Abstract: Digital signal processing of Electroencephalogram (EEG) can support the diagnosis and alarming for the benefit of humans. About one third of all epileptic patients suffer from refractory epilepsy; seizure prediction based on the EEG information content is an area of intense activity since at least twenty years. In this paper we analyze the high dimensional feature space created by a variety of feature extraction methods for prediction of epileptic seizures. We combined features selection algorithm minimum redundancy maximum relevance (mRMR) and Support Vector Machines (SVMs) architectures to study the best features set for seizure prediction. We present the comparison between the classification results obtained by a feature set composed by 147 features and a reduced set based on the first 20-ranked features using mRMR scores. We critically discuss the composition of the feature subset. The results suggest some patient specificity in features and channel selection. The best models lead us to hypothesize the preference for wider preictal periods.

Keywords: Classification Methods; Classification; Feature Selection; Seizure Prediction; Epilepsy.

1. INTRODUCTION

Around 50 million people worldwide suffer from epilepsy according to the World Health Organization. It is estimated that 30% of the cases do not have pharmacological or surgical treatment; these patients must live with seizures that may occur anytime, anywhere. The research to find sufficient information in EEG to develop an automatic alarming system is being addressed by a large community of researchers from medicine, engineering and computer science.

The problem can be formulated as a pattern recognition one. Several features based on multichannel EEG (time domain, frequency domain, etc.), build up the high dimensional feature space in which one aims to distinguish the several brain states: interictal (normal state between two seizures), pre-ictal (before the seizure), ictal (the seizure), and post-ictal (transition to normal state). Despite of the large effort done by the research community during the last decades, few significative progresses have really been done (Mormann et al., 2007). Evidences that a 'preseizure' state exists, however no conclusion has been made about
the duration of this period or the features that describe it (Sackellares, 2008).

Most of the researchers agree today that multi-features, multi-channel, analyses are needed to create an information space where eventually a good predictor can produce effective alarms before the seizure (D’Alessandro et al., 2005), allowing the patient to improve his safety, sociability and privacy. The success in seizure prediction would also be a fundamental step towards the development of closed-loop therapeutic devices, which would trigger a clinical or behavioral intervention, e.g. drug injection or electrical stimulation, after the prediction of an upcoming seizure (D’Alessandro et al., 2003).

The use of multiple electrodes and their different combination gives a spatial perspective of the brains dynamics. A multiple features space allows the description of different time series domains. The feature space computed is then classified using Support Vector Machines (SVMs). SVMs are computational intelligence classifiers considered as having a high potential for seizure prediction (Mirowski et al., 2008).

A possible limitation in large scale classification problems is known as the curse of dimensionality, related to high dimensional feature spaces which can disturb the performance of machine learning algorithms. It is possible to overcome these limitations using feature selection and reduction methods. These present various advantages, e.g. the reduction of computational costs, and improvement of the performance of the classifiers (Guyon and Elisseeff, 2003). The condition imposed in feature selection is the optimal characterization of the dataset (Peng et al., 2005) which in classification studies closely relates to the minimal classification error. Two approaches are usually considered, filters and wrappers. Computationally less demanding, filter approaches evaluate features sets based on statistical dependency. The redundancy of the subset of features selected can be a problem. The method of minimum redundancy-maximum relevance (mRMR) (Ding and Peng, 2005) overcomes this limitation and is especially well-suited to the SVM architectures used in this context. In the approach presented, the subset of features selected using mRMR allow to hypothesize about the relevance of each feature, channel and their specificity to each patient.

This paper is presented as follows. Section 2 presents the methods used in the study. Section 3 presents an overview of the results patient-by-patient. Finally, section 4 contains a brief discussion and the conclusions.

2. METHODS

The objective of this work was to classify a high-dimensional feature space extracted from long-term EEG recordings of epileptic patients, and discuss the performance of features reduction methods in seizure prediction studies. The reduced feature space was analyzed as well, identifying variations between the different datasets computed.

