

Accepted Manuscript

Title: Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods

Author: Jalil Rasekhi Mohammad Reza Karami Mollaei
Mojtaba Bandarabadi Cesar A. Teixeira Antonio Dourado



PII: S0165-0270(13)00124-6
DOI: <http://dx.doi.org/doi:10.1016/j.jneumeth.2013.03.019>
Reference: NSM 6603

To appear in: *Journal of Neuroscience Methods*

Received date: 23-1-2013
Revised date: 23-3-2013
Accepted date: 25-3-2013

Please cite this article as: Rasekhi J, Mollaei MRK, Bandarabadi M, Teixeira CA, Dourado A, Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods, *Journal of Neuroscience Methods* (2013), <http://dx.doi.org/10.1016/j.jneumeth.2013.03.019>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Preprocessing effects of 22 linear univariate features on the performance of seizure prediction methods

Jalil Rasekhi¹, Mohammad Reza Karami Mollaei¹, Mojtaba Bandarabadi², Cesar A. Teixeira², and Antonio Dourado²

¹Department of Biomedical Engineering, Faculty of Engineering, Babol Noshirvani University of Technology, Babol, Iran

²CISUC/DEI, Center for Informatics and Systems of the University of Coimbra, Department of Informatics Engineering, Polo II, 3030-290 Coimbra Portugal

Abstract

Combining multiple linear univariate features in one feature space and classifying this feature space using machine learning methods could predict epileptic seizures, in the patients suffering from refractory epilepsy. For each patient, a set of twenty-two linear univariate features are extracted from 6 electroencephalogram (EEG) signals, and make a 132 dimensional feature space. Preprocessing and normalization methods of the features, affect the outputs of the seizure prediction algorithm, and are studied in terms of alarm sensitivity and false prediction rate (FPR). The problem of approximately choosing an optimal preictal time was tackled using 4 distinct values; 10, 20, 30, and 40 minutes. Seizure prediction problem has traditionally been considered as a two-class classification problem, which is also exercised here.

The studies have been done on the features obtained from 10 patients. For each patient, 48 different combinations of methods are compared to find the best configuration. Normalization through dividing by maximum and smoothing, are found as the best configuration in most of the patients. Results also indicate that applying machine learning methods on a multidimensional feature space resulting from 22 univariate features could predict seizure onsets with high performance. In average, the seizures were predicted in 73.9% of the cases (34 out of 46 in 737.9 hours of test data), with a FPR of $0.15h^{-1}$.

Keywords: seizure prediction; epilepsy; classification; features selection; space reduction.

1. Introduction

Epilepsy is well-known by the sudden extraordinary seizures which occur quite randomly and are caused by intensive, unexplained, and coherent neuronal activities in the brain (Fisher et al., 2005). Epilepsy is caused either by brain injuries or by out of balance chemicals in the brain. In fact, anything that injures the normal brain tissue can lead to seizures. However, in more than 50% of the epileptic cases, no certain cause can be identified (Sun et al., 2001). Furthermore, the type of injury that can lead in a seizure is age-dependent. Children for example, during the very first years are quite easily affected by epilepsy, through birth-related issues, by inheritance, usual infections such as meningitis and even uncontrolled fevers. In the middle years however, epilepsy is mainly kicked off by accidents which damage brain tissue. Such accidents include but

are not limited to infections, alcohol, and side effects of medications. Finally, strokes and traumas are main triggers of the epileptic seizures throughout the last years of one's lifetime.

Almost 1% of people develop epilepsy during their lifetime. To the current figures, this translates to 60 million men and women worldwide. Epilepsy is usually controlled by medication, but not cured. Furthermore, for 30% of epileptic patients, seizures are not even controllable by the state of the art medicine (Carney et al., 2011). For these patients, only two choices will remain: to go through uncertain/risky brain surgeries, or at least receive a pre-warning in some way, so as to better escape harmful pending situations. Epileptic seizure prediction has emerged in response to the latter. For several decades, signal processing researchers have challenged brain to find algorithms for uncovering dynamics of Epilepsy by employing epileptogenic EEG signals (D'Alessandro et al., 2003; Le Van Quyen et al., 2001; Mirowski et al., 2009; Rogowski et al., 1981; Viglione and Walsh, 1975). However, efforts have not been fully successful, and promising results of seizure prediction studies were usually challenged and rejected by later works (Andrzejak et al., 2009; Mormann et al., 2007; Mormann et al., 2005; Stacey et al., 2011).

The features can be either linear or non-linear. Linear features are extracted using phase/frequency and amplitude information of the signal. Non-linear features on the other hand, try to extract the non-linear dynamics of the EEG signals. In representing EEG signals in the formal non-linear dynamics one needs two key terms namely 'state' and 'dynamics'. The state defines the system at a given time while dynamics involves the rules describing how the system will evolve over the time. The measures extracted from a single channel are called univariate, whereas those representing relations between two or more different brain regions are called bivariate or multivariate features respectively.

Several studies have used linear univariate features with promising results (McSharry et al., 2003; Mormann et al., 2005). Among these are statistical moments such as variance and kurtosis. According to these studies, variance has shown a decrease, while kurtosis has exhibited an increase during preictal time (Mormann et al., 2005). Power spectral features have also shown good capability for distinguishing preictal from interictal (Cerf and El Ouasdad, 2000; Mormann et al., 2005; Park et al., 2011). Hjorth parameters, mobility and complexity, exhibit significant increase during preictal time (Mormann et al., 2005). Decorrelation time, as the first zero-crossing of the autocorrelation function, revealed a decrease during preictal in comparison to the interictal period (Mormann et al., 2005). Although these features have shown good results in the first studies, and could provide prediction performances better than a random predictor, however when applied on the new datasets and new recordings, the brilliant results are questioned. Furthermore, their combination in a feature space would bring new hopes into the seizure prediction community. In fact, such combinatorial methods have attracted overwhelming curiosity from scientists, and their advantages over single feature solutions are identified (D'Alessandro et al., 2003; Direito et al., 2008; Feldwisch-Drentrup et al., 2010; Mirowski et al., 2009).

