QUEEN'S UNIVERSITY BELFAST

Non-linear classifiers applied to EEG analysis for epilepsy seizure detection

Martinez del Rincon, J., Santofimia, M. J., del Toro Garcia, X., Barba , J., Romero, F., Navas, P., & Lopez, J. C. (2017). Non-linear classifiers applied to EEG analysis for epilepsy seizure detection. *Expert Systems with Applications*, *86*, 99-112. https://doi.org/10.1016/j.eswa.2017.05.052

Published in:
Expert Systems with Applications

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

Publisher rights
© 2017 Elsevier.
This manuscript is distributed under a Creative Commons Attribution-NonCommercial-NoDerivs License (https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access
This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: http://go.qub.ac.uk/oa-feedback

Download date:09. Jul. 2025

Non-linear classifiers applied to EEG analysis for epilepsy seizure detection

Jesus Martinez-del-Rincon[b], Maria J. Santofimia[a,*], Xavier del Toro[a], Jesus Barba[a], Francisca Romero[c], Patricia Navas[c], Juan C. Lopez[a]

[a] *Computer Architecture and Networks Group, University of Castilla-La Mancha, Paseo de la Universidad, 4, Ciudad Real, Spain*
[b] *Centre for Secure Information Technologies, School of EEECS, Queens University Belfast, BT3 9DT, UK*
[c] *Hospital Regional Universitario Carlos Haya, Av. de Jorge Silvela, 8, 29010 Málaga, Spain*

Abstract

This work presents a novel approach for automatic epilepsy seizure detection based on EEG analysis that exploits the underlying non-linear nature of EEG data. In this paper, two main contributions are presented and validated: the use of non-linear classifiers through the so-called kernel trick and the proposal of a Bag-of-Words model for extracting a non-linear feature representation of the input data in an unsupervised manner. The performance of the resulting system is validated with public datasets, previously processed to remove artifacts or external disturbances, but also with private datasets recorded under realistic and non-ideal operating conditions. The use of public datasets caters for comparison purposes whereas the private one shows the performance of the system under realistic circumstances of noise, artifacts, and signals of different amplitudes. Moreover, the proposed solution has been compared to state-of-the-art works not only for pre-processed and public datasets but also with the private datasets. The mean F1-measure shows a 10% improvement over the second-best ranked method including cross-dataset experiments. The obtained results prove the

*Corresponding author
Email addresses: j.martinez-del-rincon@qub.ac.uk (Jesus Martinez-del-Rincon), mariajose.santofimia@uclm.es (Maria J. Santofimia), Xavier.delToro@uclm.es (Xavier del Toro), jesus.barba@uclm.es (Jesus Barba), franciscaromerocrespo@gmail.com (Francisca Romero), patricianavas@gmail.com (Patricia Navas), juancarlos.lopez@uclm.es (Juan C. Lopez)

robustness of the proposed solution to more realistic and variable conditions.

Keywords: Classification algorithms, Non-linear classifiers, SVM, Bag of words, Wavelet, Epilepsy

1. Introduction

Epilepsy is a disease that affects approximately 1% of the world's population Shoeb et al. (2004). This neurological disorder might cause a loss of consciousness, muscle jerks or, in the most severe cases, prolonged convulsions. Its effects have a significant impact on the patient's quality of life as well as other important social and economic considerations, due to health-care needs, premature death and/or loss of productivity Organization (2016).

Epilepsy diagnosis is a tedious, expensive and time-consuming task, which is performed by highly trained professionals who examine EEG data in seeking abnormal brain activity. Currently, neurophysiologists analyse long EEG logs that should ideally record as much cerebral activity as possible to increase the probability of recording seizure occurrences. This manual analysis of EEG is therefore the current bottleneck in the epilepsy diagnosis stage and, as consequence, in the process of providing a treatment for epileptic patients. Despite the great impact that epilepsy has on society, there are few computational systems or tools that support automatic analysis and categorisation of EEG recordings. The lack of reliable systems for automatic epilepsy diagnosis is not casual. In contrast, several reasons seem to be responsible for this scarcity, such as the great variability found among individuals and the overlapping among seizure and non-seizure states Echauz et al. (2008).

This work proposes and analyses two expert systems for epilepsy diagnosis that exploit the non-linear separability of the data. More importantly, this paper demonstrates their expert-system performance under realistic and variable conditions, similar to the ones that would be found in a real hospital environment. For this reason, special emphasis has been made in this paper to demonstrate the robustness of the solution regardless the training data and in cross-dataset experiments.

1.1. Previous work

Many different approaches have been proposed for automatic seizure detection and epilepsy diagnosis, for the sake of simplicity, we will mention some of the most relevant but for a thorough analysis of the state of the art, please refer to Tsiouris et al. (2015); Alotaiby et al. (2014).

The first acknowledged and widely used approach for automatic recognition of epileptic seizures based on EEG analysis was proposed in Gotman (1982, 1990) by Gotman. The approach presented in this work consists of quantitatively measuring the novelty of the EEG signal. Therefore, a continuous temporal analysis is performed that compares one epoch or EEG signal segment against a reference or background segment. Gotman's Monitor algorithm employs a set of rules for identifying and triggering seizures. The work of Wilson et al. in the Reveal algorithm Wilson et al. (2004) also relied on the analysis of EEG tendencies and a rule-based system to identify potential seizure scenarios. However, Wilson introduced analysis of frequency parameters.

Methods that combine time and spectral analysis of an EEG signal have showed an improvement in the success ratios for seizure detection in contrast to those that only focus on one domain. In this regard, the wavelet transform is one of the most frequently used signal processing algorithms for EEG analysis (see Faust et al. (2015) for a detailed summary of published research on EEG signal feature extraction using DWT).

As a common stage of all current approaches, after characterising the signal either in time or frequency, a decision must be made as to whether the EEG signal presents the characteristics of a seizure or not. This decision is supported by the use of a classifier that has as inputs several signal features that are computed from the EEG data after the pre-processing stage. There is a variety of methods that have been used to characterise the pre-processed EEG record: entropies Acharya et al. (2015), energy distribution Omerhodzic et al. (2013); Orhan et al. (2011); Patnaik & Manyam (2008); quantitative statistical variables such as the mean, standard derivation, variance, inter-quartile range and other measurements Pippa et al. (2015); autoregressive models (AR) Atyabi et al. (2016); Chen (2014); or independent component analysis Siuly & Li (2015), just to name some of the most promising approaches. The type and number of such features has a direct impact in the behaviour of the system. Thus, it is necessary to select the most appropriate techniques to maximise the recognition rates. The work in Upadhyay et al. (2016) carries out a comparative study of feature ranking techniques.

Given the complex and non-linear nature of EEG, any feature extraction technique that can detect and quantify some aspect of these non-linear mechanisms are specially relevant in distinguishing different types of EEG signals (normal, ictal, interictal). Thus, the use of Higher Order Spectra (HOS) is studied in Chua et al. (2008) Chua et al. (2011) to conclude that the analysed parameters are statistically significant therefore appropriate for the classification of EEG signals. Recurrent Quantification Analysis (RQA) Acharya et al. (2011b) parameters yields an accuracy result of 95.6% when run with Support Vector Machine (SVM) classifiers. The work in Acharya et al. (2011a) report the use of Higher Order Cumulant features (HOC). This study reports an accuracy rate of 98.5% when used with SVM classifiers. The work in Martis et al. (2013) proposes the use of a novel method, as it is the Intrinsic time-scale decomposition (ITD), to compute features for the automated classification process. Accuracy rate of 95.67% was reported in this study. Spectral and embedding entropy Kannathal et al. (2005); Acharya et al. (2012a), used to measure the system complexities, and Lyapunov exponents Guler & Ubeyli (2007) have been also employed to epilepsy detection in EEG analysis.

Regarding classification strategies, the existing literature mainly reveals two different approaches in EEG analysis for automatic seizure detection: non-linear methods, particularly Artificial Neural Networks (ANN) Alfaro-Ponce et al. (2016); Omerhodzic et al. (2013); Husain & Rao (2012); Orhan et al. (2011); Patnaik & Manyam (2008); Tzallas et al. (2007); Bao et al. (2008); N & Thanushkodi (2009) but also Decision Tress (DT) Martis et al. (2013); Polat &

Table 1: Summary of most relevant state-of-the-art works for automated EEG analysis

Reference	Features	Classifier	Accuracy(%)	Dataset Size [Sequences]
Chua et al. (2008)	HOS features	GMM	93.3	100
Acharya et al. (2011b)	RQA	SVM	94.3	300
Martis et al. (2013)	IDT	NN/DT	95.67	100
Acharya et al. (2011a)	HOC	SVM	98.5	300
Acharya et al. (2012a)	Entropy	Fuzzy Inference	98.1	300
Acharya et al. (2012b)	Entropy+HOS+others	Fuzzy Inference	99.7	300
Polat & Güneş (2007)	FFT	DT	98.72	100
Guo et al. (2009)	Relative Wavelet Energy	ANN	99.6	200
Wang et al. (2012)	DWT+Bag of Words	ANN	99.2	500
Guler & Ubeyli (2007)	DWT+lyapunov exponents	SVM	99.3	500
Janjarasjitt (2010)	Wavelet-Based Scale Variance	k-means	97.6	300
Husain & Rao (2012)	DWT-based features	ANN	98.2	500
Fathima et al. (2011)	DWT-based features	Linear classifier	99.8	500
Chen (2014)	DTCWF	Nearest Neighbour	100	500
Übeyli (2010)	Burg AR	least squares SVM	99.56	200

Güneş (2007), and linear classifiers such as Gaussian Mixture Models (GMM) Chua et al. (2011), SVM Direito et al. (2014) or k-means clustering Janjarasjitt (2010). Alternatively, other machine learning algorithms, such as Genetic Programming Bhardwaj et al. (2016) have also been proposed in this field.