2.1 The Data

The data represent long-term EEG recordings of 6 patients affected by intractable epilepsy. The data were acquired using 27 scalp electrodes according to 10-20 system. The seizures and epileptiform activity were evaluated by qualified clinical professionals. Table 1 presents an overview of the data used in the study.

<table>
<thead>
<tr>
<th>Table 1. Overview of the data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient</td>
</tr>
<tr>
<td>A</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>D</td>
</tr>
<tr>
<td>E</td>
</tr>
<tr>
<td>F</td>
</tr>
</tbody>
</table>

2.2 Multi-channel, multi-feature analysis

Each patient was analyzed using several different training sets, each with a specific combination of electrodes and preictal period. Due to the uncertainty around the preictal period four different values (10, 20, 30 and 40 minutes) were considered. These predictive horizons represent a reasonable mediation period for a clinical or behavioral intervention.

Another important variable in seizure prediction studies is the selection of the electrodes sites considered for classification. We used three combinations of six electrodes to analyze variability of the electrode selection in seizure prediction. The first combination consists of three focal electrodes, i.e., electrodes localized in the region of the seizure onset, and the remaining three electrodes far apart from the origin region. The second display represents the ‘discretization’ of the 10-20 system, that is, electrodes sites selected across the scalp - frontal, temporal and parietal bilateral electrodes. The third array is based on six electrodes randomly selected.

2.3 Feature Extraction

The inputs to machine learning algorithms are features which can be univariate, representing the analysis of a single channel, or multivariate, describing the relation between two or more channels. EEG feature extraction is performed based on a sliding-window approach. The size of the window should be long enough to capture temporal patterns of the signal, while considering the assumption of stationarity of the time series. The multi-feature approach used in this study creates a set of 147 features each 5 seconds. Table 2 summarizes the features extracted.

2.4 Training and Testing Datasets

The selection of training and testing patterns is an important subject in classification problems specially dealing with long term, unbalanced data. The target was designed using four different classes: interictal, preictal, ictal and post-ictal periods (the post-ictal period was defined as 10
First zero-crossing of the autocorrelation

Signal power in different frequency bands:

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Energy</td>
<td>Signal energy</td>
</tr>
<tr>
<td>Autoregressive Model Error</td>
<td>Forwards AutoRegressive predictive error</td>
</tr>
<tr>
<td>Decorrelation Time</td>
<td>First zero-crossing of the autocorrelation sequence</td>
</tr>
<tr>
<td>Relative Power</td>
<td>Signal power in different frequency bands: Delta band (0.1 - 4Hz); Theta band (4 - 8Hz); Alpha band (8 - 15Hz); Beta band (15 - 30Hz); Gamma band (30 - 2000Hz)</td>
</tr>
<tr>
<td>Spectral Edge</td>
<td>Frequency and Power, describing the minimum frequency and related power up to which 50% of the spectral power up to 40Hz is contained in the signal</td>
</tr>
<tr>
<td>Statistical Moments</td>
<td>Statistical analysis of the EEG signal Mean, Variance, Skewness and Kurtosis</td>
</tr>
<tr>
<td>hjorth</td>
<td>Complexity and Mobility</td>
</tr>
<tr>
<td>Wavelet Transform Coefficients</td>
<td>6 bands of decomposition of the EEG signal; power of time series described by Wavelet coefficients</td>
</tr>
<tr>
<td>Mean Phase Coherence</td>
<td>Measure of phase synchronization between two channels</td>
</tr>
</tbody>
</table>

Table 2. Overview of the features extracted

The classification of a high dimensional feature space is a difficult task that can be improved using feature selection techniques; generally we consider two approaches: filters and wrappers. The key point is the selection of a subset of features by eliminating features with low or no predictive power.

In summary, wrapper methods are based on prediction methods (e.g. SVM), and estimate the importance of the features through the evaluation of prediction algorithms. The accuracy of predictors using different input sets is used to choose the most informative set of features. Filter methods quantify a certain parameter of relevance between each feature and the target. This parameter, e.g. mutual information and statistical tests, allows selecting the features that better describe the data (a selection of top-ranked features is usually defined). However, the subset selected can present correlation between features (Peng et al., 2005).