Combination of 22 univariate features extracted from six EEG channels is studied here. Features include time domain, frequency domain, and time/frequency domain features. Machine learning (ML) approaches such as classification or clustering are widely used for the combination of features. ML methods are normally sensitive to the range of features and their span (Graf et al., 2003). Hence, preprocessing of the features prior to feeding them into machine learning algorithms may play a key role in bettering the performance of the method. To verify the effects of preprocessing blocks including outlier processing, smoothing, and normalization of the extracted features a comparative study has been carried out.

In seizure prediction problem, one has four known states namely interictal (between two consecutive seizures, the normal brain state), preictal (before seizure, when the brain is evolving to seizure), ictal (the seizure) and postictal (after seizure, the brain evolves to normal state). And the goal of seizure prediction is to detect the preictal state. The most common strategy with seizure prediction methods is to first extract some features from the windowed EEG signals over time and then to classify the resulting set of features as a two class problem, namely preictal and non-preictal. It should be emphasized that the adequate choice of preictal time is of high significance for obtaining good prediction results. Considering four preictal times, three normalization methods, removing of the outlier on/off, and smoothing on/off, a total of 48 different combinations of a typical seizure prediction algorithm should be taken into account and investigated, as depicted in Fig.1. Results are compared to find out which combination better predicts seizure onsets, with regard to prediction sensitivity (SEN) and false prediction rate (FPR).

2. Materials and methods

2.1. Subjects

10 subjects were selected from European database on epilepsy (Klatt et al., 2012), 15-57 years of age, and suffering from refractory partial epilepsy. The recordings had been carried out in two Epilepsy centers at Portugal (University Hospitals of Coimbra) and France (Pitié-Salpêtrière Hospital of Paris). 8 of the recordings were made using scalp electrodes through the 10-20 standard system, and two were recorded invasively. Further details regarding the patients are listed in Table 1.

Table 1. Information for the 10 studied patients

ID	Sex	Age	Onset age	Rec. type	Rec. time (h)	No. Seiz	Seizure type	Localize	Samp. rate	Hospital
1	M	57	6	Scalp	110.2	8	CP(7), UC(1)	Right T	512	Coimbra
2	M	20	--	Scalp	89.8	13	CP(12), UC(1)	Left T	512	Coimbra
3	M	15	3	Scalp	113.5	7	CP(6), UC(1)	Right T	512	Coimbra
4	M	16	12	Scalp	91.9	7	SG(1), UC(6)	Left T	512	Coimbra
5	M	35	1	Scalp	141.5	4	CP(3), UC(1)	Right T	512	Paris
6	M	41	19	Scalp	140.6	6	CP(4), SP(1), UC(1)	Right T	400	Paris
7	F	21	1	Scalp	252.6	9	UC(9)	Right T	400	Paris
8	M	46	19	Scalp	91.7	7	UC(6), CP(1)	Right T	400	Paris
9	F	23	15	Invasive	145.1	13	CP(13)	Right T	400	Paris
10	M	36	1	Invasive	211.7	12	UC(12)	Right F	400	Paris
A/T.	--	31	8.5	---	1388	86	---	---	---	---

* Localize: Localization of seizures; T: Temporal lobe, F: Frontal lobe, P: Parietal lobe (between frontal and occipital).

* Seizure type: Type of the clinical seizures; SP: Simple Partial, CP: Complex Partial, SG: Secondarily Generalized, UC: Unclassified.

2.2. Seizure prediction

The block diagram of the utilized seizure predictor is depicted in Fig. 1, where prediction relies on the SVM classifier trained to discriminate different states of the incoming EEG signal. The diagram also includes a univariate feature extraction stage using 5-sec windows, a preprocessing unit, as well as a regularization box for decision-making.

Figure 1. Block diagram for employed SVM seizure predictor

As seen in this figure, the feature extractor is first applied to the raw EEG signal, and the resulting univariate feature vector is further processed by some method in preprocessing unit. The features are subsequently labeled based on the selected preictal time, and then fed to the SVM classifier so as to discriminate the resulting sequence into predefined brain states. The outputs of the SVM are also regularized by a moving average approach in order to reduce the number of false alarms.

2.3. Feature extraction

For each patient, 22 linear univariate features are extracted from six channels, three close to the seizure focus and three far away. The raw EEG data is first segmented using 5-sec non-overlapping windows, and then processed by an infinite impulse response (IIR) forward-backward Butterworth 50 Hz notch filter to cancel out strong 50Hz distortions caused by AC power supply. The forward-backward scheme is used to eliminate the undesirable phase shift. Linear univariate features, extracted from EEG signal, can be characterized according to both phase/frequency and amplitude information of the signal of each channel. A short introduction to the utilized features is presented in the following sections:

2.3.1. Spectral Power features

The overall bandwidth of the EEG signal is typically divided into particular frequency bands namely Delta ' δ ' (less than 4 Hz), Theta ' θ ' (4-8 Hz), Alpha ' α ' (8-12 Hz), Beta ' β ' (13-30 Hz), and Gamma ' γ ' (greater than 30 Hz). Experiments have shown that brain activity increases inside some frequency bands during particular situations, whereas may decrease for other frequency bands. Mormann (Mormann et al., 2005) pointed out a decrease of power in the Delta band during the transient stage from interictal to ictal onset, in relative to the inter-ictal state. This was also coupled with a relative power decrease in the other sub-bands.

Spectral power of each sub-band is calculated by first applying Fast Fourier Transform (FFT) on the segmented signals and then summing up the resulting Fourier coefficients falling within that particular sub-band. On the other hand, normalized spectral power provides a more robust measure to the fluctuations in the patient's daily life, for revealing the pre-ictal state (Van Laar et al., 2011). To isolate the epileptogenic related variations from inclined brain signal levels caused by daily life changes, it has been suggested to employ normalized power values instead of absolute values. The use of normalized power decreases the dependency of the sub-band power values to the total power, providing improved measures for comparison. Spectral power features are normalized by dividing to the total power.

