>1</b>	<b>100</b>
C, E	0.9500	99.50	<b>1</b>	<b>100</b>	<b>1</b>	<b>100</b>	0.9950	99.50
D, E	0.9950	99.50	0.9950	99.50	<b>1</b>	<b>100</b>	0.9950	99.50
AB, E	0.9950	99.70	<b>1</b>	<b>100</b>	0.9950	99.70	<b>1</b>	<b>100</b>
CD, E	0.9950	99.70	0.9950	99.70	<b>1</b>	<b>100</b>	0.9950	99.70
ACD, E	0.9933	99.50	0.9950	99.80	<b>1</b>	<b>100</b>	0.9950	99.80
BCD, E	0.9933	99.50	<b>0.9950</b>	<b>99.80</b>	<b>0.9950</b>	<b>99.80</b>	<b>0.9950</b>	<b>99.80</b>
ABCD, E	0.9938	99.60	<b>0.9950</b>	<b>99.80</b>	<b>0.9950</b>	<b>99.80</b>	0.9900	99.60

Table 10: Classification of EEG signals using ANN, LDA, and SVM (with linear kernel) based dissimilarity as features and  $k$ -NN based dissimilarity as distances. The SPWV distribution with Hanning smoothing window and the  $d_{LSD_1}$  distance were used. 3-class classification problem. Performance evaluation using the 10-fold cross validation.

classifier	ANN		LDA		$k$ -NN		SVM	
	AUC	Accuracy (%)	AUC	Accuracy (%)	AUC	Accuracy (%)	AUC	Accuracy (%)
A, C, E	<b>0.9975</b>	<b>99.70</b>	<b>0.9975</b>	<b>99.70</b>	<b>0.9975</b>	<b>99.70</b>	<b>0.9975</b>	<b>99.70</b>
A, D, E	<b>0.9975</b>	<b>99.70</b>	<b>0.9975</b>	<b>99.70</b>	<b>0.9975</b>	<b>99.70</b>	<b>0.9975</b>	<b>99.70</b>
AB, CD, E	0.9942	99.40	<b>0.9964</b>	<b>99.60</b>	0.9950	99.40	0.9938	99.20

The results obtained show that the complexity of the first and the second steps are  $O(n)$  and  $O(n^3)$ , respectively. The complexity of the third step depends on the type of classifier used. Note that the first and the second are the most time consuming. For example, with a learning database of 400 signals, 914 seconds are required to extract TFDs from the EEG signals, and 1076 seconds to calculate the dissimilarity matrix.

Firstly, the time complexities of the different steps of the algorithm are estimated separately. Now, in order to estimate the time complexity of the proposed approach, the running time of the whole algorithm is measured according to the number of signals in the learning database.

Figure 6 plots the curves of the running time of the proposed approach combined with the LDA, ANN, SVM and  $k$ -NN classifiers. One can clearly note that the time complexity of the proposed approach is  $O(n^3)$  whatever the classifier. This is due to the fact that the algorithm for calculating the dissimilarity matrix is the one that determines the complexity time of the proposed approach.

**-Memory complexity:**

The algorithms of the proposed approach use numerical data like TFDs and dissimilarity matrix. This data is firstly stored in the memory and is subsequently used by the algorithms. To assess the memory consumption of the proposed approach, the memory required to store the data is studied according to the size of the learning database (Figure 7). Note that the data is stored as numerical values with a double-precision floating point. Based on the fitted curve in figure 7, one can note that the space complexity of the proposed approach is  $O(n)$ .

4.5. Robustness of the approach

To get an idea about the influence of noise on the applicability of the proposed approach, the performance of classifiers are evaluated according to the signal-to-noise ratios (SNRs) of clean EEG signals.

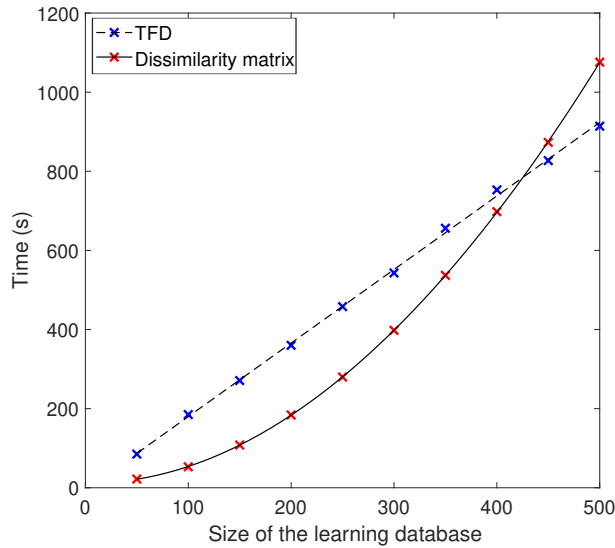


Figure 4: The running time of the first (TFD) and the second (Dissimilarity matrix) steps, as a function of the learning database size (number of signals)

Table 11: Time complexities of the three steps of the algorithm used in the proposed approach.