However, these previous techniques have been evaluated in simple and relatively small datasets such as the University of Bonn dataset in which only one type of variation or activity modality is present, which explains the high accuracy rates achieved by simple and linear methods. Furthermore, the methods are retrained for each datataset and parameters have been manually tuned for the testing set rather than using automatic optimisation techniques as in other fields Valipour (2016); Valipour & Singh (2016); Yannopoulos et al. (2015); Valipour (2012b,a). This results in overfitting to the specific dataset, which means that a significant performance drop is expected when testing in a different dataset or under more realistic and challenging scenarios with different activity variations and noise presence.

Table 1 summarises a comparative analysis of the most relevant works of the state of the art for automated EEG analysis for epilepsy diagnosis.

1.2. Proposed system

The present work proposes two systems for automatic epilepsy seizure detection. Both systems are based on EEG analysis and inspired by non-linear classifiers and the Bag-of-Words model Joachims (1997), which has been previously used in fields such as natural language understanding or computer vision Cheng et al. (2010); Gilbert et al. (2009) to deal with multiple sources of noise and variation. The goal is to analyse the behaviour of both systems and study their suitability and robustness using datasets with different characteristics in terms of noise, signal attenuation, presence of artifacts, or the type of activity being recorded (ictal, inter-ictal, normal with artifacts, etc.). Furthermore, cross-dataset testing will be employed to ensure that the results are representative of the real expected performance. The accuracy of the results obtained by the proposed system is compared to the performance of a linear classifier and the state of the art. Our proposal outperforms the most representative and relevant state-of-the-art works and its performance is stable across datasets. Moreover, this proposal has been demonstrated to be computationally efficient.

2. Background

2.1. Wavelet transforms

The analysis of EEG for seizure detection is mostly performed in the time and frequency domains. The simplest and most straight-forward technique, as performed by neurophysiologists, is the visual inspection of the EEG time series, which does not require any additional manipulation of the EEG data. Additional information in the time domain can be obtained by means of simple calculations on the time series, such as the average, median, and standard deviation values. Nevertheless, it is generally more interesting to analyse transients and changes in the EEG signals by means of calculating the rate of change, moving average, autocorrelations, and autoregressions.

The frequency content of the EEG signals provides very valuable information, but it is difficult to extract from the visual analysis in the time domain.

Moreover, certain manipulations and signal processing techniques, such as filtering, convolution operations, and Fourier analysis, are better addressed in the frequency domain.

Wavelet transforms provide the most suitable tool for time-frequency analysis of non-stationary and transient signals. They can remove noise and reveal trends, similarities, repeated patterns and discontinuities, to ultimately outline the occurrence of certain events of interest. The wavelet transform, in contrast to Fourier analysis, consists of the decomposition of the original signal into scaled (stretched or compressed) and shifted versions of the original wavelet waveform, also known as the mother wavelet. The wavelet transform behaves as a frequency microscope that provides detailed information about different frequency bands as well as temporal information. Computationally efficient algorithms of the Discrete Wavelet Transform (DWT), based on the multi-resolution analysis concept, provide the decomposition of the original signal into low-frequency approximations and high-frequency detailed coefficients. Iterative decompositions of the resulting low-frequency approximations provide local detail in certain frequency bands in the time-frequency domain. The DWT decomposition is illustrated in the following example (Figure 1), in which an EEG signal that contains an epileptic seizure is analysed[1]. A fourth-order Daubechies with five levels of decomposition is shown. The approximation A5 and different levels of detail, from D1 to D5, show the frequency content of the different frequency bands of interest.

The Wavelet transform has been employed in several previous studies in the field of epilepsy analysis and is used for the extraction of features from EEG data. Table 2 compiles some of the references and the type

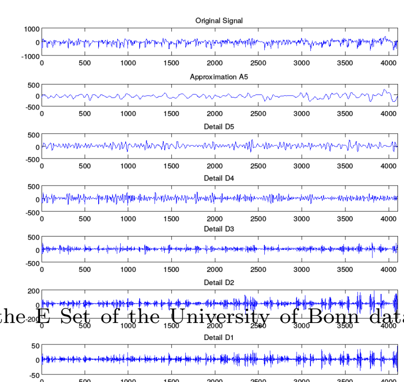


Figure 1: The 4th-order Daubechies 5-level decomposition of an EEG signal that contains an epileptic seizure

[1]This signal corresponds to one channel of the E Set of the University of Bonn dataset that records epileptic seizure activity.

Table 2: Wavelet transform implementations in previous studies

Reference	Wavelet	Decomposition levels
Omerhodzic 2010Omerhodzic et al. (2013)	4th order Daubechies	5
Janjarasjitt 2010Janjarasjitt (2010)	25th order Daubechies	5
Adeli 2003Adeli et al. (2003a)	4th order Daubechies Harmonic Wavelet	6
Fathima 2011Fathima et al. (2011)	2nd order Daubechies	4
Husain 2012Husain & Rao (2012)	4th order Daubechies	4
Ataee 2006Ataee et al. (2006)	4th order Daubechies	4−6
Orhan 2011Orhan et al. (2011)	2nd order Daubechies	6
Subasi 2007Subasi & Erelebi (2005)	4th order Daubechies	5

of Wavelet and number of decomposition levels employed. The fourth-order Daubechies Wavelet with 4 to 6 levels of decomposition is the most common choice found in the literature.

2.2. Bag of Words

The Bag-of-Words (BoW) model was originally proposed in the field of Natural Language Understanding Joachims (1997). However, this field is not the only field in which this technique has succeeded. In contrast, it has also been applied to the computer vision field, for image recognition, in which good performance rates have been achieved Cheng et al. (2010); Gilbert et al. (2009). Image recognition is not very different from the pattern recognition tasks that are required for seizure detection based on EEG signal analysis and, in fact, this technique has been explored for biomedical time series classification Wang et al. (2013). They are both digital signals in which the salient points of the signal serve to identify a sought-after pattern.

The working hypothesis of this study is, therefore, that with some adjustment, the same approach that is applied to Natural Language Understanding and Computer Vision can be applied to epilepsy seizure detection. The good results obtained in these fields of knowledge can also be reproduced in the field of EEG analysis for seizure detection. To prove this working hypothesis, a BoW-inspired system must be implemented and tested to determine whether the obtained accuracy rates improve on the state-of-the-art results.

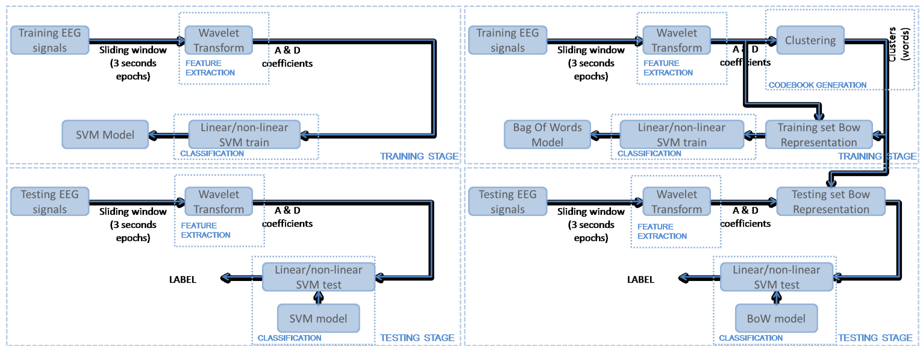
Essentially, the first step of the BoW model consists of calculating an attribute-value representation, in which each word that appears in the document has an associated value that reflects the number of times the word appeared in the text. In the context of EEG analysis, each word is considered to be a feature, and each document is represented by means of a feature vector. A document is therefore described by means of the word distribution, which is used to characterise the type of content of that document.

The required adjustments are intended to adapt the original approach, in which words are considered to be representation units, to the approach proposed here, in which EEG signal segments, or epochs, are equivalent to words in a document. Similar to the role that word order plays in documents, the epoch order can also be considered to be irrelevant and is therefore overlooked.

3. Methods

This section describes the characteristics of the EEG non-linear classifiers proposed here. Figure 2 outlines the stages that are involved in the process of signal characterisation and categorisation for both systems: an SVM classification framework and a BoW-inspired methodology that extends the previous pipeline. Both methods have most of their stages in common. The difference between the SVM method and the BoW-inspired one is that the codebook generation stage is omitted for the SVM. The classifier is therefore trained with the feature vector set computed after applying the wavelet transform decomposition.

From the seizure detection viewpoint, the process of codebook generation consists in identifying the different codewords appearing in the different EEG channels of a given record. Therefore, codewords are the different clusters in which the feature vectors characterising EEG channels can be grouped in. After



(a) using an SVM classifier method and (b) implementing a BoW-inspired method

Figure 2: Proposed EEG classification frameworks

having generated the codebook, the next step consists in obtaining the histogram that characterises the EEG signal channel. In order to do so, the proposal made here resorts to clustering the feature vector in the optimum number of clusters in which these data can be grouped in, and then, measuring the distance to each of the computed clusters. The next step consists in training the classifier using examples, with segments corresponding to normal activity and those others corresponding to epileptic activity. The adopted learning strategy uses a Support Vector Machine (SVM) classifier to compute the final classification model.

For both systems, different non-linear classifier kernels have been applied to compute their accuracy rates. The different kernels are also described in section 3.3.1. Several stages are common to both processes and both systems, as seen from Figure 2, such as the signal segmentation, the wavelet transform stages and the adopted learning strategy based on Support Vector Machine (SVM) classifiers. The stages represented in the figure are discussed in detail in the following subsections.

These different stages are grouped into two major processes: training and testing. First, a learned model is trained using examples of segments that correspond to both normal and seizure activity. This model is then used in the testing phase to classify a new, unseen signal.

In this framework, it is important to note that each individual signal channel is considered in isolation and is split into 3-second epochs, with a window overlap of one second between epochs (see Figure 3). The accuracy rate therefore refers to the number of epochs that can be correctly identified. This approach is the typical strategy used in the literature Fathima et al. (2011); Janjarasjitt (2010).