2.3.2. Statistical moments

Statistical analysis is generally carried out by calculating four well-known parameters namely: “mean”, “variance”, “skewness”, and “kurtosis”. The four parameters, also known respectively as 1st, 2nd, 3rd and 4th statistical moments, provide information regarding the amplitude distribution of time series. While “mean” and “variance” provide information on the location and span of the distribution, “skewness and “kurtosis” provide information on the shape of time series (Mormann et al., 2007). For symmetric amplitude distributions (ADs), skewness is zero, whereas it will be nonzero for asymmetric ADs, and AD will be tilting to the right or left respectively, in case the skewness becomes negative or positive. Kurtosis on the other hand, reveals the degree of flatness (or peakedness) of the AD of a time series. These statistical measures were investigated in our study, since other researches (Mormann et al., 2005; Rogowski et al., 1981) have shown that they significantly change during preictal period in comparison to the interictal state. More specifically, a decrease in variance coupled with an increase for kurtosis has been reported during preictal period (Mormann et al., 2005).

2.3.3. Hjorth parameter

Hjorth introduced a set of three time domain parameters to quantitatively describe EEG signals which include activity, mobility and complexity (Hjorth, 1970). Mormann (Mormann et al., 2005) have reported a significant increase in the mobility and complexity during preictal period. Therefore we have employed mobility and complexity parameters extracted from EEG signal. The mobility is defined as the root mean square (RMS) of slopes of the EEG signal divided by the RMS of amplitudes inside a moving window. Mobility is known to provide an estimate of the mean frequency. The complexity on the other hand, gives a measure of the RMS of the rate of slope changes with reference to an ideal possible curve. It gives an estimate of the bandwidth of the signal.

2.3.4. Long-term energy

Accumulated energy, also known as Long-term energy has been used in some studies and showing ability for predicting seizures. It has been shown that the bursts of long-term energy increase as the seizures are approached (Litt et al., 2001). The idea of monitoring long-term energy bursts to predict seizures is based on the hypothesis that seizure generation in the temporal lobe epilepsy, is developed by a chain of EEG events occurring for some hours. The feature is obtained by first evaluating the instantaneous energy of EEG sequence $E_i(\mathbf{n})$,

$$E_i(\mathbf{n}) = \mathbf{x}(\mathbf{n})^2 \quad (1)$$

then passing the resulting energy sequence through a length-M moving average to obtain (2):

$$E_k = \frac{1}{M} \sum \mathbf{x}(\mathbf{n})^2 \quad (2)$$

and finally using a second length-N averaging window on resulting E_k sequence this time, to obtain the accumulated energy(3):

$$AE_m = \frac{1}{N} \sum E_k + AE_{m-1} \quad (3)$$

This process is equivalent to integrating over the instantaneous EEG energy values (Litt et al., 2001).

2.3.5. Autoregressive error

Another method commonly used in analyzing EEG signals is auto-regressive (AR) model. In this modeling approach the current value of a given time series is determined only according to the weighted sum of previous values of the same time series as well as noise. This is a linear model and assumes the signal to be stationary. However EEG signals are non-stationary and should be segmented into shorter semi-stationary parts. Auto-regressive modeling has been reported to locate preictal changes prior to seizure onset (Salant et al., 1998). In our study the mean square error (MSE) between the original EEG time series and the synthesized output from the related order-10 AR model, is used as a feature. Burg’s method (Priestley, 1981) was employed to build the AR model.

2.3.6. Decorrelation time

Autocorrelation of signal $\mathbf{x}(\mathbf{t})$ is a function defined as the correlation between $\mathbf{x}(\mathbf{t})$ and its shifted-by- τ versions. It is therefore a pure function of this time shift. Autocorrelation is useful in detecting the stationarity of the time series of interest. “Decorrelation time” on the other hand is defined as the first zero-crossing of the autocorrelation function (Box and Jenkins, 1990). Mormann (Mormann et al., 2005) employed decorrelation time on epileptogenic EEG signals and could distinguish preictal from interictal periods by seeking a decrease in this measure.

2.3.7. Spectral edge frequency and spectral edge power

The spectral power of EEG signals distributes across all frequencies. However, a good part of this power is limited to <40Hz frequencies. A useful quantization of how the power is actually scattered within these frequencies is by calculating features known as spectral edge frequency and spectral edge power. Spectral edge frequency stands for the frequency below which α percent of the overall power of the signal is located. α can hold any values ranging from 0 to 100. In the context of seizure prediction however, α value of 50% has been successfully employed, describing the minimum frequency up to which 50% of the overall power of the 0-40Hz band is contained. Obviously, the power covered under spectral edge frequency is called spectral edge power (Stanski et al., 1984).

2.3.8. Wavelet coefficients

Wavelet transform is one of the most applied mathematical transforms in the field of signal processing. Considering its multi-resolution nature, this transform has found significant role in processing non-stationary EEG signals. In wavelet analysis, signal is decomposed into a sum of scaled and translated variations of a mother wavelet. A mother wavelet is simply a wavy function carefully constructed so as to have certain mathematical properties. In a comparative study (Direito et al., 2011b) on 22 univariate features, and using feature selection methods, the selected features highlighted the significance of spectral information, especially within low frequency bands of the wavelet coefficient analysis. In another study (Direito et al., 2011a) also argued on the importance of wavelet coefficients vs. other univariate features. The Daubiches-4 (db4) has shown good localization properties of EEG signal in the time and frequency domains (Petrosian et al., 2000) and is used in this work. More recently special wavelets have been developed for seizure prediction, showing promising results (Bandarabadi et al., 2011).

2.4. Preprocessing of features

Although numerous studies have used different feature processing techniques to improve the performance of their prediction algorithms, none carried out a comparative study of the preprocessing effects. To fulfill this objective, we have considered four options for the preprocessing stage: outlier processing, smoothing using moving average (MA), normalization, and preictal time. Fig. 2 presents the possible combinations of the preprocessing steps. The three options are presented below in more detail.

