33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Optimized Feature Subsets for Epileptic Seizure Prediction Studies

Bruno Direito, Francisco Ventura, César Teixeira, and António Dourado, *Member IEEE*

Abstract— **The reduction of the number of EEG features to give as inputs to epilepsy seizure predictors is a needed step towards the development of a transportable device for real-time warning. This paper presents a comparative study of three feature selection methods, based on Support Vector Machines. Minimum-Redundancy Maximum-Relevance, Recursive Feature Elimination, Genetic Algorithms, show that, for three patients of the European Database on Epilepsy, the most important univariate features are related to spectral information and statistical moments.**

I. Introduction

Epilepsy is one of the most prevalent neurological disorders, and affects approximately 1% of the world's population [1]. This disorder is characterized by the occurrence of episodic abnormal cerebral electrical activity referred to as seizures. Advances in epileptic seizure prediction would represent a significant improvement in the daily life of refractory epilepsy patients and an important step towards the development of closed-loop therapeutic systems.

Despite the common agreement that a 'preseizure' state exists, i.e., that the transition from a normal ('inter-ictal') to seizure ('ictal') state occurs through a gradual transformation, few significant progresses have been made. During the last decades numerous features were proposed to characterize the EEG signals of the 'pre-ictal' period but the absence of reproducible results and statistical significance of the proposed measures have been a major obstacle in the development of clinical applications [2].

In recent years, the analysis of high-dimensional feature spaces [3], and the use of machine learning methods has been proposed [4]. Support Vector Machines (SVMs) is considered as a promising approach, with the advantage to create individually tailored solutions.

The analysis of high dimensional feature spaces in pattern recognition is usually constrained by the curse of dimensionality, which can disturb the performance of machine learning methods. To face these limitations it is important to select the appropriate subset of features [5].

A promising study was presented by [6][7], using a genetic algorithm and a high dimensional feature space to identify a patient specific optimized subset of features to compare the 'preseizure' and the 'no-preseizure' classes. The results however presented a high false prediction rate.

In the present work, we applied three different feature selection techniques to a high dimensional feature set to obtain an optimal predictor based on a subset of features selected from a candidate set of electrodes and quantitative features. A filter method, a wrapper method (both based on SVMs) and a genetic algorithm (to optimize the SVMs' parameters and the subset of features selected) were used to improve the classification performance and understanding about the feature sets.

The low-dimensional sets obtained were used as the input of SVMs to predict seizures in out-of-sample data, i.e., prospectively.

The paper is presented as follows. Section II presents the datasets used throughout the study and a brief description of the feature selection and machine learning methods used. A summary of the results is presented in section III. Finally, the concluding remarks are presented in section IV.

II. Methods

A. Dataset

The data used for this study consists in multichannel long-term EEG recordings obtained from 3 epilepsy patients. The patients were monitored during several days using scalp electrodes placed according to the 10-20 system. The seizures and epileptiform activities were annotated by a trained technician and reviewed by a neurologist. An overview of the data is presented in table I. The dataset is part of the database [8] developed by the EPILEPSIAE[1] project.

Each patient dataset was divided into three subsets: training, testing and validation sets. The training set is composed by the period of time containing three seizures.

This work was supported by the European Union FP7 Framework Program, EPILEPSIAE project (211713). BD would like to acknowledge the support from the Portuguese Foundation for Science and Technology (SFRH / BD / 47177 / 2008)

B. Direito, F. Ventura, C. Teixeira and A. Dourado are with the University of Coimbra, Science and Technology Faculty of the university of Coimbra, Pólo II, Coimbra, Portugal (e-mail: {bdireito, fventura, cteixei, dourado}@dei.uc.pt).

[1] EU FP7 Project EPILEPSIAE- Evolving Platform for Improving the Living Expectations of Patients Suffering from IctAl Events - http://www.epilepsiae.eu/

TABLE I
DATASET DESCRIPTION

Patient	Epileptic *focus*	Duration (hours)	Number of seizures	Number of channels
A	Fronto-temporal	78	12	31
B	frontal	138	11	24
C	frontal	252.6	9	27

The testing and validation sets, both contain at least one seizure.

Another important variable in seizure prediction is the considered pre-ictal period. This is especially relevant in machine learning since it is necessary to define a target. Two different periods were considered based on our previous experience: thirty and forty minutes before the seizure onset.

One of the main challenges in machine learning is to deal with unbalanced datasets. The inter-ictal period is by far longer than the relatively short pre-ictal, ictal and post-ictal periods. To deal with the biased datasets and to improve the computational performance of the algorithms, the inter-ictal class was under-sampled in the training set.

B. Feature Extraction

The features considered in this study were computed using EPILAB [9]. This software includes a variety of feature extraction methods and prediction algorithms that allow a rapid design and training of different algorithms using long-term EEG signals.