Algorithm step	TFD	dissimilarity matrix	LDA	ANN	SVM	$k$ -NN
Big-O notation	$O(n)$	$O(n^3)$	$O(n)$	$O(n^3)$	$O(n^3)$	$O(n)$
Function name	linear time	Quadratic time	linear time	Quadratic time	Quadratic time	linear time

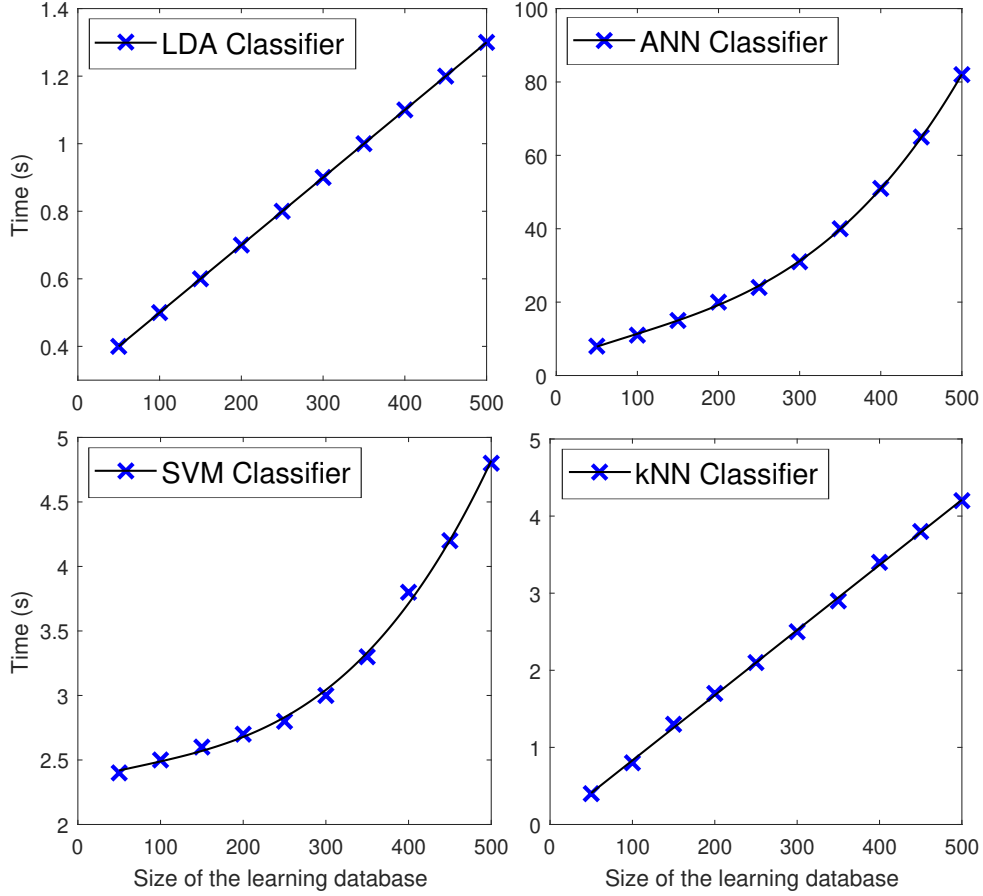


Figure 5: The running time of the third step using the classifier: LDA, ANN, SVM and  $k$ -NN, as a function of the learning database size (number of signals)

The SNR is given as:  $SNR(dB) = 10\log(P_s/P_n)$ , where  $P_s$  and  $P_n$  are the signal and noise powers, respectively. We have considered EEG signals of the entire database corrupted with white Gaussian noise at a given SNR range from 0 dB to 50dB. Classifiers are turned using 2-class (instance: ABCD-E), 3-class (instance: AB-CD-E) and 5-class classification problems and their performances are checked using 10-fold cross validation and accuracy. Figures 8-a, 8-b and 8-c plot the classification accuracy as function of the SNR.

As shown in Figures 8-a and 8-b, the proposed approach performs better in the 2-class and 3-class classification problems, the classification accuracy values are generally insignificantly deteriorated as the SNR decreases. For example, at 0db SNR, there is less than 3% and 10% deterioration for the 2-class and 3-class classification problems, respectively. In addition, for each classifier, there is no significant change in the classification accuracy under clean conditions and at 1 dB SNR.

As shown in figure 8-c, the performance of the proposed approach in the 5-class classification problem is far from satisfactory when the SNR is less than 10db. The classification accuracy values are significantly deteriorated as the SNR decreases. This is because of the high dependency between classes C and D, where most errors are due to misclassification between them. However, for each classifier, there is no significant change in the classification accuracy under clean conditions and at 20 dB SNR.