Figure 2. The flowchart of the possible combinations of the methods to be compared

2.4.1 Outlier preprocessing

Outlier samples, although very few in numbers, yet occupy a significant portion of the feature space. In this regard, they can affect the output of the subsequent normalization methods by altering the accumulated distribution of the non-outlier (healthy) samples. Several methods have been introduced to detect outliers, reduce their effect, and thus to improve classification results. We have employed Grubb's method (Grubbs and MD., 1974) to separate outliers from the rest of samples. The values of detected outliers are replaced by the 1.1 times of the maximum value in the rest (non-outlier) of the data, to move them further away from the assembly of non-outliers. With Grubbs method, outliers are detected one at a time. In every iteration, each outlier is removed from the dataset upon detection, and the test is repeated until there are no outliers left. The test starts by the hypothesis of having no outliers. The two-sided Grubbs' test statistic is defined as:

$$G = \frac{\max_{i=1, \dots, N} |X_i - \bar{X}|}{\sigma} \quad (4)$$

where \bar{X} and σ represent the mean and standard deviation of the samples respectively. According to this test, the hypothesis of no outliers is rejected at significance level α if the following inequality holds:

$$G > \frac{N-1}{\sqrt{N}} \sqrt{\frac{t_{\alpha/(2N)-2}^2}{N-2+t_{\alpha/(2N)-2}^2}} \quad (5)$$

where $t_{\alpha/(2N)-2}^2$ is the upper critical value of the t-distribution with N-2 degrees of freedom and a significance level of $\alpha/(2N)$.

2.4.2. Smoothing of the features

The features obtained from non-stationary sources such as brain, tend to express noisy behavior. EEG features, although very turbulent, yet they fluctuate around a certain value within short periods of time. These short periods of time may be easily dominated by daily life activities as well as by different tasks being carried out by the brain. A straightforward result of this domination is reflected by increased/decreased levels of overall signal power. By smoothing of the features, rebellious nature of the features is damped, and can be better classified either visually or by machine learning methods. The smoothing process of each feature “ α ” is achieved by a 1-min length moving average window, covering 12 consecutive 5-sec epochs.

$$\bar{\alpha}_k = \frac{1}{m+1} \sum_{j=0}^m \alpha_{k-j}, \quad m = 11 \quad (6)$$

2.4.3. Normalization

The normalization of the vectors of the input space can be considered as the most basic type of preprocessing. In the contest of statistics and signal processing, as the range of values of data varies widely, feature scaling is utilized to adjust the sets of independent variables or features of data into a standard range. Some machine learning algorithms, especially large margin classifiers such as SVM, are sensitive to the scaling of data (Graf et al., 2003). To improve classification efficiency, one needs to normalize either data points or the classifier’s kernel itself. Consequently for nonlinear kernels, the only possibility is to normalize data (Graf et al., 2003). In this work normalization is employed as a preprocessing step for SVM classifier. To normalize features, three different methods are employed:

- Dividing by maximum value of the feature:

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (7)$$

- Dividing by mean of the absolute values of the feature:

$$x_{norm} = \frac{x - \text{mean}(x)}{\text{mean}(\text{abs}(x))} \quad (8)$$

- Dividing by standard deviation of the feature:

$$x_{norm} = \frac{x - \text{mean}(x)}{\text{std}(x)} \quad (9)$$

Where x and x_{norm} stand for the feature sequence being normalized and its normalized version respectively.

2.4.4. Epileptic brain states and Preictal time

Seizure prediction experts categorize epileptic brain states into four distinct classes: interictal, preictal, ictal and postictal, which have different dynamical properties (Fig. 3). In seizure prediction problem the main goal is to detect the preictal state, which precedes the ictal states. Therefore seizure onset can be predictable upon detection of preictal state. To achieve this, one strategy would be to build a two-class problem: preictal state against all other classes as a single class, or to employ the original four classes and build up a quad-class problem. The two-class approach is employed in this study.

Figure 3. Epileptic brain states

In supervised learning methods, such as SVM, training data should be selected carefully. Preictal samples' set, which is one of the classes in seizure prediction problem, is chosen based on a pre-determined preictal time. Preictal time is the time interval which is suspected to have the most important transitional information and may contain traces of forthcoming seizure. The seizure onset is initially marked by experts. Preictal time is however a matter of choice and can have a fixed value ranging from several seconds to tens of minutes, but it is surely ending at the seizure onset. In this study four different preictal times are used: 10min, 20min, 30min and 40min. The proper choice of preictal time is very essential to the overall performance of a seizure predictor, and should be covering almost all transient changes preceding a seizure onset, but excluding other EEG dynamics such as in ictal and interictal states as much as possible. The seizure prediction performances, achieved using these four preictal times, are investigated to obtain the best preictal time for each patient.

2.5. Classification

Support Vector Machines (SVMs), introduced in 1995 by Vapnik (Vapnik, 1995), are widely used in difficult nonlinear classification problems such as seizure prediction (Mirowski et al., 2008; Park et al., 2011), and still remain as one of the most power tools in signal processing and machine learning. The SVM classifier with Gaussian kernel is employed here to classify the features.

$$K(x, y) = \exp\left(-\frac{|x - y|^2}{2\sigma^2}\right) \quad (10)$$

where σ is the scale parameter (controlling the span of the kernel). Also, x and y are feature vectors in the input space. Gaussian kernel has two parameters C and γ that should be optimized. A grid search is carried out to find out the best SVM parameters (C , γ) for each 48 possible combinations of methods in Fig1. Each combination is grid-searched via 100 training/test iterations; hence a total of $48 \times 100 = 4800$ searches for each patient.

2.6. Regularization

As a common experience with hard classification problems, such as that of the two-class SVM classifier, usually not all samples are correctly classified. This issue becomes severe with EEG prediction where sample feature lengths are in the range of seconds. In such cases false alarms can be raised, lowered, then raised again, ... many times just by random. Regularization methods are designed to reduce this noisy behavior by also taking the most recent classifications into account. Using a sliding window with a size related to the preictal time value, a measure called firing power is computed (Teixeira et al., 2011). This measure quantifies the number of samples classified as preictal member, and is defined as:

$$fp[n] = \frac{\sum_{k=n-\tau}^n o[k]}{\tau} \quad (11)$$

where $fp[n]$ is the firing power at the discrete time n , τ is the number of samples per a time length equal to the preictal time, and $o[k]$ is the output of the two-class classifier. As an example of the value of τ , if the features are extracted using a 5-sec epochs, and considering a preictal time of 10 minutes, a time length of 10 minutes involves $10 \times 60 / 5$ sample features, or $\tau = 120$. That is, the firing power at every moment is calculated using the past 120 samples. $o[k] = 1$ and $o[k] = 0$ represent preictal and interictal states respectively. A firing power of one indicates that all of the samples extracted from the past preictal time have been classified as preictal. Furthermore, $fp[n]$ is a normalized function between zero and one. It is important to notice that, "regularized output" is some value between 0 and 1, and hard decision-making is necessary as the final step to generate alarm. The threshold 0.5 is selected here. There are two limitations on raising alarms:

1- After firing power is raised above the threshold value for the first time, and an alarm is generated, alarm generation is blocked for as long as a preictal time. This is important, since it is reasonable only to raise alarms prior to the seizures not after.