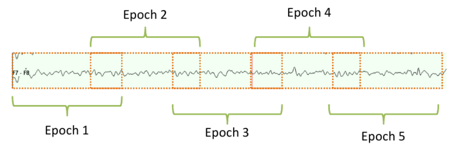


Figure 3: Signal segmentation into 3 second epochs

3.1. Feature extraction

Even though using the raw EEG signal channels as input for the classifier is possible, the use of these full segments is a poor representation of the input data. This drawback is due to the large amount of redundant information that is contained in an epoch and its high dimensionality, which make the learning and classification task more difficult. It is therefore necessary to find a better representation. Feature vector computation is the process of identifying the salient features of a signal segment and translating them into a quantitative set of features that characterise that segment. The process of computing these quantitative values is not unique; moreover, the performance and accuracy rate of the process can be greatly affected by the method by which these characterising values are selected and obtained.

This work proposes the use of a wavelet decomposition approach to minimise the amount of information that is required to characterise a segment as well as to magnify those signal aspects, or features, that are related to the presence of epileptiform activity.

Among the different wavelet transform types and decomposition level configurations, this work made use of the Daubechies wavelet Omerhodzic et al. (2013)]. The clinically and physiologically relevant activity of the brain is framed in the frequency range of 0.3 to 30 Hz. More specifically, brain activity can be categorised into a set of typical wave types, each of which lies within a

Table 3: Decomposition levels and frequency bands

Decomposed signal	Frequency bands	Decomposition Level
D1	43.4 - 86.8	1 (noises)
D2	21.7 - 43.4	2 (gama)
D3	10.8 - 21.7	3 (beta)
D4	5.40 - 10.8	4 (alpha)
D5	2.70 - 5.40	5 (theta)
A5	0.00 - 2.70	5 (delta)

predetermined frequency band.

The theoretical foundation for identifying those frequency bands out of the different decomposition levels is derived from Nyquist's theorem. The frequency bands of each decomposition level are comprised in the range stated by $[f_m/2 : f_m]$, such that $f_m = f_s/2^{l+1}$, where f_s is the sampling rate frequency and l is the level of decomposition Omerhodzic et al. (2013).

Given the dominant frequency components of the brain signal, the number of decomposition levels is set to five Adeli et al. (2003b). The Daubechies 4 (db4) wavelet transform is applied, decomposing the signal into details D1-D5 plus one final approximation A5, as listed in Table 3.

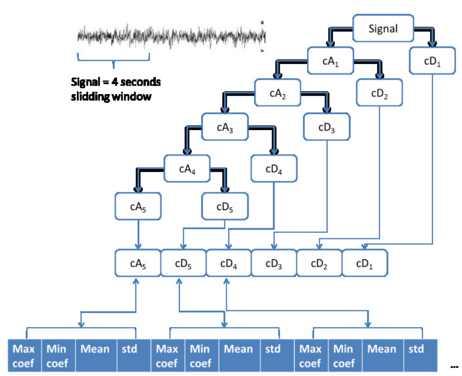


Figure 4: Signal decomposition using the DWT and feature extraction

However, the number of values that correspond to these coefficients is still too large for the purposes of a feature extraction process, which could be affected by the curse of dimensionality. For that reason, rather than using all of the coefficient values, the coefficient set dimensionality is reduced by selecting a small number of values that is believed to be the most characteristic set. Based on Kandaswamy et al. (2004); Gotman (1990), four statistical operations are performed over the original coefficient value set, and the following values are selected: the maximum value; the minimum value; the mean value; and the standard deviation value.

The complete feature extraction process is visually depicted in Figure 4. First, the Daubechies wavelet transform is used to decompose a given segment of the original signal into the 6 frequency subbands D1-D5,A5. Then, for each band, 4 statistical values are generated form all the coefficients comprised in each band, i.e. maximum, minimum, mean and standard deviation. The resulting feature vector $x_i \in \Re^{24}$ is composed of 24 values, with four statistical values for each of the six wavelet coefficient sets that correspond to the different decomposition bands.

3.2. Bag-of-Words feature representation

BoW is proposed in this paper as one of the main novelties in the EEG analysis field. BoW has shown its excellent properties in the fields of computer vision and text analysis to automatically learn and extract discriminative features in complex data, where manual feature selection or manually design features are not possible or provide little discriminative properties. This is largely the case of EEG where the interesting neural activity can be difficult to describe, may appear in many different varying shapes or may be largely hidden by noise. Features extracted from the EEG signals in the literature are largely based on simple statistics, being wavelet features one of the most advance techniques. In this sense, BoW can provide a relevant framework to the field to improve the current state of the art.

This subsection describes the processes of clustering and codebook generation that are involved in the BoW-inspired system. From the BoW perspective, EEG signals play the role of a text document in which each signal segment, quantised as a feature vector, can be characterised as a set of words in a specific configuration. The aim of this new feature representation is to better address the non-linear nature of the data by mapping to a new representation or space where the classifier can be better applied. The BoW representation can be therefore understood as a non-linear transformation function.

3.2.1. Codebook generation

The first step is the generation of words to be used to represent the initial signal. This process is called codebook generation and consists of identifying the most common and repetitive patterns, or *words*, that appear in a set of signals, or *document*. Thus, each word represents a frequent and characteristic spectro-temporal feature that can be used to codify our signal, to obtain the most representative groups. Under this definition, words are the different cluster centers c_k in which the feature vectors of the EEG segments can be grouped, and the codebook C, or *vocabulary*, is the entire set of words that can appear in the whole dataset.

Clustering

Clustering the feature vectors according to their common features allows us to obtain those representative words that repeat over the dataset. This clustering also removes undesired feature value variations due to noise in the signal because each group will allow a certain variability or deviation from the cluster center. At the same time, the outlier segments that are not very representative will be filtered because they will not have sufficient critical mass to compose their own cluster. This process can be considered equivalent to the elimination of the *typos* from the text.

Two different clustering techniques were tested in this paper, and an empirical comparison is presented in the results section. No assumption regarding the number of clusters, their allowed variability or the memberships of the feature vectors to the hypothetical words was made.

The first clustering approach used in this work is *k-means* clustering Kanungo et al. (2002). In this algorithm, initial seeds for each of the K clusters are initialised to a random sample in the dataset. Then, an iterative process is applied to refine their positions and characteristics until convergence is achieved. At each iteration, each sample, defined by its feature vector $x_i \forall i \in dataset$, is assigned to the closest cluster, and the cluster center c_k is recalculated as the average of all of the samples assigned to it.

$$c_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i \tag{1}$$

where n_k is the number of data samples that correspond to cluster k.

In contrast to previous work Gotman (1990), in our implementation, the number K of clusters is not predetermined beforehand but is calculated for each new training set under consideration. The implemented approach is intended to maximise the distances among the clusters, the *inter-class distance*, while minimising the distance between the elements that are inside a cluster, the *intra-class distance*:

$$Interclass(K) = \sum_{k=1}^{K} \sum_{i=1, i \neq k}^{K} \parallel c_i - c_k \parallel^2 \tag{2}$$

$$Intraclass(K) = \sum_{k=1}^{K} \frac{1}{n_k} \sum_{i=1}^{n_k} \parallel x_i - c_k \parallel^2 \tag{3}$$

To accomplish this goal, we predefine a maximum number of clusters, which ranges from 1 to 8 clusters, and we evaluate the optimisation function for each of the considered numbers of clusters:

$$\arg\max_x \left\{ \frac{Interclass(K) - Intraclass(K)}{max(Interclass(K), Intraclass(K)} \right\} \tag{4}$$

However, although k-means works well with isolated and compact clusters Jain et al. (1999), its performance decreases for a more complex clustering space. In addition, another disadvantage of the k-means algorithm is its stochastic initialisation, which results in a high sensitivity to the selection of the initial seed. As a result, the clustering can converge to a local minimum of the optimisation function if the initial partition is not properly chosen Jain et al. (1999).

To obtain a more robust grouping, a second clustering algorithm has been implemented and tested, based on the expectation−maximization (EM) algorithm McLachlan & Krishnan (2007). EM is an iterative methodology that allows finding the most likely estimates of parameters in statistical models. An EM iteration alternates between performing an expectation (E) step, which creates a function for the expectation of the log-likelihood as evaluated using the current estimate of the parameter k, and a maximisation (M) step, which recomputes the parameter k that maximises the expected log-likelihood found in the E step. This framework allows us to estimate and fit a Mixture of Gaussian (MoG) ϕ to our data x and calculate the associated parameters, the mean and covariance $c_k, \sum_k$, while minimising the error. The minimisation of the error is equivalent to maximising the probability of expressing our data as a function of the MoG.

$$p(x|\phi_k) \propto e^{-(x-c_k)^T \cdot \sum_k^{-1} \cdot (x-c_k)} \tag{5}$$

This equation 5 defines a probability that decreases exponentially with the Mahalanobis distance of a given data point x to a Gaussian ϕ_k, where

$$\phi_k = N(c_k, \sum_k) \tag{6}$$

being $N()$ a Gaussian or Normal distribution.

In our approach, the number of clusters k is automatically learned during the clustering process by applying the Figueiredo-Jain GMM automatic estimation Figueiredo & Jain (2002).

The strength of EM is that it can derive elliptical clusters (Gaussians) instead of spherical clusters that are estimated by k-means, and thus, it is more general and versatile when adapting to complex clustering spaces. Moreover, by integrating the automatic estimation of the number of clusters K in an iterative process, not only the computational cost is reduced by avoiding repetitions of the clustering process a number of times but also the sensitivity to the stochastic initialisation is removed.

The resulting vocabulary C will be the set of cluster centers that result from clustering the training set:

$$C = \{c_k\}_{k \in K} \tag{7}$$

in the case of k-means clustering, or by their centroids and their covariances:

$$C = \{c_k, \sum_k\}_{k \in K} \tag{8}$$

in the case of the EM algorithm.

3.2.2. Bag-of-Words representation

Once the vocabulary has been defined, the next step consists of redefining the feature vectors, which are originally composed of statistical values derived from the wavelet decomposition coefficients, as a function of our vocabulary. This codification or quantisation process generates a new descriptor, which is composed of words or *bag of words* and is finally fed into the classifier.