These results clearly indicate that the proposed approach is robust to noise. Even though unsatisfac-



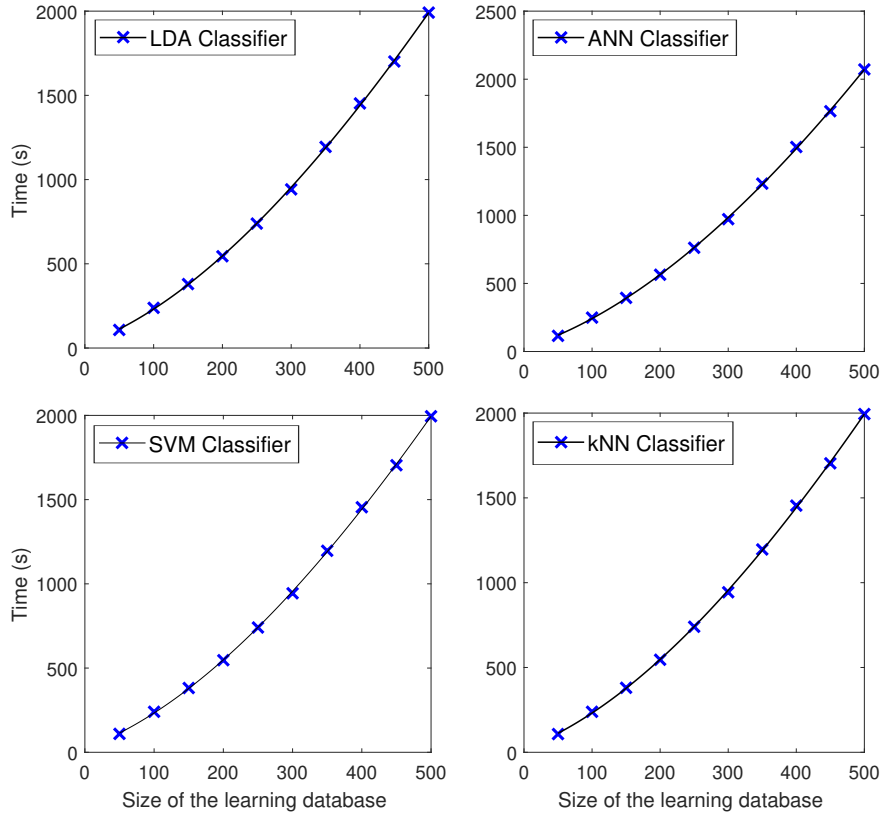


Figure 6: The running time plot of the whole algorithm of the proposed approach combined with LDA, ANN, SVM and the  $k$ -NN classifiers, as a function of the learning database size (number of signals).

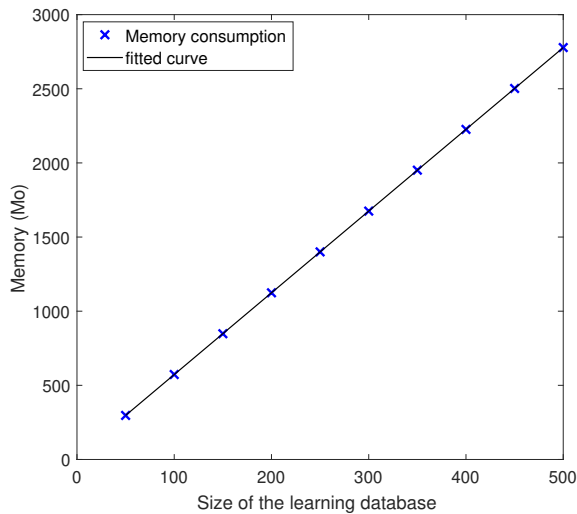


Figure 7: Plots and fitted curve of the memory consumption by the proposed approach, as a function of the learning database size (number of signals).

tory accuracy is obtained in the 5-class classification problem, a low SNR cannot affect the applicability of the proposed approach to discriminate EEGs signals of healthy people from those of epileptic patients.