2- When alarm is raised and a preictal time is passed, alarm can not be raised just by noticing a firing power above the threshold. Firing power should be allowed to fall below the threshold value first, and after that alarm generation will be permitted upon a positive threshold pass.

3. Results

To optimize the prediction capability of the system presented in Fig. 1, all results are investigated to choose the best options for the four parameters of (a) Preictal time, (b) Preprocessing method, (c) Normalization approach and (d) The number of classes used in SVM classification. The comparison of resulting set of combinations would also reveal more information on the significance of each parameter as well as its effects on the overall prediction capability. Sensitivity of the raised alarms and false prediction rate are used to compare and present the results. Sensitivity is the measure of the system's ability for correct prediction of the seizure, and is defined as the fraction of correctly predicted seizures within the total

seizures. In every predicting system, false predictions are inevitable and represented by false prediction rate (FPR) parameter. The false prediction rate is the number of false predictions per time interval.

For each patient, recorded dataset is divided into two parts, one for the training of SVM classifier, and the other for test. Train set starts from the beginning of the dataset and ends by the end of third seizure. The related continuous preictal and non-preictal intervals of this training dataset are used in the training phase. Rest of dataset, forms the second part and is used in the test phase. Table II presents the three highest ranked combinations for each patient, producing the best results. Talking in average, the seizures were predicted in 73.9% of the cases, with a FPR of $0.15h^{-1}$.

Table II. Results for the 3 best combinations of the methods

ID	Preict. Time	Preprocessing	Normal. Method	SS	FPR	# test Seiz.	Test time h
1	10	Smooth	x/max	60.00	0.08	5	64.5
	20	Smooth	x/max	60.00	0.20		
	20	No change	x/max	60.00	0.22		
2	40	Smooth	x/max	50.00	0.04	6	48.5
	30	Smooth	x/max	50.00	0.09		
	10	Outlier & Smooth	x/max	50.00	0.13		
3	20	No change	x/max	75.00	0.25	4	24
	30	Smooth	x/max	75.00	0.29		
	40	Smooth	x/max	75.00	0.29		
4	10	No change	x/max	75.00	0.07	4	83.6
	30	No change	x/max	75.00	0.07		
	40	No change	x/max	75.00	0.07		
5	40	Outlier & Smooth	x/mean	100	0.02	1	65.1
	40	Smooth	x/mean	100	0.05		
	40	Outlier & Smooth	x/max	100	0.06		
6	40	Smooth	x/std	100	0.15	2	40.8
	20	Outlier & Smooth	x/max	100	0.18		
	30	No change	x/max	100	0.18		
7	20	No change	x/max	100.00	0.07	6	139.6
	40	Outlier	x/max	100.00	0.08		
	30	No change	x/max	100.00	0.09		
8	30	Outlier & Smooth	x/max	100.00	0.07	2	58.8
	10	Smooth	x/max	100.00	0.10		
	10	Smooth	x/mean	100.00	0.10		
9	30	No change	x/max	57.14	0.16	7	52.7
	30	Smooth	x/max	57.14	0.16		
	40	No change	x/max	57.14	0.16		
10	10	Smooth	x/max	77.78	0.39	9	160.3
	20	No change	x/std	77.78	0.41		
	40	No change	x/max	66.67	0.19		

Fig. 4 shows the average results of SS and FPR, achieved for each combination when applied over all patients. This figure provides a more generalized perspective for the achieved results from each combination. Obviously the methods involving smoothing, outlier removal and using longer preictal time are providing better results. The lower table identifies the configuration corresponding to the results in the upper points. For example point number 4 (SEN=65 and FPR=0.21) corresponds to a preictal of 10m with outlier removal and max smoothing.

Figure 4. Average of SS and FPR for diferente combinations

4. Discussion

Capability of linear univariate features and effects of different preprocessing methods on these features was studied here. In this section, we will discuss the results.

4.1. Twenty-two linear univariate features

Contrary to some studies which had previously questioned the prediction ability of most of these univariate features, their combination in a single high dimensional feature space, coupled with machine learning methods, showed good capability. In this study, the combination of 22 univariate features in a high dimensional feature space and classifying the resulting feature space using SVM has produced promising results. The 22 linear univariate features were extracted from 6 recording channels, 3 on the seizure focus, and 3 on its opposite side. The feature space has a dimension of 132, and is classified in two classes; preictal state, and non-preictal state.

To find relative behavior of features during preictal time and during normal state (interictal), the values of 22 features were averaged separately over preictal and interictal samples. Averaging was performed patient by patient for a single focal channel, and the preictal time was considered as 30 minutes. The results are tabulated in Table IV.