The new chosen descriptors represent the feature vector in terms of its distance to each of the words or cluster centroids. Since clusters are characterised differently depending on the applied clustering technique, two different distances were used: Euclidean distance for k-means clusters and Mahalanobis distance, for the EM clusters.

After the generation of the descriptor, the aforementioned feature vector of 24 values is now reduced to a new vector whose dimensionality depends on the optimum number of clusters for that specific signal. This arrangement can be seen as a non-linear transformation of the data.

3.3. Classification

Finally, the chosen feature representation, either the statistical values that result from wavelet decomposition or the BoW representation, are fed into a classifier that distinguishes among the different classes of samples. In our seizure detection problem, this classification is performed to distinguish normal and seizure EEG signals, and thus, a binary classifier is used (normal/epileptic). In our implementation, an SVM paradigm has been used Vapnik (1995); Janjarasjitt (2010); Kıymık et al. (2005). The choice of SVM in comparison with more traditional approaches, such as regression, neural networks and discriminant analysis (DA) Ripley & Hjort (1995), is supported by the reported advantages of the SVM Auria & Moro (2008): it does not require regularity in the data and thus can be applied to data that follow an unknown distribution; it delivers a unique solution because the optimality problem is convex in contrast to neural networks; it can be easily extended to non-linear non-parametric problems by replacing the linear kernel; it scales relatively well to high-dimensional data; and the trade-off between the classifier complexity and error can be controlled explicitly.

To classify a new test descriptor, the SVM should be already trained in a supervised mode with a training set that is composed of both positive and negative examples of normal and epileptic EEG. As an output of the training phase, an hyperplane that is capable of separating the two classes with the maximum margin, called the maximum-margin hyperplane, is obtained. The position of a new test descriptor with regard to this hyperplane will be the criterion for assigning it an identity as normal or epileptic.

3.3.1. Non-linear classification

For linear data, a hyperplane can be used to split the data. However, the assumption of linearity is often wrong (see Figure 5). In these cases, the dataset is inseparable in a linear space, and the classification fails. Although the decision of taking a linear classifier is supported by the literature Gotman (1990); Orhan et al. (2011), where little attention has been paid to the classification technique to be applied, and linear classifiers have reported excellent results in EEG analysis, our working hypothesis about the non-linearity of the data will be evaluated by proposing the usage of non-linear SVM.

An extension of SVM was developed Husain & Rao (2012) to solve non-linear problems by the "*kernel trick*". Given a training set $F = \{f\{x_i\}, y(x_i) \in \{-1, 1\}\}_{i \in dataset}$, where f is the BoW descriptor that corresponds to the training sample i and y is its class, this methodology applies a kernel function K to the descriptors, which maps them into a higher dimensional non-linear space by means of a non-linear function φ.

$$f(x), f(x') \leftarrow \mathsf{K}\left(f(x), f(x') = \varphi\left(f(x)\right), \varphi\left(f(x')\right)\right) \tag{9}$$

In this new space, the data are linearly separable, and the SVM framework can be applied. This process is illustrated in Figure 5.

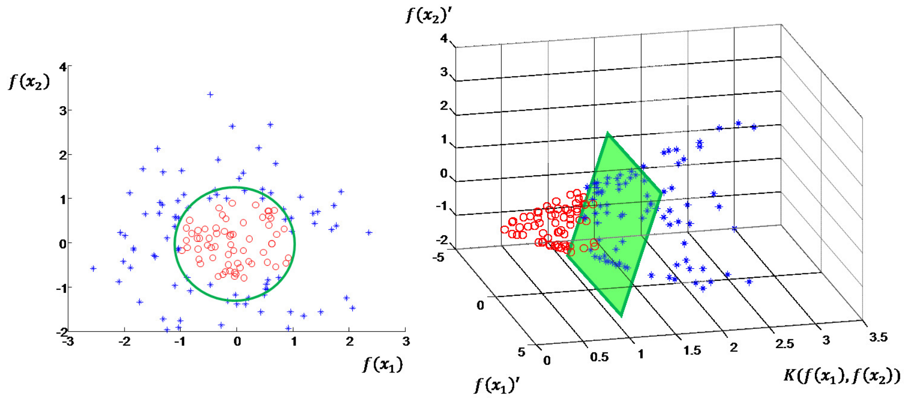


Figure 5: Non-linearly separable data (left) and its mapping into a linearly separable space through a non-linear kernel (right)

Different kernel functions can be applied to obtain the best possible transformation, and even a function that is personalised to the data can be used. Among the most common transformations are Cristianini & Shawe-Taylor (2000).

Polynomial of order d:

$$\mathsf{K}\left(f(x), f(x')\right) = \left(f(x) \cdot f(x')\right)^d \tag{10}$$

$$\mathsf{K}\left(f(x), f(x')\right) = \left(f(x) \cdot f(x') + 1\right)^d \tag{11}$$

Gaussian Radial Basis Functions (RBF):

$$\mathsf{K}\left(f(x), f(x')\right) = exp\left(\frac{-\|f(x) - f(x')\|^2}{2\sigma^2}\right) \tag{12}$$

Perceptron multi-layer:

$$\mathsf{K}\left(f(x), f(x')\right) = tanh\left(\tau \cdot f(x) \cdot f(x') + c\right) \tag{13}$$

4. Results

This section describes the results that were obtained from testing the proposed system in three different datasets, which encompass situations with artifacts, different noise levels, highly attenuated signals and different activity variations. This section is intended to evaluate both types of systems under different circumstances, to determine the system that better suits the characteristics of a real scenario, and to compare it against the state of the art.

The proposed system has been implemented in Matlab, using the gmmbayestb-v1.0[2] for automatically learning the number of clusters based on the Figueiredo-Jain GMM automatic estimation Figueiredo & Jain (2002). Additionally, we have employed the Matlab support for the SVM classifier and its different kernels.

4.1. EEG Data

Different datasets have been used in this work for training and testing purposes. First, the system was trained using the data described in Andrzejak et al. (2001), which is an open-access dataset made available by the University of Bonn. This dataset comprises a series of *clean* EEG signal channels that were recorded from both healthy and epileptic patients during ictal and inter-ictal periods. It is organised into five different sets, which are labeled from A to E. The A set records eyes opened and healthy patient activity, and the B set records the activity with eyes closed and healthy patients; the C set records inter-ictal activity from the healthy part of the brain, the D set records also inter-ictal activity but from the epileptic hemisphere of the brain, and finally, the E set records epileptic seizure activity. This work concentrates on sets A and E for

[2] `http://www.it.lut.fi/project/gmmbayes/doc/gmmbayestb-v1.0/gmmbayestb-v1.0/`

learning by example, following a similar procedure as in other approaches in the literature Janjarasjitt (2010); Fathima et al. (2011). Each set contains 100 individual channels of 23.6 seconds, at a sample rate of 173.6 Hz. This dataset is an artifact-free dataset, which was recorded with a 128-channel amplifier system. Each dataset records the activity of 5 different patients. In total, sets A and E contain a total of 1200 segment each, considering here 3-second segments. Half of these segments, randomly chosen, has been used for the codebook generation as well as for the SVM classifier model training.

Initially, the same University of Bonn dataset was used for testing using the remaining 50%. In addition, to obtain accurate results that are closer to a real scenario, where the system cannot be retrained for each new environment or patient, two other datasets are used as cross-dataset evaluation. The testing stage is extended to additional datasets that were not considered during the training stage or adapted to them. This allows us to obtain a more reliable evaluation of the real performance of the method. Moreover, it is expected that these datasets contain different variations, noise and artifacts from the ones used for training.

This work therefore resorts to a second only-testing dataset, which was made available by the Epilepsy Center of the University Hospital of Freiburg, Germany Winterhalder et al. (2003). This dataset records data from 21 patients who suffer from medically intractable focal epilepsy. The data are labeled according to the type of activity they record, which conforms to our set of labels I and J. Moreover, each labeled activity is stored in a single file in which the signal channels are differentiated. For each patient, this dataset provides records of 2-5 hours of ictal activity, sampled at a frequency rate of 256 Hz. We have employed a reduced version of the dataset that records 3693,2 seconds of ictal activity. It should be highlighted that this dataset contains artifact-free data that were recorded from intracranial sensors.

Finally, a third dataset is also used for testing exclusively. The purpose of this third only-testing dataset is to consider real EEG data that was recorded from non-ideal environments in which there were numerous artifacts and attenuated signals. Such factors are missing in the two previous datasets. This last source of EEG data used in this work comes from the Hospital Regional Universitario Carlos Haya (HRUCH) Malaga, Spain. EEG data were recorded with XLTEK Neuroworks at a sampling rate of 512 Hz, although the signals are band-pass filtered in the range of 2 to 200 Hz. This dataset is comprised of four sets, labeled F, G, H, and K (to continue with the University of Bonn nomenclature). The F set records the activity of a healthy patient, although with many artifacts (due to cable disturbances and blinking). The G set records inter-ictal activity, also with many blinking artifacts. These two sets are sampled at a frequency of 511.99 Hz. The H set records a partial seizure, recording from both the healthy and the epileptic part of the brain. The seizure takes place at the left temporal lobe of the brain. The data were sampled at a frequency of 200 Hz. The K set records a tonic-clonic general seizure, also downsampled at 200 Hz. These three sets sum to a total of 277,71 seconds of recording.

Because of the many artifacts and attenuated signals, the HRUCH dataset can be considered to be the most complex dataset, and it can provide a clear idea of how good is the performance of the proposed system outside of the lab, in a real environment. The data, as provided by the HRUCH dataset, are the type of data that a framework for seizure detection will be required to address. Table 4 summarises the most relevant features of the sets used for testing purposes.