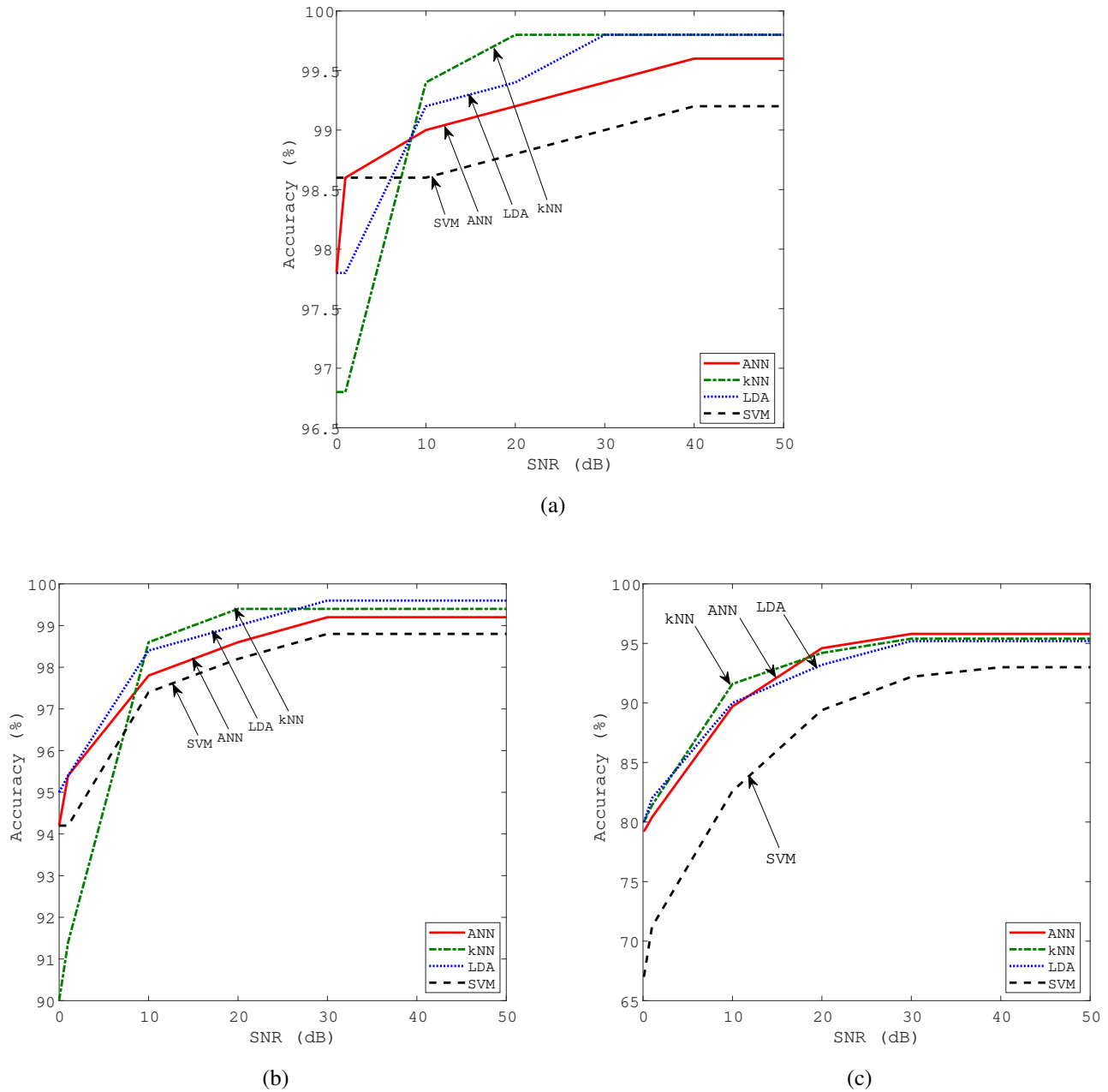


Figure 8: Classification accuracy as function of the SNR. (a) Classifiers evaluated using the 2-class classification problem. (b) Classifiers evaluated using the 3-class classification problem. (c) Classifiers evaluated using the 5-class classification problem.

#### 4.6. Comparison with the literature results

The proposed approach is combined with four different classifiers. The performance of these classifiers is evaluated against 13 classification problems. Experimental results showed that the used classifiers achieved highest accuracies in most of the studied classification problems. However, in order to compare the developed approach with the literature approaches, it is necessary to select the most efficient classifier. Thus, the selection strategy based on accuracy competence allowed the LDA classifier to be selected because it is the one reaching the highest accuracies in the most classification problems (10 out of 13 classification problems). For this reason, only the classification accuracies of the LDA classifier are compared to the literature results.

The comparison of the results in terms of accuracy reported in literature for EEG epileptic classification approaches published from 2008 to 2018 [3, 4, 10–28] with those of the proposed approach is presented in Tables 12, 13, 14, and 15. All of these approaches are evaluated on the same database [47] as for the proposed approach. The results of the proposed approach that have been compared are those obtained by the LDA classifier. Although Dhiman et al. [18] and Das et al. [21] reported a 100% classification rate for classification instances A-E, B-E, C-E, D-E, AB-E, ABCD-E, and A-D-E, respectively, the proposed approach outperforms approximately the same results for the remaining classification instances but without using features extraction. Yet, the proposed approach has the best classification rate for classification instances A-B-C-D-E (98%), ACD-E (99.80%), BCD-E (99.80%), CD-E (99.80%), A-C-E (99.70%), and AB-CD-E (99.60%) compared to the reported results by Ahnaf et al. [19], Mousavi et al.[10], and Kumar et al. [4].