Table IV. The average ratio (in percent) of feature changes for preictal state to those of interictal state

Feature name	Patient ID	500	2100	5100	6000	1308603	1318803	1327903	1328803	1272703	1329003
Error AR coefficient		18.87	28.97	-19.70	-12.63	0.61	68.32	24.25	-40.12	11.74	0.68
Decorrelation time		-2.22	6.45	-10.79	3.80	54.77	-13.14	-18.87	9.34	0.03	-0.95
Long-term energy		29.66	34.26	-15.92	-16.33	10.00	111.12	-22.49	-40.32	1.21	-0.39
Hjorth - Mobility		56.44	-1.63	39.80	1.90	-57.58	62.84	-70.63	40.66	1.96	-14.44
Hjorth - Complexity		51.46	-0.65	38.47	5.84	-60.38	59.08	-69.92	34.68	-5.95	-18.14
Relative power Delta		-10.92	-0.74	-1.43	0.88	16.38	-15.41	21.62	-19.47	5.39	3.54
Relative power Theta		-9.56	-29.87	19.75	35.22	-1.80	-20.93	-9.35	-3.43	-18.64	-1.64
Relative power Alpha		-1.43	-34.62	21.72	25.16	19.58	-12.21	-34.47	9.47	-19.99	-7.47
Relative power Beta		17.85	-27.54	15.61	1.11	33.51	18.51	-63.32	47.30	-18.16	-7.56
Relative power Gamma		63.07	-2.07	38.41	-6.42	-44.85	76.22	-73.69	50.91	14.20	-10.24
Spectral edge freq.		3.11	-13.30	11.90	11.32	-2.53	0.32	-13.34	11.59	-12.70	-3.11
Spectral edge power		19.09	33.77	-18.66	-20.44	20.71	81.93	47.69	-44.51	25.44	3.77
Mean		0.13	3.23	-0.28	-4.78	-0.48	-5.17	-1.47	-4.44	-1.76	-1.83
Variance		32.10	32.63	-15.11	-16.66	9.25	116.19	13.35	-42.19	1.72	-0.56
Skewness		-6.33	4.72	-0.54	-2.36	0.58	4.00	-7.28	1.15	1.06	-0.91
Kurtosis		28.87	-9.86	-4.23	-12.31	35.31	79.25	-38.21	1.90	26.24	1.71
Wavelet energy band1		103.54	17.53	46.37	-0.25	-55.06	136.49	-52.95	-19.73	5.31	-20.53
Wavelet energy band2		110.43	15.21	42.00	-5.43	-22.63	160.14	-55.10	-17.59	6.11	-13.47
Wavelet energy band3		108.95	12.16	31.78	-11.72	-9.11	167.19	-51.39	-14.38	8.91	-14.58
Wavelet energy band4		51.39	-0.83	11.94	-1.16	50.54	101.83	-17.29	-17.13	15.70	-6.46
Wavelet energy band5		3.87	-7.34	-6.67	29.22	3.39	91.84	9.03	-38.56	8.91	-6.46
Wavelet energy band6		4.00	4.84	-9.00	14.33	-8.45	105.80	12.53	-34.69	18.69	-2.34

In table IV, for each patient, the five features that have most significant changes during the preictal time in comparison to the interictal, are shown in bold. These five features include relative power of the Gamma band (appearing 6 times), Hjorth Mobility (5 times), Wavelet energy band1 (5 times), Wavelet energy band2 (5 times), Spectral edge power (4 times), Variance (4 times).

4.2. Preprocessing methods

In summary, smoothing and outlier preprocessing methods could improve the results. Outlier samples have side effects by damping the feature space during normalization, thus removing or putting a threshold on them will let the feature values to better spread across the feature space. "Dividing by the maximum value" of the feature produces better results, compared to "dividing by standard deviation or mean value" of that feature.

In terms of preictal time, it is evident from the diagram of Fig. 4 that highest preictal time (40min) provides better sensitivity but highest FPR for higher numbers of combinations. Lowest preictal time (10min) on the other hand, generates lowest sensitivity but lowest FPR for higher numbers of combinations as well. By using higher preictal times, although sensitivity is increased, but more chance is also given to the false positives to be raised. The similar conclusion is also made for the lowest preictal time. Across this diagram average SEN and average FPR patterns are closely following each other, and setting a trade-off between best SEN and best FPR values. The pattern is particularly not followed by average FPR for small number of combinations (combinations of 39, 40 and 41).

4.3. Type of recording

In order to compare the effect of using different types of recordings, iEEG and EEG recordings were investigated. iEEG has localized information of a specific part of the brain, and lacks useful data regarding the general brain state. On the other hand, the EEG signal recorded via skull involves higher quality information about this general brain state. The results indicate that EEG recordings generate slightly better predictions compared to that of iEEG recordings using 22 univariate features. In fact, iEEG based predictions achieve an average SEN of 68.7% and FPR of $0.33h^{-1}$, while predictions using EEG recordings generate an average SEN of 76.67% and FPR of $0.08h^{-1}$. It is therefore strongly suggested to employ EEG recordings for making predictions, as they better reflect the general state of brain. Moreover, this becomes particularly important when the recording channels are selected in a more separated way.

4.4. Number of channels

The reasons behind selecting three channels from focal area and three from an area far from the seizure focus are: (1) Non-focal areas although having much weaker preictal traces, but carry valuable information concerning the general brain state. This additional information thus can be useful in better detecting epileptogenic transient changes, (2) Employing

higher numbers of recording channels is avoided in this work. More channels are not favorable to be implemented in portable systems, are energy demanding, and harder to be realized as implants, (3) To reduce complexity on one hand, and profiting from better spatial information on the other hand, only six channels were selected.

5. Conclusions

By Combining multiple linear low discriminative univariate measures in one feature space and classifying them using machine learning methods epileptic seizures could be predicted for 10 patients with long term EEG records with high sensitivity and a low FPR. Preprocessing and normalization methods of the features were studied in terms of alarm sensitivity and false prediction rate (FPR). Smoothing combined with normalizing by maximum of the feature produced better results in most cases.

Future work could focus on reducing the dimension of the feature space, by incorporating feature ranking methods. This might reduce the destructive effects of non discriminative features on the performance of classifiers.

Acknowledgements. Part of this work has been done in the Center for Informatics and Systems of the University of Coimbra, Portugal (CISUC) during a research stay of Mr. Jalil Raseki. The authors acknowledge and thank the access to the EPILEPSIAE (FP7 211713 Grant, www.epilepsiae.eu) database in CISUC during this stay. MB would particularly like to acknowledge the Portuguese Foundation for Science and Technology (FCT- SFRH / BD / 71497 / 2010).