In summary, out of the 21728 segments of 3 seconds used in this work, 600 have been used for training purposes (half of the A and E sets). For those used for testing, 10374 segments correspond to normal activity and 10154 segments to epileptic activity. The system performance has been tested with a total amount of 4065,31 seconds, 215,81 seconds of normal activity (label 1) and 3849,5 of seizure activity (label 2). No additional filters have been applied to any of the datasets, apart from the anti-aliasing filter applied by the equipment used to record the University of Bonn and HRUCH datasets. To address the different frequency rates that range from 173.6 to 511.99 Hz, all of the datasets are automatically resampled at 173.6 Hz, the frequency of the training set.

Table 4: Datasets used in this work

Set label	Source	Number of segments	Time in seconds	Label
A	University of Bonn	600	11,8	1
B	University of Bonn	1200	23,6	1
C	University of Bonn	1200	23,6	1
D	University of Bonn	1200	23,6	1
E	University of Bonn	600	11,8	2
F	HRUCH	345	7,61	1
G	HRUCH	5829	125,60	1
H	HRUCH	894	39,5	2
I	Freiburg	4431	3600	2
J	Freiburg	1920	93,2	2
K	HRUCH	2309	105	2

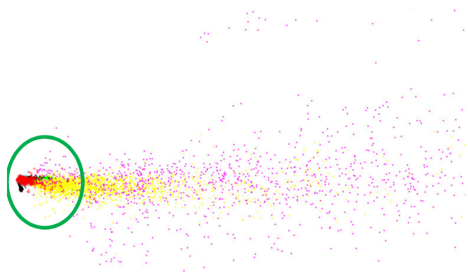
4.2. Qualitative analysis of the data

PCA projections of the data are computed and represented in Figure 6 to graphically demonstrate that the data from different datasets (i.e., the different modalities present in the data, in different colours) are not linearly separable. This can be noticed in Figure 6b, where the data and modalities seem hard to separate with linear and simple classifiers in this original space, while non-linear separation using kernels may give better results.

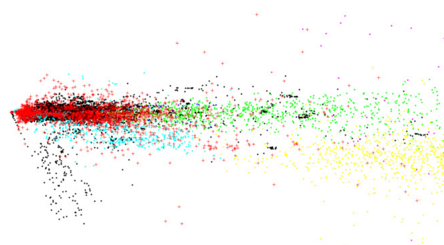
By displaying the same PCA representation of the data space but this time after BoW has been applied (see Figure 6d), a more linear space can be observed where modalities are less mixed. Therefore, it can be inferred that BoW helps to linearise the space, simplifying the separation process performed by the classifier.

Similar conclusion can be extracted by comparing Figures 7a and 7b, which show the same projections but now differentiating the healthy and epileptic samples by class rather than every single modality.

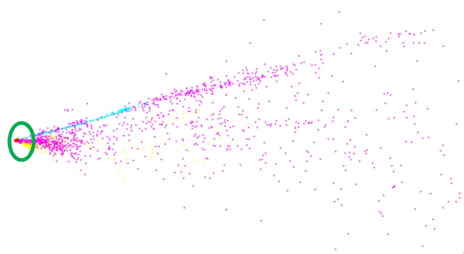
Please note that, although these figures are only indicative because no PCA is actually performed by the classifier, it gives an indication of the distribution of both the input feature space and the BoW transformed feature space. Only two dimensions are represented to facilitate human visual interpretation of the data.



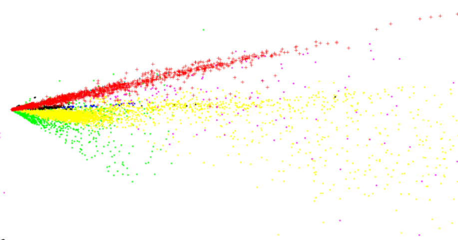
(a) PCA projection of all of the training and testing samples for the wavelet feature representation



(b) Zoomed area in the green circle



(c) PCA projection of all of the training and testing samples for the BoW feature representation

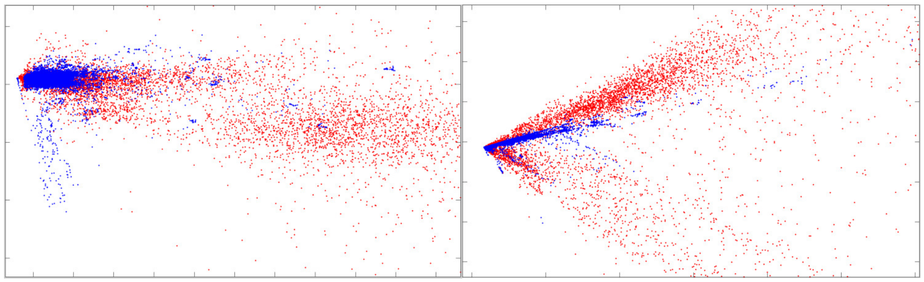


(d) Zoomed area in the green circle

Figure 6: PCA projection of all of the training and testing samples into the 2 most significant PCA dimensions for the (a,b) wavelet feature and (c,d) BoW feature representation. The right column shows a zoom of the area in the green circle. Colour-dataset correspondence legend: red dot=A, green=E, blue=F, black=G, cyan=H, yellow=I, magenta=J, red cross=K

4.3. Clustering evaluation

The first implementation decision to be evaluated is the clustering methodology to be applied to generate the codebook. Both k-means and EM were evaluated on the 3 testing sets (Bonn, Freiburg and HRUCH). The empirical



(a) (b)

Figure 7: PCA projection of all of the training and testing samples (zoom versions in Figure 6) into the 2 most significant PCA dimensions for the a) wavelet feature and b) BoW feature representation. Blue indicates healthy samples, while red indicates seizure samples

results confirm clearly the theoretical advantage of using EM instead of k-means (see Figure 8).

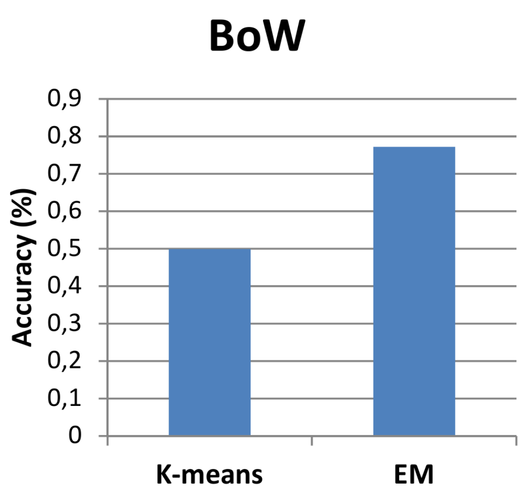


Figure 8: Average accuracy rate obtained by BoW over the whole testing set (A to K) by using K-means and EM clustering

Since EM relies on a stochastic process to initialise the clustering process, an experiment was performed to evaluate the impact that the selected initial cluster might have in the overall performance of the system. Quantitative experiments in section 4.5 were repeated 10 times for the BoW approach (SVM-only appro-

Table 5: Standard deviation for accuracy results after 10 iterations

STD for BoW		
	A SET	E SET
No D1 band	0,0002635	0,00932274
Quadratic kernel	0	0,00811947
Polynomial kernel	0,00035136	0,00403399
RBF kernel	0,00035136	0,00901815
Perceptron kernel	0,0002635	0,01237935

aches do not use the clustering) for sets A and E and the standard deviation (STD) between experiments was measured (see Table 5). The average standard deviation (STD) is 0.0044 for BoW, which represents an almost negligible influence of this initialisation on the final performance of the system.

4.4. Window size evaluation

Another parameter that must be verified is the size of the signal segment or epochs into which the EEG channels have been split. The 3-seconds window size has been empirically demonstrated by analyzing the performance of the system under different window sizes. The following graphics summarise the variations in the accuracy rate, which were experienced by varying the size of the sliding window, from two-second windows to five-second windows. For the purpose of conciseness, only the SVM and BoW implementation of the RBF kernel is presented in this paper. Because it will be justified later, on average, the RBF kernel provides better results than any of the other implementations, and for that reason, only its value is represented here.

Figure 9 summarises the accuracy rates that are obtained by systems testing the different sets considered here. Although the results are not totally conclusive, it can be observed that the maximums are normally achieved in 3 or 4 seconds. This timing is especially notable for sets A and E, where a maximum accuracy rate of 100% is achieved.

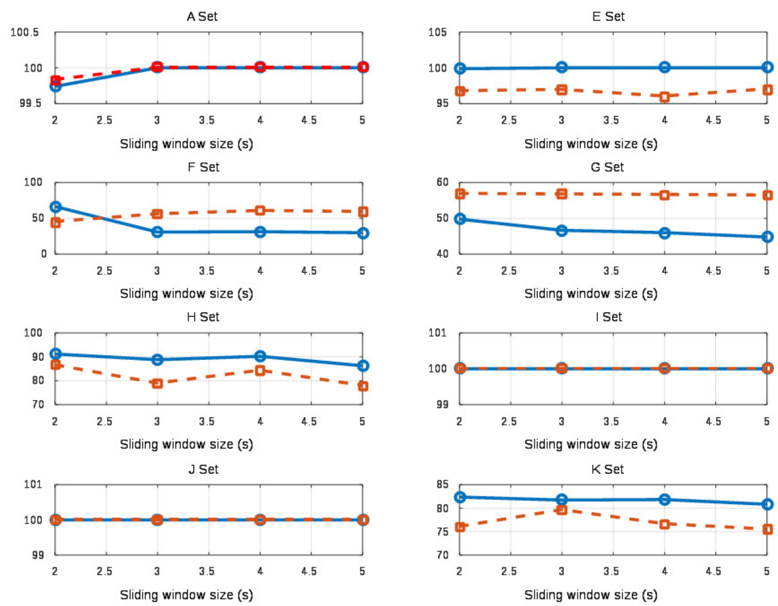


Figure 9: Classification accuracy evolution with the sliding window size, for tested size values of 2, 3, 4 and 5 seconds. The solid blue line represents the SVM-only system results, and the dashed green line represents the BoW+SVM system results