#### 4.7. Discussion

Table 12 shows a brief summary of studies that present various approaches to epileptic seizure detection. It is very important to note that almost all-known classifiers were used for detection of epileptic seizures in EEG signals. All these classifiers were performed using the approach of features extraction. However, when the features are not appropriate for the given classification problem, the performances obtained are unsatisfactory. Because of the improper features, the classification algorithm cannot generate high performances. Therefore, it is mandatory to find and extract suitable features from the raw signals to be able to obtain good classification results. On the other hand, our study showed that the detection of epileptic seizures in EEG signals can now be carried out using the dissimilarity-based approach and without using features extraction. Except for the choice of an appropriate TFD and dissimilarity measure, the dissimilarity-based approach has the advantage of being generalized for all medical signals, which is not the case for the feature-based approach.

Table 12: Various approaches used for detection of epileptic seizures applied to dataset [47]. Works published from 2004 to 2018.

Authors	Year	Feature	Classifier
Mousavi et al.[10]	2008	AR model , wavelet decomposition	Multi-Layer Perceptron(MLP)
Tzallas et al. [11]	2009	Features based TFD	SVM
Liang et al.[13]	2010	Fast Fourier Transform, approximate entropy analysis	SVM
Kumar et al. [14]	2014	Digital Wavelet Transform based fuzzy approximate entropy	SVM
Kang et al. [16]	2015	Short-Time Fourier Transform	Quadratic Discriminant Analysis
Dhiman et al. [18]	2016	Wavelet Packet Transform	Twin SVM
Ahnaf et al. [19]	2016	Tunable-Q factor wavelet transform	Bootstrap aggregating
Das et al. [21]	2016	Dual-tree complex wavelets, inverse Gaussian	SVM
Alcin et al. [20]	2016	STFT, Gray Level Co-occurrence Matrix, Fisher Vector	Extreme Learning Machine
Bhati et al. [22]	2017	Time–frequency localized three-band synthesis filter bank	MLPNN
Biju et al. [23]	2017	Energy/entropy of IA and SD of IF from Hilbert-Huang Transform	ANN
Li et al. [24]	2017	DWT-based EA	Neural Network Ensemble
Sharma et al. [25]	2017	analytic time-frequency flexible wavelet transform and fractal dimension	LS-SVM
Tiwari et al. [30]	2017	Key-Point Descriptor of EEG Signals	SVM
Zahra et al. [26]	2017	Multivariate EMD	ANN
Jaiswal et al. [27]	2017	Local Neighbor Descriptive Pattern and One-dimensional Local Gradient Pattern feature extraction techniques	ANN
Li and al. [28]	2018	wavelet packet transform	Least squares support vector machine
<b>This work</b>	<b>2018</b>	<b>dissimilarity-based TFD</b>	<b>ANN, LDA, <math>k</math>-NN, SVM</b>

Table 13: Comparison of accuracy of the various methods for detection of epileptic seizures applied to dataset from [47]. Comparison of previously reported results for 2 class classification problem. Classification instances out of (A, B, C, D, E).

Work	A, E	B, E	C, E	D, E	AB, E	CD, E	ACD, E	BCD, E	ABCD, E
[14]	<b>100</b>	<b>100</b>	99.60	95.80	-	-	98.15	98.22	97.38
[16]	<b>100</b>	<b>100</b>	99.86	97.97	-	-	-	-	-
[18]	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	-	-	-	-	<b>100</b>
[19]	<b>100</b>	-	<b>100</b>	<b>100</b>	-	-	-	-	99.60
[21]	<b>100</b>	-	<b>100</b>	<b>100</b>	-	-	-	-	<b>100</b>
[22]	-	-	-	-	-	99.33	-	-	-
[23]	-	-	-	<b>100</b>	-	-	-	-	-
[25]	<b>100</b>	<b>100</b>	99	98.50	<b>100</b>	98.67	-	-	99.20
[30]	-	-	-	-	<b>100</b>	99.45	-	-	99.31
[27]	99.82	99.25	99.10	99.07	-	98.88	-	-	98.72
<b>This work</b>	<b>100</b>	<b>100</b>	<b>100</b>	99.50	<b>100</b>	<b>99.70</b>	<b>99.80</b>	<b>99.80</b>	99.80

Table 14: Comparison of accuracy of the various methods for detection of epileptic seizures applied to dataset from [47]. Comparison of previously reported results for 3-class classification problem. Classification instances out of (A, B, C, D, E).

Work	A,C,E	A,D,E	AB,CD,E
[10]	91-96	-	-
[21]	-	<b>100</b>	96.28
[19]	-	98.67	98.40
[24]	-	98.78	-
[30]	-	-	98.80
[27]	-	98.22	-
[28]	-	99.60	-
<b>This work</b>	<b>99.70</b>	99.70	<b>99.60</b>

Table 15: Comparison of accuracy and AUC of the various methods for detection of epileptic seizures applied to dataset from [47]. Comparison of previously reported results for 5-class classification problem. Classification instances: A, B, C, D, E.