The minimum-redundancy maximum-relevance (mRMR) algorithm presented by (Ding and Peng, 2005), ranks a set of features minimizing the redundancy among the subset of features while maximizing the relevance of the features. The idea is to minimize the redundancy of a subset selected according to a metric of relevance. The first step of mRMR algorithm is based on an F-test, as a relevance measure, and computation of the Pearson’s correlation among features as a redundancy measure. After selecting the first feature, i.e., the feature with maximum value of relevance (F-test ranked set) with the target, the remaining set of features is iteratively selected based on the mRMRscore (1). In this work, the approach used was F-test correlation difference (FCD), (Ding and Peng, 2005).

\[
mRMRscore = \max_{i \in IS} \left( F(i, s) - \frac{1}{|S|} \sum_{j \in S} |c(i, j)| \right) \tag{1}
\]

In this work we selected a subset based on the 20 best-ranked features. The size fo the subset considered two main concerns, identification of the importance of the electrode sites and feature concepts.

2.6 Classification

The classification of the reduced feature space is performed using SVMs. It is a margin classifier that draws a hyperplane in the feature space defining a decision boundary between samples of different target classes. The use of kernels to construct nonlinear decision boundaries is one of the main advantages of this algorithm. Consider a n-dimensional features space and \((x_i, c_i)\), where \(x_i\) is the \(i\)-th \(n\)-dimensional feature vector and \(c_i\) its class; the goal is to find the maximum-margin hyperplane that divides the points according to their class. Any \(n\)-dimensional hyperplane can be written as the set of points \(x\) satisfying (2):

\[
w . x + b = 0 \tag{2}
\]

where \(\cdot\) denotes the dot product. The weight vector \(w\) is perpendicular to the hyperplane. The parameter \(|w|\)
Fig. 2. Comparison between the best classification results obtained using the complete set of features and the ones obtained with the optimized subset of features determines the offset of the hyperplane from the origin. The purpose is to compute $w$ and $b$ to maximize the margin between the hyperplanes that separate the data into the different classes. Mathematically this can be presented by (3):

$$\min P(w, b) = \frac{1}{2} ||w||^2 + C \sum_i H[y_i, f(x_i)]$$ (3)

The first term represents the maximization of the margin while the second corresponds to a minimization of the training error.

In this study we used the MATLAB implementation of libSVM library (Chang and Lin, 2001). To optimize the $C$ and $\gamma$ parameters we used a grid search on $C \in \{2^5, 2^7, \ldots, 2^{15}\}$ and $\gamma \in \{2^{-5}, 2^{2}, \ldots, 2^{10}\}$. Two different kernels were used in this study: radial basis function and polynomial function of third degree. The F-measure was used to evaluate the computational models. This measure represents the weighted harmonic mean of precision and recall. Mathematically it is defined by (4).

$$F = \frac{2 \cdot \text{precision} \cdot \text{recall}}{(\text{precision} + \text{recall})}$$ (4)

Figure 1 presents an overview of the methodology used in this study.

3. RESULTS

This section summarizes the results, presenting the feature subsets selected using the mRMR algorithm, feature type and channels, and the comparison between classification results using the reduced and the complete feature sets.

Each dataset processed was constituted by 147 Features extracted from 6 channels. The 20-features subset selected was analyzed considering the feature type and the channels selected.

Fig. 2 presents the comparison between the best models (we used the F-measure to evaluate the models) using the reduced set and the complete set of features. In summary, there is no significant difference between the datasets, except in patient C. A patient by patient discussion is presented.

Patient A: The best model using the complete set of features presented an F-measure value of 0.091 while the model with the reduced set performed slightly worse - F-measure of 0.076. The best model (reduced set) was based on a preictal period of 40 minutes and the channel array considered was the discretization of the 10-20 system. The most important features considered in the subset were the frequency sub-band relative power (all the 5 different sub-bands) and the four statistical moments. According to Fig. 9.a, the features based on electrode PZ appeared 10 times. Fig. 3 summarizes the subset of features.