Table 6: Accuracy results for the different experiments that were conducted

System	BoW(%)	BoW (%)	BoW (%)	BoW (%)	BoW(%)	SVM(%)	SVM (%)	SVM (%)	SVM (%)	SVM (%)
Set \ Kernel	Linear (dot product)	Quadratic	Polynomial order 3	RBF	Perceptron	Linear (dot product)	Quadratic	Polynomial order 3	RBF	Perceptron
A	**100,00%**	100,00%	100,00%	100,00%	82,00%	100,00%	100,00%	100,00%	**100,00%**	82,00%
B	93,00%	93,00%	89,00%	93,00%	**98,00%**	79,00%	78,00%	78,00%	77,00%	**98,00%**
C	**99,00%**	98,00%	98,00%	98,00%	71,00%	90,00%	91,00%	92,00%	92,00%	71,00%
D	**86,00%**	79,00%	84,00%	86,00%	75,00%	78,00%	50,00%	80,00%	80,00%	75,00%
E	95,00%	96,00%	100,00%	97,00%	79,00%	100,00%	100,00%	100,00%	**100,00%**	79,00%
F	70,00%	73,00%	77,00%	73,00%	**78,00%**	75,00%	73,00%	72,00%	72,00%	**78,00%**
G	**93,00%**	59,00%	66,00%	58,00%	82,00%	22,00%	30,00%	36,00%	44,00%	82,00%
H	84,00%	80,00%	27,00%	84,00%	24,00%	46,00%	31,00%	42,00%	**87,00%**	24,00%
I	**100,00%**	93,00%	29,00%	100,00%	100,00%	100,00%	88,00%	87,00%	**100,00%**	88,00%
J	99,00%	91,00%	33,00%	100,00%	100,00%	98,00%	47,00%	99,00%	**100,00%**	100,00%
K	73,00%	77,00%	26,00%	**81,00%**	31,00%	41,00%	41,00%	19,00%	80,00%	31,00%
Mean	**90,21%**	85,35%	66,26%	88,15%	74,57%	75,35%	66,10%	73,12%	**84,72%**	73,46%

4.5. Quantitative results

An exhaustive evaluation of each of the proposed methods -with and without BoW, with and without different non-linear kernels- was performed. The accuracies of the classification results obtained for different experiments are depicted in Table 6. Since the models are trained using half of the A and E datasets, first and fifth lines in the table can be considered intra-dataset while all others are cross-dataset experiments. Parameters are kept the same for all datasets.

5. Discussion

Several conclusions can be observed from Table 6. First, we can see how BoW drastically improves the accuracy of the system, on average and for each set, with regard to the equivalent model of SVM. This improvement is because the BoW strategy creates a more discriminative space in which the classification can be performed, while focusing on the key features. This arrangement is shown in Figure 6, where all of the positive and negative samples of all (A to K) datasets are projected into the 2 most significant dimensions of a PCA space.

Second, a similar accuracy to BoW can be obtained with a more conventional approach and a careful selection of non-linear classifiers in the same feature space. This approach gives lower, but similar, accuracy on average and provides some of the best possible accuracies on the individual sets (E, J, H). The good performance of the RBF kernels, especially in the SVM version, is supported theoretically because of the fact that if the kernel used is a Gaussian RBF, then the resulting feature space is a Hilbert space that has an infinite dimension. In this space, our maximum margin classifiers are well regularised and large or even infinite dimensions do not spoil the results, which mitigates the curse of dimensionality. However, it is important to note that there is a drop in SVM-RBF performance with respect to some of the noisiest datasets (G). This drop could suggest the convenience of the BoW approach for addressing the (considerably) more difficult conditions.

Moreover, the use of non-linear classifiers for BoW methods is revealed to be unnecessary because similar results are obtained for all of the possible kernels, except for the $3rd$-order polynomial approach, in which overfitting to the training seems to have happened. This redundancy occurs because BoW has already reduced significantly the dimensionality and non-linearity of the feature space as shown in Figure 7b and makes the use of non-linear kernels in the classifier redundant. This is a significant advantage since selecting a suitable kernel is not trivial and relies largely on empirical tuning as shown in Burges (1998) and in our own exhaustive experiments.

In the overall and considering all datasets and the cross-dataset setup, comprising different variations and artifacts, it can be observed how the BoW method combined with an SVM classifier with a linear kernel yields a mean accuracy of 90,21%. BoW implementation provides the best results on average when compared to the equivalent linear or non-linear SVM implementation. This is justified by the success of BoW in creating a more discriminative and linear space in which the classification of the EEG data can be better performed. The use of BoW also avoids the non-trivial selection of a kernel and parameter tuning that is required in the SVM classifier.

Table 7: Accuracy means for A and E sets

System	BoW(%)	BoW (%)	BoW (%)	BoW (%)	BoW(%)	SVM(%)	SVM (%)	SVM (%)	SVM (%)	SVM (%)
Set \ Kernel	Linear (dot product)	Quadratic	Polynomial order 3	RBF	Perceptron	Linear (dot product)	Quadratic	Polynomial order 3	RBF	Perceptron
Mean A and E sets	97,59%	98,09%	**99,84%**	98,38%	80,60%	99,84%	**100,00%**	**100,00%**	**100,00%**	80,60%

5.1. Comparison with the state of the art

Two different comparisons with state-of-art methods were performed. In the first comparison, results reported by state-of-art methods on the public A and E dataset are compared against or best performing methods. These dataset are the most widely used in the literature and for that reason they are commonly used as reference framework Chen (2014). Our best performing methods were chosen by selecting a BoW and a non-linear SVM systems from Table 7, which summarises the accuracy rates obtained for the different techniques when applied only to sets A and E.

Table 8 compares the accuracy results that are obtained from state-of-the-art methods with the ones in Table 7 for sets A and E (normal and ictal activity) according to the reported results in their corresponding papers. Different approaches are implemented by the studies listed in this table, such as Neural Networks, Wavelet analysis, or a different implementation of the BoW model Wang et al. (2012). From the observed data, it can be concluded that the current implementation using both systems, BoW and SVM, achieves state-of-the-art performance.

Although these results are important as a reference against the state of the art, there are some limitations in this comparison. First, different authors used different experimental setups and training/testing splits, which makes those numbers not directly comparable. In this regard, in order to measure the impact that both the training/testing split and the particular subset selected for training may have in the final results, an additional experiment was carried out, in which the training and testing configuration is modified. A leaving-10%-out cross validation approach was implemented in which the experiment was run

Table 8: Comparative analysis with previous work using the University of Bonn Dataset (CV: cross validation)

Reference	Accuracy (%)	Training/testing Setup
Polat and Gulness Polat & Güneş (2007)	98.72	5 and 10 fold CV
Guo et al. Guo et al. (2009)	99.6	50-50
Wang et al. Wang et al. (2012)	99.5	10 fold CV
Janjarasjitt et al. Janjarasjitt (2010)	97.6	66-33
Husain et al. Husain & Rao (2012)	98.2	60-40
Fathima et al. Fathima et al. (2011)	99.8	66-33
Chen et al. Chen (2014)	100	50-50
Übeyli Übeyli (2010)	99.56	50-50
This work: BoW + SVM Non-linear SVM (RBF kernel)	99.85 100	50-50

10 times. With the obtained accuracy, the STD was calculated, obtaining a mean value of 7.3483e-04 (±0.07%) for the BoW and 2.2631e-04 (±0.02%) for the SVM. This variation may imply a crucial difference between being the best performing method or not.

As a second limitation, this single dataset only contains a type of variation in the epileptic activity, which implies a simple problem and explains the high accuracy rates obtained. Finally, since authors only report results on these datasets or retrain for each dataset, there is a risk of overfitting which means that the reported result may be artificially high and not a true reflection of the real performance. Very little discussion, if any, is provided on those papers regarding the required tuning of these methods and their parameters to reach those results. This issue also arises in the adjustment of our proposed methods, since by selecting the best methods for A and E datasets in Table 7 we are not necessarily taking the best overall method 6, and their performance will drop when evaluated under more challenging conditions: BoW+Polynomial kernel drops from 99.84% to 66.26% and the SVM-RBF drops for 100% to 84.72%.

In order to provide a better and more reliable comparison, a new set of experiments were performed in which some representative and up-to-date works of the state of the art have been implemented and evaluated in the same experimental setup, including cross-dataset evaluation, and in more complex and realistic datasets with the presence of noise and artifacts. For instance, our datasets consists of intracranial data that is characterised by high-amplitude signals with low noise, and a different dataset was obtained from a real scenario, in which the EEG signals that were recorded from the scalps were considerably attenuated. No other state-of-the-art methods have been tested under those conditions, and few have followed our approach of training with a completely different dataset from the testing one. This testing strategy proves the robustness of our methods against different patients, capturing device-related and environmental changes.

Evaluation was performed focusing on cross-dataset experiments, in which the system is trained on the standard public set and evaluated in the other without adaptation or tuning. Thus, training was performed using half of the segments in A and E datasets, whereas testing was carried out on all other segments and datasets. The main reason is to demonstrate that accurate results were not dependent on which dataset was used to train the system and the reported results are a better reflection of the performance in realistic scenarios.