References

- Andrzejak RG, Chicharro D, Elger CE, Mormann F. Seizure prediction: Any better than chance? *Clinical Neurophysiology*, 2009; 120: 1465-78.
- Bandarabadi M, Teixeira CA, Sales F, Dourado A. Wepilet, optimal orthogonal wavelets for epileptic seizure prediction with one single surface channel. *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011: 7059-62.
- Box GEP, Jenkins G. *Time Series Analysis, Forecasting and Control*. Holden-Day, Incorporated, 1990.
- Carney PR, Myers S, Geyer JD. Seizure prediction: Methods. *Epilepsy & Behavior*, 2011; 22, Supplement 1: S94-S101.
- Cerf R, El Ouasdad EH. Spectral analysis of stereo-electroencephalograms: preictal slowing in partial epilepsies. *Biol. Cybern.*, 2000; 83: 399.
- D'Alessandro M, Esteller R, Vachtsevanos G, Hinson A, Echaz J, Litt B. Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. *Biomedical Engineering, IEEE Transactions on*, 2003; 50: 603-15.
- Direito B, Dourado A, Vieira M, Sales F. Combining Energy and Wavelet Transform for Epileptic Seizure Prediction in an Advanced Computational System. *BioMedical Engineering and Informatics, 2008. BMEI 2008. International Conference on*, 2008: 380-5.
- Direito B, Duarte J, Teixeira C, Schelter B, Le Van Quyen M, Schulze-Bonhage A, Sales F, Dourado A. Feature selection in high dimensional EEG features spaces for epileptic seizure prediction. *Proc. of the 18th IFAC World Congress*, 2011a.
- Direito B, Ventura F, Teixeira C, Dourado A. Optimized feature subsets for epileptic seizure prediction studies. *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011b: 1636-9.
- Feldwisch-Drentrup H, Schelter B, Jachan M, Nawrath J, Timmer J, Schulze-Bonhage A. Joining the benefits: combining epileptic seizure prediction methods. *Epilepsia*, 2010; 51: 1598-606.
- Fisher RS, Boas WvE, Blume W, Elger C, Genton P, Lee P, Engel J. Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia*, 2005; 46: 470-2.
- Graf ABA, Smola AJ, Borer S. Classification in a normalized feature space using support vector machines. *Neural Networks, IEEE Transactions on*, 2003; 14: 597-605.
- Grubbs FE, MD. ABRLAPG. *Procedures for Detecting Outlying Observations in Samples*. Defense Technical Information Center, 1974.
- Hjorth B. EEG analysis based on time domain properties. *Electroencephalography and clinical neurophysiology*, 1970; 29: 306-10.
- Klatt J, Feldwisch-Drentrup H, Ihle M, Navarro V, Neufang M, Teixeira C, Adam C, Valderrama M, Alvarado-Rojas C, Witon A, Le Van Quyen M, Sales F, Dourado A, Timmer J, Schulze-Bonhage A, Schelter B. The EPILEPSIAE database: An extensive electroencephalography database of epilepsy patients. *Epilepsia*, 2012; 53: 1669-76.
- Le Van Quyen M, Martinerie J, Navarro V, Boon P, D'Havé M, Adam C, Renault B, Varela F, Baulac M. Anticipation of epileptic seizures from standard EEG recordings. *The Lancet*, 2001; 357: 183-8.
- Litt B, Esteller R, Echaz J, D'Alessandro M, Shor R, Henry T, Pennell P, Epstein C, Bakay R, Dichter M, Vachtsevanos G. Epileptic Seizures May Begin Hours in Advance of Clinical Onset: A Report of Five Patients. *Neuron*, 2001; 30: 51-64.

- McSharry PE, Smith LA, Tarassenko L. Comparison of predictability of epileptic seizures by a linear and a nonlinear method. *Biomedical Engineering, IEEE Transactions on*, 2003; 50: 628-33.
- Mirowski P, Madhavan D, LeCun Y, Kuzniecky R. Classification of patterns of EEG synchronization for seizure prediction. *Clinical Neurophysiology*, 2009; 120: 1927-40.
- Mirowski PW, Yann L, Madhavan D, Kuzniecky R. Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG. *Machine Learning for Signal Processing, 2008. MLSP 2008. IEEE Workshop on*, 2008: 244-9.
- Mormann F, Andrzejak RG, Elger CE, Lehnertz K. Seizure prediction: the long and winding road. *Brain*, 2007; 130: 314-33.
- Mormann F, Kreuz T, Rieke C, Andrzejak RG, Kraskov A, David P, Elger CE, Lehnertz K. On the predictability of epileptic seizures. *Clinical Neurophysiology*, 2005; 116: 569-87.
- Park Y, Luo L, Parhi KK, Netoff T. Seizure prediction with spectral power of EEG using cost-sensitive support vector machines. *Epilepsia*, 2011; 52: 1761-70.
- Petrosian A, Prokhorov D, Homan R, Dasheiff R, Wunsch D. Recurrent neural network based prediction of epileptic seizures in intra-and extracranial EEG. *Neurocomputing*, 2000; 30: 201-18.
- Priestley MB. *Spectral analysis and time series / M.B. Priestley*. Academic Press: London ; New York ;, 1981.
- Rogowski Z, Gath I, Bental E. On the prediction of epileptic seizures. *Biol. Cybern.*, 1981; 42: 9-15.
- Salant Y, Gath I, Henriksen O. Prediction of epileptic seizures from two-channel EEG. *Med. Biol. Eng. Comput.*, 1998; 36: 549-56.
- Stacey W, Le Van Quyen M, Mormann F, Schulze-Bonhage A. What is the present-day EEG evidence for a preictal state? *Epilepsy research*, 2011; 97: 243-51.
- Stanski D, Hudson R, Homer T, Saidman L, Meathe E. Pharmacodynamic modeling of thiopental anesthesia. *Journal of Pharmacokinetics and Biopharmaceutics*, 1984; 12: 223-40.
- Sun DA, Sombati S, DeLorenzo RJ. Glutamate Injury-Induced Epileptogenesis in Hippocampal Neurons: An In Vitro Model of Stroke-Induced "Epilepsy". *Stroke*, 2001; 32: 2344-50.
- Teixeira CA, Direito B, Feldwisch-Drentrup H, Valderrama M, Costa RP, Alvarado-Rojas C, Nikolopoulos S, Le Van Quyen M, Timmer J, Schelter B, Dourado A. EPILAB: A software package for studies on the prediction of epileptic seizures. *Journal of Neuroscience Methods*, 2011; 200: 257-71.
- Van Laar JOEH, Peters CHL, Houterman S, Wijn PFF, Kwee A, Oei SG. Normalized spectral power of fetal heart rate variability is associated with fetal scalp blood pH. *Early Human Development*, 2011; 87: 259-63.
- Vapnik VN. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., 1995.
- Viglione S, Walsh G. Proceedings: epileptic seizure prediction. *Electroencephalogr Clin Neurophysiol*, 1975; 39: 435-6.