Work	Instance	Accuracy (%)
[11]	A, B, C, D, E	89
[13]	A, B, C, D, E	85.90
[20]	A, B, C, D, E	96.40
[26]	A, B, C, D, E	87.20
<b>This work</b>	A, B, C, D, E	<b>98.00</b>

Tables 13, 14, and 15 present the comparison of the results in terms of classification accuracy reported in literature for EEG epileptic classification with those of the proposed approach. The proposed approach achieved high classification performances in the 13 classification problems studied (2, 3, and 5-class classification problems). For the reported 2-class classification problems, the accuracy obtained from the proposed approach is the best presented for the instances (A-E, B-E, C-E, CD-E, ACD-E and BCD-E) and is the second-best presented for the instances D-E and ABCD-E. For the reported 3-class classification problems, the accuracy obtained from the proposed approach is the best presented for the instances (AB-CD-E and A-C-E) and is the second-best presented for the instance A-D-E. For the reported 5-class classification problem, the accuracy obtained from the proposed approach is the best presented for the instance A-B-C-D-E.

For most of the classification problems of EEG signals studied, classifiers combined with the proposed approach achieved highest accuracies and AUC. Errors were mainly due to the misclassification between classes C and D. That can be explained by the high dependence between classes C and D (interictal EEGs of epileptic patients recorded by intracranial electrodes). However, that does not affect the applicability of the proposed approach to discriminate:

- EEGs signals of the healthy people and ictal patients.
- EEGs signals of the healthy people and interictal patients.

- EEGs signals of the ictal and interictal patients.

## 5. Conclusion

Several solutions have been proposed to automate the classification of EEG signals for epileptic seizure detection. In the traditional approach, firstly, feature vectors are extracted from the TFD of EEG signals. Secondly, classifiers are ran on the feature space in order to construct a decision rule. The success of this approach heavily depends on the best choice of the feature.

In this paper, we proposed an approach for the classification of EEG signals based on dissimilarities between their TFDs, called dissimilarity-based TFD. This approach did not require direct access to the features of signals. Classifiers based on this approach estimated the class label of a test signal based on the dissimilarity between the TFD of the test signal and TFD of labeled training signals. Experimental results showed that the proposed approach achieved highest accuracy in most of the classification problems studied. A high value of 98% was achieved for the 5-class problem. Further, in most classification problems with 2 and 3-class, it also yielded a satisfactory accuracy of approximately 100%. The comparison between the proposed approach and literature studies, confirmed that the proposed approach had a potential in the classification of EEG signals and the detection of epileptic seizures.

The robustness of the proposed approach was explored with the addition of noise to the EEG signals at various SNRs. The experimental results showed when  $\text{SNR} \geq 1\text{dB}$ , the classification accuracies were more than 98%, 95% and 82% for the 2-class, 3-class and 5-class classification problems. These results clearly indicated that the proposed approach was robust to noise. And the latter could not affect the applicability of this approach for the diagnosis of epilepsy using EEG signals.

The dissimilarity-based TFD approach required the use of a TFR and a dissimilarity measure to compute a dissimilarity matrix. Thus, a classifier combined with this approach used the dissimilarity matrix in order to build its decision rule. Indeed, the success of this classifier was directly related to a wise choice of TFD and a dissimilarity measure. A bad choice of one of these two could considerably degrade the accuracy of the classifier. On the other hand, the dissimilarity matrix represented the pairwise dissimilarities between TFDs of the learning objects. Thus, the choice of the TFR and the dissimilarity measure could depend on the learning objects.

The main problem of the dissimilarity-based approach remained in the high dimensionality of the spatial features related to the resulting dissimilarity. As shown in the experimental results, complexities of the algorithms used in the proposed approach with the different classifiers were  $O(n^3)$  for time complexity and  $O(n)$  for space complexity.

We suggest three types of solutions to solve the problem of the complexity of the proposed approach. The first consists in building the dissimilarity space using all available objects, and subsequently applying some standard dimensionality reduction techniques (such as principal component analysis, t-distributed stochastic neighbour embedding ...). The second one is to reduce the dissimilarity space dimension by using an approach of prototype selection [49]. The third one can be conducted by directly choosing a small set of objects.

## Acknowledgements

The authors would like to gratefully acknowledge the Champagne-Ardenne Region for its financial support to the project, with which this work has been conducted.

## Conflict of interest

We declare that we have no conflict of interest.