We have selected four of the most representative works of the state of the art. The method "DTCWT+SVM" Chen (2014) proposed the use of a novel approach based on the use of a dual-tree complex wavelets (DTCWT) combined with an SVM classifier. The method labelled PE+SVM Li et al. (2014) employs permutation entropy and an SVM classifier to explore changes in the EEG. The method labelled as DWT+KNN Guo et al. (2011) applies genetic programming to a reduced dimension feature vector obtained after a discrete wavelet transform (DWT) with the purpose of improving the discriminative performance of K-nearest neighbour (KNN) classifier. Finally, the method labelled as DWT+ANN Tzallas et al. (2007, 2009) proposes the use of time-frequency

Table 9: Comparative analysis with previous work using the complete dataset space employed in this work

	A-E sets (University of Bonn dataset)				FGHK sets (HRUCH dataset)				IJ sets (Freiburg dataset)			
Method \ Set	Acc	Prec	Recall	F1	Acc	Prec	Recall	F1	Acc	Prec	Recall	F1
BoW+SVM	**94,68%**	**70,06%**	94,00%	**80,29%**	**81,64%**	79,24%	**74,49%**	**76,79%**	99,12%	**100,00%**	99,38%	99,69%
DTCWT+SVM Chen (2014)	83,60%	57,26%	71,00%	63,39%	60,37%	**100,00%**	20,78%	34,41%	42,14%	**100,00%**	41,18%	58,33%
PE+SVM Li et al. (2014)	70,93%	11,36%	14,33%	12,68%	40,26%	45,81%	64,99%	53,74%	8,41%	**100,00%**	6,42%	12,07%
DWT+KNN Guo et al. (2011)	91,00%	52,91%	**100,00%**	69,20%	68,02%	82,89%	52,40%	64,21%	**100,00%**	**100,00%**	**100,00%**	**100,00%**
DWT+ANN Tzallas et al. (2007)	90,55%	51,68%	**100,00%**	68,14%	72,34%	96,59%	44,23%	60,68%	98,65%	**100,00%**	99,17%	99,58%

distributions with an artificial neural network (ANN) classifier. Since no implementation was provided by the authors, we implemented them keeping their parameter setting and configuration as closed as possible to their specifications. Whenever there were missing details regarding the implementation, the configuration details of our system were adopted to ensure a fair comparison. Similarly, the experimental setup and training/testing split used is identical for all compared methods.

Table 9 presents the obtained results for all the datasets used in this work, including our private datasets. Accuracy (Acc), precision (Prec), recall and F-measure (F1) are used as evaluation metrics. Table 10 summarises the mean F1-measure obtained for each of the evaluated methods across the three different datasets.

From the obtained results, it can be concluded that our method outperforms the other methods in almost all cases, with the exception of the Freiburg dataset, in which the results obtained are very similar to the best result reported in literature. However, the excellent performance of all methods when using the Freiburg dataset, which only contains intracrannial ictal (positive) segments, may indicate the simplicity of this set and/or the particularities of intracrannial ictal cases. By considering all dataset together (see Table 10), we can conclude that our method is more reliable and robust since it does not depend on the characteristics of the signal to be classified (intracrannial or scalp, noisy or noise-free and with artifacts or artifact-free).

5.2. Computational efficiency

Table 10: Mean F1-measure for the evaluated methods

Method	Mean F1-measure	Standard Deviation
BoW+SVM (Ours)	85,59 %	12,33 %
DTCWT+SVM Chen (2014)	52,04%	15,47 %
PE+SVM Li et al. (2014)	26%	23,88 %
DWT+KNN Guo et al. (2011)	77,80%	19,38 %
DWT+ANN Tzallas et al. (2007)	76,13%	20,64 %

As additional advantage regarding the computational cost (see Figure 10), BoW reduces the feature vector dimension. This reduction makes easier the convergence of the classifier, which results in a lower training computational cost, having a more noticeable effect when the number of available samples increases. Both proposed systems have equivalent testing costs.

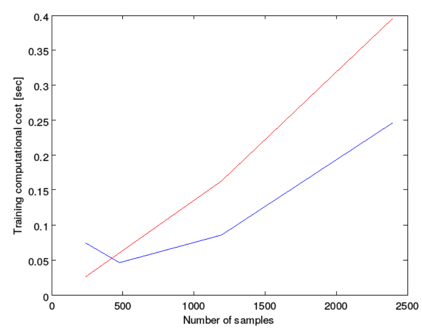


Figure 10: Training time of BoW+LinearSVM (blue) and SVM-RBF (red) systems and their trend with the number of training samples

6. Conclusions

This work presents two systems for the automatic analysis of EEG recordings, which aim toward epilepsy seizure detection. As the first contribution, one of the proposed systems consists of a non-linear implementation of a Support Vector Machine (SVM) classifier that makes use of well-known techniques and non-linear kernels to optimise the feature extraction and learned classification model. The second contribution is inspired by a successful model for Natural Language Understanding and Computer Vision, known as Bag-of-Words (BoW), which is adapted to the field and added into the framework.

The proposed systems were validated in a wide spectrum of data, including public standard datasets and complex private datasets, in which different type of activities, noise and artifacts appear. Our proposed system performs at state-of-art level when evaluated in standard datasets under an intra-dataset setup. More importantly, cross-dataset experiments have been used to evaluate the performance of the proposed BoW approach against some of the most relevant and representative state-of-the-art methods for all the datasets considered in this work.

The main advantage of the proposed solution consists in its robustness to real-environmental conditions, as demonstrated by the performance results. In terms of computational cost and, without having carried out any optimisation works, we can also highlight as an additional advantage the reduction of the feature vector dimension carried out by the BoW. This reduction makes easier the convergence of the classifier, which results in a lower training computational cost, having a more noticeable effect when the number of available samples increases.

The results prove, as main advantage, the robustness of our method to real-environmental conditions without having carried out any optimisation works. The robust performance of the BoW implementation when facing these types of realistic EEG signals suggests the suitability of this model for deployment in real hospital environments to reduce the bottleneck of EEG analysis in epilepsy diagnosis stage.

As future work we aim to evaluate our system for automatic analysis of long EEG test, such as those of sleep deprivation, which currently relies on specialists supervising the testing results. This will require the introduction of time-series modeling in the BoW representation since our current implementation of BoW fails to represent the underlying temporal and causal information that is inherent to time series such as EEG signals and that may be needed to detect more complex and subtle neural activity. Additionally, due to the computational efficiency of the proposed method, we will work on the implementation of a hardware-specific version of the algorithm for FPGAs.

Acknowledgment

This work has been partly funded by the Spanish Ministry of Economy and Competitiveness under project REBECCA (TEC2014-58036-C4-1-R), by the Regional Government of Castilla-La Mancha under project SAND (PEII_2014_046_P) and by the University of Castilla-La Mancha R&D Plan under the access contracts to the Spanish system of science, technology and innovation call, which is partially funded by the European Social Fund (31/07/2014 Resolution, published in the DOCM on the 25 of August 2014).

References

Acharya, U., Molinari, F., Vinitha Sree, S., Chattopadhyay, S., Kwan-Hoong, N., & Suri, J. (2012a). Automated diagnosis of epileptic eeg using entropies. *Biomed. Signal Process Control*, *7*, 401–408.

Acharya, U., Sree, S., Ang, P., Yanti, R., & Jasjit, S. (2012b). Application of non-linear and wavelet based features for the automated identification of epileptic eeg signals. *International journal of neural systems*, *22*.

Acharya, U. R., Fujita, H., Sudarshan, V. K., Bhat, S., & Koh, J. E. (2015). Application of entropies for automated diagnosis of epilepsy using EEG signals: A review. *Knowledge-Based Systems*, *88*, 85 – 96.

Acharya, U. R., Sree, S. V., & Suri, J. S. (2011a). Automatic detection of epileptic eeg signals using higher order cumulant features. *International Journal of Neural Systems*, *21*, 403–414.

Acharya, U. R., Vinitha Sree, S., Chattopadhyay, S., & Yu, A. P. C., W.and Alvin (2011b). Application of recurrence quantification analysis for the automated identification of epileptic eeg signals. *International Journal of Neural Systems*, *21*, 199–211.

Adeli, H., Zhou, Z., & Dadmehr, N. (2003a). Analysis of EEG records in an epileptic patient using wavelet transform. *Journal of Neuroscience Methods*, *123*, 69 – 87.

Adeli, H., Zhou, Z., & Dadmehr, N. (2003b). Analysis of eeg records in an epileptic patient using wavelet transform. *Journal of neuroscience methods*, *123*, 69–87.

Alfaro-Ponce, M., Argelles, A., & Chairez, I. (2016). Pattern recognition for electroencephalographic signals based on continuous neural networks. *Neural Networks*, *79*, 88 – 96.

Alotaiby, T. N., Alshebeili, S. A., Alshawi, T., Ahmad, I., & Abd El-Samie, F. E. (2014). Eeg seizure detection and prediction algorithms: a survey. *EURASIP Journal on Advances in Signal Processing*, *2014*, 1–21.

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Physical Review E*, *64*, 061907.

Ataee, P., Avanaki, A. N., Shariatpanahi, H. F., & Khoee, S. M. (2006). Ranking features of wavelet-decomposed eeg based on significance in epileptic seizure prediction. In *Signal Processing Conference, 2006 14th European* (pp. 1–4).

Atyabi, A., Shic, F., & Naples, A. (2016). Mixture of autoregressive modeling orders and its implication on single trial EEG classification. *Expert Systems with Applications*, *65*, 164 – 180.

Auria, L., & Moro, R. A. (2008). *Support Vector Machines (SVM) as a Technique for Solvency Analysis*. Discussion Papers of DIW Berlin 811 DIW Berlin, German Institute for Economic Research.

Bao, F. S., Lie, D. Y.-C., & Zhang, Y. (2008). A New Approach to Automated Epileptic Diagnosis Using EEG and Probabilistic Neural Network. In *Proceedings of the 2008 20th IEEE International Conference on Tools with Artificial*

Intelligence - Volume 02 ICTAI '08 (pp. 482–486). Washington, DC, USA: IEEE Computer Society.

Bhardwaj, A., Tiwari, A., Krishna, R., & Varma, V. (2016). A novel genetic programming approach for epileptic seizure detection. *Computer Methods and Programs in Biomedicine*, *124*, 2 – 18.

Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.*, *2*, 121–167.

Chen, G. (2014). Automatic eeg seizure detection using dual-tree complex wavelet-fourier features. *Expert Syst. Appl.*, *41*, 2391–2394.

Cheng, Z., Qin, L., Huang, Q., Jiang, S., & Tian, Q. (2010). In *Pattern Recognition (ICPR), 2010 20th International Conference on* (pp. 3228–3231).

Chua, K. C., Chandran, V., Acharya, R., & Lim, C. M. (2008). Automatic identification of epilepsy by hos and power spectrum parameters using eeg signals: A comparative study. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 3824–3827). doi:10.1109/IEMBS.2008.4650043.