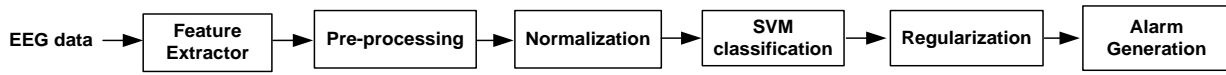


Figure 1. Block diagram for employed SVM seizure predictor

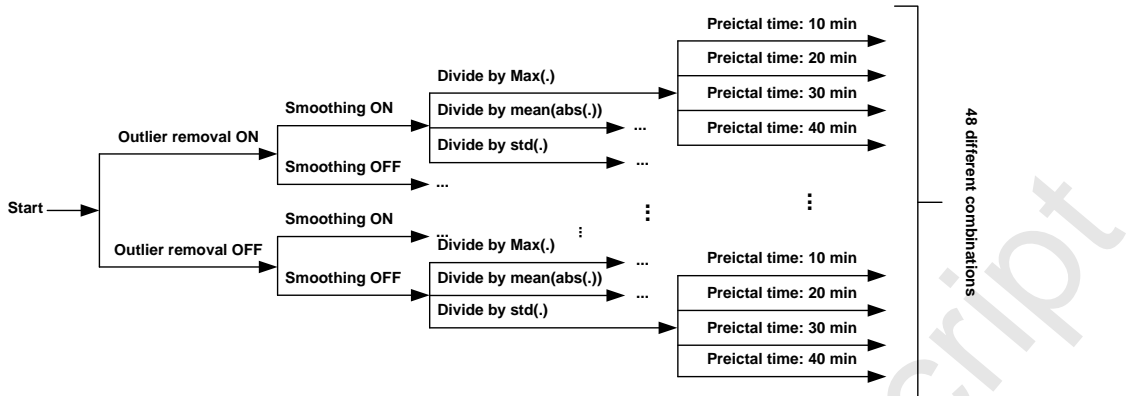


Figure 2. The flowchart of the possible combinations of the methods to be compared

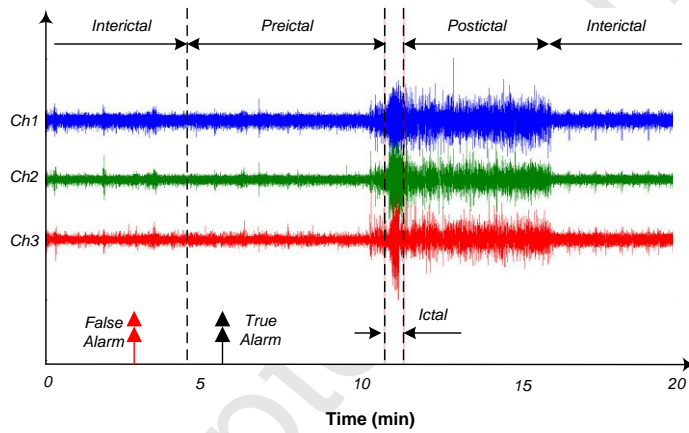


Figure 3. Epileptic brain states

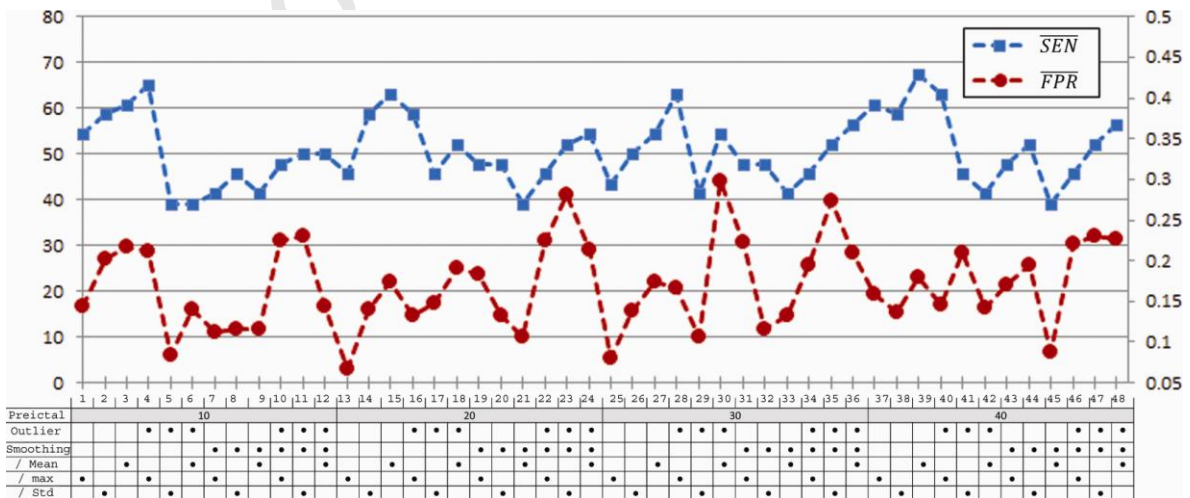


Figure 4. Average of SS and FPR for different combinations

Highlights

- 22 linear univariate features are combined in one feature space.
- SVM classifier is used to discriminate the feature space to pre/non-pre ictal states.
- The effects of preprocessing and normalization of the features were studied.
- Smoothing combined with normalizing by max of the feature can produce better results.
- Sensitivity of 73.9% with false prediction rate of 0.15h⁻¹ was achieved in average.

Accepted Manuscript