## References

- [1] P. E. McSharry, T. He, L. A. Smith, L. Tarassenko, Linear and non-linear methods for automatic seizure detection in scalp electro-encephalogram recordings, *Medical and Biological Engineering and Computing* 40 (4) (2002) 447–461.
- [2] F. Mormann, R. G. Andrzejak, C. E. Elger, K. Lehnertz, Seizure prediction: the long and winding road, *Brain* 130 (2) (2006) 314.
- [3] J.-L. Song, W. Hu, R. Zhang, Automated detection of epileptic EEGs using a novel fusion feature and extreme learning machine, *Neurocomputing* 175, Part A (2016) 383 – 391.
- [4] T. S. Kumar, V. Kanhangad, R. B. Pachori, Classification of seizure and seizure-free EEG signals using local binary patterns, *Biomedical Signal Processing and Control* 15 (2015) 33 – 40.
- [5] Y. Chen, E. K. Garcia, M. R. Gupta, A. Rahimi, L. Cazzanti, Similarity-based classification: Concepts and algorithms, *Journal of Machine Learning Research* 10 (Mar) (2009) 747–776.
- [6] R. P. Duin, E. Pekalska, The dissimilarity representation for structural pattern recognition, in: *Iberoamerican Congress on Pattern Recognition*, Springer, 2011, pp. 1–24.
- [7] E. Pękalska, R. P. Duin, *The dissimilarity representation for pattern recognition: foundations and applications*, World Scientific, 2005.
- [8] R. P. Duin, E. Pękalska, The dissimilarity space: Bridging structural and statistical pattern recognition, *Pattern Recognition Letters* 33 (7) (2012) 826–832.
- [9] R. P. Duin, M. Loog, E. Pekalska, D. M. Tax, Feature-based dissimilarity space classification, in: *Recognizing Patterns in Signals, Speech, Images and Videos*, Springer, 2010, pp. 46–55.
- [10] S. Mousavi, M. Niknazar, B. V. Vahdat, Epileptic seizure detection using ar model on EEG signals, in: *Biomedical engineering conference, 2008. CIBEC 2008. Cairo International, IEEE, 2008*, pp. 1–4.
- [11] A. T. Tzallas, M. G. Tsipouras, D. I. Fotiadis, Epileptic seizure detection in EEGs using time-frequency analysis, *IEEE Transactions on Information Technology in Biomedicine* 13 (5) (2009) 703–710.
- [12] K. Polat, S. Güneş, Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast fourier transform, *Applied Mathematics and Computation* 187 (2) (2007) 1017 – 1026.
- [13] S.-F. Liang, H.-C. Wang, W.-L. Chang, Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection, *EURASIP Journal on Advances in Signal Processing* 2010 (1) (2010) 853434.
- [14] Y. Kumar, M. Dewal, R. Anand, Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine, *Neurocomputing* 133 (2014) 271 – 279.

- [15] K. Fu, J. Qu, Y. Chai, Y. Dong, Classification of seizure based on the time-frequency image of EEG signals using HHT and SVM, *Biomedical Signal Processing and Control* 13 (2014) 15 – 22.
- [16] J.-H. Kang, Y. G. Chung, S.-P. Kim, An efficient detection of epileptic seizure by differentiation and spectral analysis of electroencephalograms, *Computers in biology and medicine* 66 (2015) 352–356.
- [17] R. Sharma, R. B. Pachori, Classification of epileptic seizures in EEG signals based on phase space representation of intrinsic mode functions, *Expert Systems with Applications* 42 (3) (2015) 1106 – 1117.
- [18] R. Dhiman, J. Saini, Priyanka, Biogeography based hybrid scheme for automatic detection of epileptic seizures from EEG signatures, *Applied Soft Computing* 51 (2017) 116 – 129.
- [19] A. R. Hassan, S. Siuly, Y. Zhang, Epileptic seizure detection in EEG signals using tunable-q factor wavelet transform and bootstrap aggregating, *Computer Methods and Programs in Biomedicine* 137 (2016) 247 – 259.
- [20] O. F. A., S. Siuly, V. Bajaj, Y. Guo, A. Sengur, Y. Zhang, Multi-category EEG signal classification developing time-frequency texture features based fisher vector encoding method, *Neurocomputing* 218 (2016) 251 – 258.
- [21] A. B. Das, M. I. H. Bhuiyan, S. S. Alam, Classification of EEG signals using normal inverse gaussian parameters in the dual-tree complex wavelet transform domain for seizure detection, *Signal, Image and Video Processing* 10 (2) (2016) 259–266.
- [22] D. Bhati, M. Sharma, R. B. Pachori, V. M. Gadre, Time–frequency localized three-band biorthogonal wavelet filter bank using semidefinite relaxation and nonlinear least squares with epileptic seizure EEG signal classification, *Digital Signal Processing* 62 (2017) 259 – 273.
- [23] B. K.S., H. A. Hakkim, J. M.G., Ictal EEG classification based on amplitude and frequency contours of IMFs, *Biocybernetics and Biomedical Engineering* 37 (1) (2017) 172 – 183.
- [24] M. Li, W. Chen, T. Zhang, Classification of epilepsy EEG signals using dwt-based envelope analysis and neural network ensemble, *Biomedical Signal Processing and Control* 31 (2017) 357 – 365.
- [25] M. Sharma, R. B. Pachori, U. R. Acharya, A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension, *Pattern Recognition Letters* 94 (Supplement C) (2017) 172 – 179.
- [26] A. Zahra, N. Kanwal, N. ur Rehman, S. Ehsan, K. D. McDonald-Maier, Seizure detection from eeg signals using multivariate empirical mode decomposition, *Computers in Biology and Medicine* 88 (Supplement C) (2017) 132 – 141.
- [27] A. K. Jaiswal, H. Banka, Local pattern transformation based feature extraction techniques for classification of epileptic eeg signals, *Biomedical Signal Processing and Control* 34 (Supplement C) (2017) 81 – 92.
- [28] M. Li, W. Chen, T. Zhang, A novel seizure diagnostic model based on kernel density estimation and least squares support vector machine, *Biomedical Signal Processing and Control* 41 (2018) 233 – 241.