Chua, K. C., Chandran, V., Acharya, U. R., & Lim, C. M. (2011). Application of higher order spectra to identify epileptic eeg. *Journal of Medical Systems*, *35*, 1563–1571.

Cristianini, N., & Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. New York, NY, USA: Cambridge University Press.

Direito, B., Teixeira, C., Bandarabadi, M., Sales, F., & Dourado, A. (2014). Automatic warning of epileptic seizures by SVM: the long road ahead to success. *IFAC Proceedings Volumes*, *47*, 1158 – 1163. 19th IFAC World Congress.

Echauz, J., Wong, S., Smart, O., Gardner, A., Worrell, G., & Litt, B. (2008). Computation Applied to Clinical Epilepsy and Antiepileptic Devices. In I. Soltesz, & K. Staley (Eds.), *Computational Neuroscience in Epilepsy* (pp. 530 – 558). San Diego: Academic Press.

Fathima, T., Bedeeuzzaman, M., Farooq, O., & Khan, Y. U. (2011). Wavelet based features for epileptic seizure detection. *MES Journal of Technology and Management*, *2*, 108 – 112.

Faust, O., Acharya, U. R., Adeli, H., & Adeli, A. (2015). Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis. *Seizure*, *26*, 56 – 64.

Figueiredo, M. A. T., & Jain, A. K. (2002). Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*, 381–396.

Gilbert, A., Illingworth, J., & Bowden, R. (2009). Fast realistic multi-action recognition using mined dense spatio-temporal features. In *2009 IEEE 12th International Conference on Computer Vision* (pp. 925–931).

Gotman, J. (1982). Automatic recognition of epileptic seizures in the EEG. *Electroencephalography and Clinical Neurophysiology*, *54*, 530 – 540.

Gotman, J. (1990). Automatic seizure detection: improvements and evaluation. *Electroencephalography and Clinical Neurophysiology*, *76*, 317–324.

Guler, I., & Ubeyli, E. (2007). Multiclass support vector machines for eeg-signals classification. *IEEE Trans. Inf. Technol. Biomed.*, *11*, 117–126.

Guo, L., Rivero, D., Dorado, J., Munteanu, C. R., & Pazos, A. (2011). Automatic feature extraction using genetic programming: An application to epileptic eeg classification. *Expert Syst. Appl.*, *38*, 10425–10436. URL: http://dx.doi.org/10.1016/j.eswa.2011.02.118. doi:10.1016/j.eswa.2011.02.118.

Guo, L., Rivero, D., Seoane, J. A., & Pazos, A. (2009). Classification of eeg signals using relative wavelet energy and artificial neural networks. In *Proceedings of the first ACM/SIGEVO Summit on Genetic and Evolutionary Computation* (pp. 177–184).

Husain, S. J., & Rao, K. S. (2012). Epileptic Seizures Classification from EEG Signals using Neural Networks. In *International Proceedings on Computer Science and Information Technology* (pp. 269–273). volume 37.

Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Comput. Surv.*, *31*, 264–323.

Janjarasjitt, S. (2010). Classification of the epileptic eegs using the wavelet-based scale variance feature. *International Journal of Applied Biomedical Engineering*, *3*, 43–50.

Joachims, T. (1997). *Text Categorization with Support Vector Machines: Learning with Many Relevant Features*. Technical Report 23 Universität Dortmund, LS VIII-Report.

Kandaswamy, A., Kumar, C., Ramanathan, R., Jayaraman, S., & Malmurugan, N. (2004). Neural classification of lung sounds using wavelet coefficients. *Computers in Biology and Medicine*, *34*, 523 – 537.

Kannathal, N., Lim, C., Acharya, U., & Sadasivan, P. (2005). Entropies for detection of epilepsy in eeg. *Comput. Methods Programs Biomed.*, *80*, 187–194.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.*, *24*, 881–892.

Kıymık, M. K., Güler, I., Dizibüyük, A., & Akın, M. (2005). Comparison of stft and wavelet transform methods in determining epileptic seizure activity in eeg signals for real-time application. *Computers in biology and medicine*, *35*, 603–616.

Li, J., Yan, J., Liu, X., & Ouyang, G. (2014). Using permutation entropy to measure the changes in eeg signals during absence seizures. *Entropy*, *16*, 3049. URL: http://www.mdpi.com/1099-4300/16/6/3049. doi:10.3390/e16063049.

Martis, R. J., Acharya, U. R., Tan, J.-H., Petznick, A., Tong, L., Chua, C. K., & Ng, E. Y. K. (2013). Application of intrinsic time-scale decomposition (itd) to eeg signals for automated seizure prediction. *International journal of neural systems*, *23 5*, 1350023.

McLachlan, G. J., & Krishnan, T. (2007). Extensions of the em algorithm. In *The EM Algorithm and Extensions* (pp. 159–218). John Wiley & Sons, Inc.

N, S., & Thanushkodi, D. K. (2009). Automated Epileptic Seizure Detection in EEG Signals Using FastICA and Neural Network . *International Journal of Soft Computing and Its Applications*, *1*, 1–14.

Omerhodzic, I., Avdakovic, S., Nuhanovic, A., & Dizdarevic, K. (2013). Energy distribution of EEG signals: EEG signal wavelet-neural network classifier. *CoRR*, *abs/1307.7897*.

Organization, W. H. (2016). Epilepsy Fact sheet, number 999. URL: http://www.who.int/mediacentre/factsheets/fs999/en/.

Orhan, U., Hekim, M., & Ozer, M. (2011). EEG signals classification using the k-means clustering and a multilayer perceptron neural network model. *Expert Systems with Applications*, *38*, 13475 – 13481.

Patnaik, L., & Manyam, O. K. (2008). Epileptic EEG detection using neural networks and post-classification. *Computer Methods and Programs in Biomedicine*, *91*, 100 – 109.

Pippa, E., Mporas, I., & Megalooikonomou, V. (2015). Automatic estimation of the optimal ar order for epilepsy analysis using eeg signals. In *Bioinformatics and Bioengineering (BIBE), 2015 IEEE 15th International Conference on* (pp. 1–4).

Polat, K., & Güneş, S. (2007). Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation*, *187*, 1017–1026.

Ripley, B. D., & Hjort, N. L. (1995). *Pattern Recognition and Neural Networks*. (1st ed.). New York, NY, USA: Cambridge University Press.

Shoeb, A., Edwards, H., Connolly, J., Bourgeois, B., Treves, T., & Guttag, J. (2004). Patient-specific seizure onset detection. In *Engineering in Medicine and Biology Society, 2004. IEMBS '04. 26th Annual International Conference of the IEEE* (pp. 419–422). volume 1.

Siuly, S., & Li, Y. (2015). Designing a robust feature extraction method based on optimum allocation and principal component analysis for epileptic EEG signal classification. *Computer Methods and Programs in Biomedicine*, *119*, 29 – 42.

Subasi, A., & Erelebi, E. (2005). Classification of EEG signals using neural network and logistic regression. *Computer Methods and Programs in Biomedicine*, *78*, 87 – 99.

Tsiouris, K. M., Tzallas, A. T., Markoula, S., Koutsouris, D., Konitsiotis, S., & Fotiadis, D. I. (2015). A Review of Automated Methodologies for the Detection of Epileptic Episodes Using Long-Term EEG Signals. *Handbook of Research on Trends in the Diagnosis and Treatment of Chronic Conditions*, (pp. 231–261).

Tzallas, A. T., Tsipouras, M. G., & Fotiadis, D. I. (2007). Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Intell. Neuroscience*, *2007*, 18:1–18:13.

Tzallas, A. T., Tsipouras, M. G., & Fotiadis, D. I. (2009). Epileptic seizure detection in eegs using time-frequency analysis. *Trans. Info. Tech. Biomed.*, *13*, 703–710.

Übeyli, E. D. (2010). Least squares support vector machine employing model-based methods coefficients for analysis of eeg signals. *Expert Systems with Applications*, *37*, 233–239.

Upadhyay, R., Padhy, P., & Kankar, P. (2016). A comparative study of feature ranking techniques for epileptic seizure detection using wavelet transform. *Computers & Electrical Engineering*, *53*, 163–176.

Valipour, M. (2012a). Comparison of surface irrigation simulation models: full hydrodynamic, zero inertia, kinematic wave. *Journal of Agricultural Science*, *4*, 68.

Valipour, M. (2012b). Sprinkle and trickle irrigation system design using tapered pipes for pressure loss adjusting. *Journal of Agricultural Science*, *4*, 125.

Valipour, M. (2016). Optimization of neural networks for precipitation analysis in a humid region to detect drought and wet year alarms. *Meteorological Applications*, *23*, 91–100.

Valipour, M., & Singh, V. P. (2016). Global experiences on wastewater irrigation: Challenges and prospects. In *Balanced Urban Development: Options and Strategies for Liveable Cities* (pp. 289–327). Springer.

Vapnik, V. N. (1995). *The Nature of Statistical Learning Theory*. New York, NY, USA: Springer-Verlag New York, Inc.

Wang, J., Liu, P., She, M. F., Nahavandi, S., & Kouzani, A. (2013). Bag-of-words representation for biomedical time series classification. *Biomedical Signal Processing and Control*, *8*, 634 – 644.

Wang, J., Liu, P., She, M. F., Nahavandi, S., & Kouzani, A. Z. (2012). Bag-of-words representation for biomedical time series classification. *CoRR*, .

Wilson, S. B., Scheuer, M. L., Emerson, R. G., & Gabor, A. J. (2004). Seizure detection: evaluation of the reveal algorithm. *Clinical Neurophysiology*, *115*, 2280 – 2291.

Winterhalder, M., Maiwald, T., Voss, H., Aschenbrenner-Scheibe, R., Timmer, J., & Schulze-Bonhage, A. (2003). The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. *Epilepsy & Behavior*, *4*, 318–325.

Yannopoulos, S. I., Lyberatos, G., Theodossiou, N., Li, W., Valipour, M., Tamburrino, A., & Angelakis, A. N. (2015). Evolution of water lifting devices (pumps) over the centuries worldwide. *Water*, *7*, 5031–5060.