Patient B: Using the complete set of features, the best model achieved an F-measure value of 0.070; using the optimized subset, the F-measure was 0.081. This model presents a preictal period of 30 minutes and channel configuration similar to patient A (10-20 system discretized). The subset highlights the frequency sub-bands theta, gamma and beta (see Fig. 4). Similarly, the features based on the EEG signal of electrode PZ were the most preeminent (see Fig. 9.b).

Patient C: The comparison between the classification results obtained using the subset and the complete set of features yielded a decrease in the F-measure; the classification results achieved an F-measure value of 0.218 while achieving only 0.081 with 20 features. This model was based on a preictal period of 40 minutes and considered a channel configuration with three focal electrodes and three outside this area. The bivariate measure, Mean Phase Coherence, was selected three times, while Relative Power
appears again as the main concept chosen. Features based on the electrode T3 appeared 15 times in the subset of features.

Patient D: The performance of the best models using the complete and reduced set of features is similar - 0.086 using all the features computed and 0.081 using the 20 feature subset. Different features appear in the subset; time-frequency analysis performed by Wavelet Coefficients appears four times (Fig. 5). Electrode T5 is the main contributor of information in this subset of features (Fig. 9.d). The model used had the same configuration as the previous patient (three electrodes located near the focus and three outside the focal area).

Patient E: The performance of the best models is quite similar (on the one hand, the complete set achieved a value of 0.092, on the other hand, using the feature subset the performance was 0.088). The Wavelet Transform Coefficients appear 8 times in the reduced set of features. Only five different concepts appear in this subset. The preictal duration in this model is 40 minutes and channel configuration based on three focal electrodes. Channel F7 appears 9 times, however a balance between the rest of the channel is noticeable (Fig. 9.e).

Patient F: Comparing the performance of the classifiers using the different sets we noticed that the results were similar (0.016 based on the reduced set and 0.017 using the complete set). Mean Phase Coherence, Wavelet Coefficients and Statistical moments are repeatedly selected in the reduced set of features. Channels F7 and C3 are the most important contributors in a random array of electrodes, i.e., the model selected is based in six randomly selected electrodes. The preictal period selected was 40 minutes.

4. CONCLUSION

In this paper we analyzed the features extracted from the EEG signals of six epileptic patients. We used a feature selection method based on mRMR algorithm and SVMs architectures to perform classification. The ability to find an optimal combination of feature-channel would represent an important step towards seizure prediction.

We compared the classification results using a reduced subset of the 20 best-ranked features and the dataset-composed by 147 features computed using three different distributions of 6 channels. The comparison shows that the classification results are very similar. The dimension reduction performed by feature selection improved the computational cost and allowed us an easier interpretation of the feature (evaluation of the subset). The reduction of computational cost is also important for real-time implementation (training and testing) of predictors.

Inferior classification results of the shorter preictal periods (the best models pointed to 30 and 40 minutes) suggest...
that the optimal period considering the optimized subset of features is around 40 minutes.

The mRMR based subsets computed, pointed to different combination of electrodes, without specific patterns, e.g. no significant match was found between focus location and the channels selected. The channels selected were not confined to any specific area which may indicate the importance of regions outside the ictal region for machine learning studies.

We found that, using this method, specific features are more important (Relative Power and Wavelet Coefficients in particular) than others and only one patient presented features from all electrodes of the configurations chosen in the study.

These findings suggest the importance of individual, patient specific, solutions; each patient requires a specific combination of features and electrodes. Another important aspect is related to the variability of computational models between patients.

Future developments will take into account the frequency band of the spectral measures; in fact, frequency-based features were among the most selected. Other concerns include the increase of the number of patient used in this work and the comparison of these findings with other feature selection methods, e.g. recursive feature elimination and genetic algorithms.

REFERENCES