- [29] E. Sejdić, I. Djurović, J. Jiang, Time–frequency feature representation using energy concentration: An overview of recent advances, *Digital Signal Processing* 19 (1) (2009) 153–183.
- [30] A. K. Tiwari, R. B. Pachori, V. Kanhangad, B. K. Panigrahi, Automated diagnosis of epilepsy using key-point-based local binary pattern of eeg signals, *IEEE Journal of Biomedical and Health Informatics* 21 (4) (2017) 888–896.
- [31] M. M. Deza, E. Deza, Encyclopedia of distances, in: *Encyclopedia of Distances*, Springer, 2009, pp. 1–583.
- [32] F. Esposito, D. Malerba, V. Tamma, H. Bock, Similarity and dissimilarity, in: *Analysis of Symbolic Data*, Springer, 2000, pp. 139–197.
- [33] G. Wu, E. Y. Chang, Z. Zhang, An analysis of transformation on non-positive semidefinite similarity matrix for kernel machines, in: *Proceedings of the 22nd International Conference on Machine Learning*, Vol. 8, 2005.
- [34] E. Pękalska, R. P. Duin, Dissimilarity representations allow for building good classifiers, *Pattern Recognition Letters* 23 (8) (2002) 943 – 956.
- [35] K. Gröchenig, *Foundations of time-frequency analysis*, Springer Science & Business Media, 2013.
- [36] L. Cohen, *Time-frequency analysis*, Vol. 778, Prentice hall, 1995.
- [37] L. Cohen, Time-frequency distributions-a review, *Proceedings of the IEEE* 77 (7) (1989) 941–981.
- [38] P. Flandrin, W. Martin, A general class of estimators for the wigner-ville spectrum of non-stationary processes, in: *Analysis and Optimization of Systems*, Springer, 1984, pp. 15–23.
- [39] F. Hlawatsch, P. Flandrin, The interference structure of the wigner distribution and related time-frequency signal representations, *The Wigner Distribution—Theory and Applications in Signal Processing* (1997) 59–133.
- [40] F. Hlawatsch, T. G. Manickam, R. L. Urbanke, W. Jones, Smoothed pseudo-wigner distribution, choi-williams distribution, and cone-kernel representation: Ambiguity-domain analysis and experimental comparison, *Signal Processing* 43 (2) (1995) 149–168.
- [41] M. Basseville, Distance measures for signal processing and pattern recognition, *Signal processing* 18 (4) (1989) 349–369.
- [42] S. Aviyente, Divergence measures for time-frequency distributions, in: *Seventh International Symposium on Signal Processing and Its Applications*, 2003. *Proceedings.*, Vol. 1, 2003, pp. 121–124.
- [43] I. Vincent, C. Doncarli, E. L. Carpentier, Non stationary signals classification using time-frequency distributions, in: *Proceedings of IEEE-SP International Symposium on Time- Frequency and Time-Scale Analysis*, 1994, pp. 233–236.
- [44] A. P. Bradley, The use of the area under the roc curve in the evaluation of machine learning algorithms, *Pattern recognition* 30 (7) (1997) 1145–1159.
- [45] R. Kohavi, A study of cross-validation and bootstrap for accuracy estimation and model selection, in: *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2, IJCAI'95*, Morgan Kaufmann Publishers Inc., 1995, pp. 1137–1143.



- [46] T. Fawcett, An introduction to roc analysis, *Pattern recognition letters* 27 (8) (2006) 861–874.
- [47] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C. E. Elger, Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state, *Phys. Rev. E* 64 (2001) 061907.
- [48] C. Cortes, V. Vapnik, Support-vector networks, *Machine learning* 20 (3) (1995) 273–297.
- [49] E. Pękalska, R. P. Duin, P. Paclík, Prototype selection for dissimilarity-based classifiers, *Pattern Recognition* 39 (2) (2006) 189 – 208, part Special Issue: Complexity Reduction.