The features are based on a non-overlapping 5-second sliding window segments: signal energy, decorrelation time, relative spectral power, energy of the wavelet transform coefficients, spectral edge power and frequency, Hjorth mobility and complexity, mean, variance, skewness, kurtosis, and autoregressive modeling error. The complete set includes 22 features per electrode.

C. Feature Selection Methods

Feature selection methods identify a subset of univariate features useful to optimize the seizure predictor and ranking all potentially relevant feature-electrode combinations.

Three techniques have been applied (an overview of the methods is presented in Fig. 1).

1) Minimum Redundancy Maximum Relevance (mRMR)

The mRMR algorithm implemented [10], is an iterative procedure that ranks a set of features minimizing the redundancy (among the subset of features) while maximizing the relevance of the features.

The first step of mRMR algorithm is based on a statistical 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. In this work, the approach used to compute the coefficient was the F-test correlation difference (FCD) (eq.1)

$$mRMRscore = \max_{i \in \Omega S} \left\{ F(i,s) - \left(\frac{1}{|S|}\right) \sum_{j \in S} |c(i,j)| \right\} \qquad (1)$$

where $F(i,s)$ represents the relevance coefficient and the second term represents the redundancy between features. The first 132 ranked features were selected as inputs for the predictor.

2) Recursive Feature Elimination (RFE)

A different approach is used in the RFE-SVM algorithm described by [11]. RFE-SVM follows an iterative procedure based on the following three steps: (i) training a classifier (optimization of the SVM parameters and weights of each feature); (ii) compute the ranking criterion based on the weights computed (w_i^2) and (iii) removing the features with lowest ranking criterion.

Although SVMs can use non-linear kernels to solve complex decision boundaries, we limit our method to the linear kernel, because with this kernel it is possible to correlate the weight vector obtained in the support vectors computation with the importance of each feature to our problem.

3) Elitist Non-dominated Sorting Genetic Algorithm (NSGA-II)

Evolutionary Algorithms (EA) are inspired by biological evolution; each possible solution is represented by a coded "chromosome". Elitist Non-dominated Sorting Genetic Algorithm (NSGA-II) [12] represents a multi-objective EA based on non-dominated sorting approach.

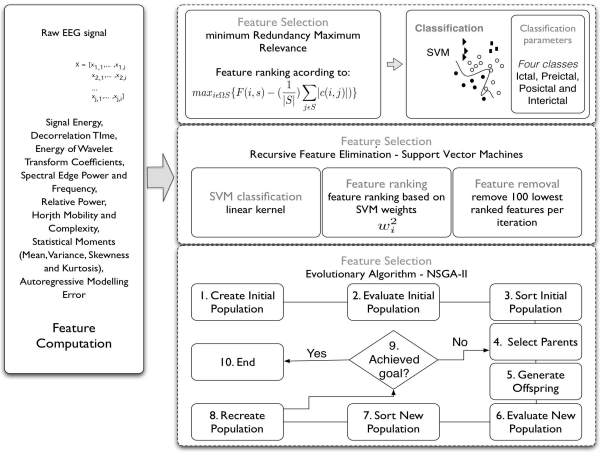

Figure 1 - Overview of the methods used in this work.

Using this approach, the features combined serve as inputs to the classifier (SVM) based selection process.

In the present study, NSGA-II has four different fitness functions: sensitivity, specificity, number of inputs and number of channels. In other words, the method tried to optimize the SVM parameters maximizing the sensitivity and specificity while minimizing the number of inputs and number of channels (Fig. 1). The input of the GA in our study corresponds to a binary string that codes the SVMs' parameters (C and γ) [13] and the subset of features.

The size of the features subset obtained using NSGA-II is variable and dependent on the algorithm stopping criterion, in this paper, the algorithm stops after 1000 iterations. The initial population considered 50 individuals.

D. Classification Strategy

SVM is a margin classifier that draws a hyperplane in the feature space defining a decision boundary between samples of different target classes. In this study we used the MATLAB interface of the libSVM library [13] and MATLAB's Parallel Computing Toolbox. The values accepted to optimize the C and γ parameters were comprised between 2^0 and 2^{16} and 2^{-16} and 2^{16} respectively.

III. RESULTS AND DISCUSSION

Table II summarizes the results of the best models obtained using the different methods. The unbalance between the number of samples of each class, lead us to consider the tradeoff between sensitivity and specificity. Therefore the best models were selected based on the overall ability to correctly classify pre-ictal samples while achieving good accuracy results.

Patient A - The best classification results were obtained using the subset computed using mRMR (Fig. 2 - a.1 and a.2). The statistical moments, i.e. variance, skewness, mean and kurtosis, appear to have some correlation to the target defined. Additionally, spectral features (especially the high frequency bands) also appear in the subset. Electrographical records suggest the importance of 'F9' 'FT9' 'T9' electrode sites in ictogenesis, however, our methods do not present any particular pattern concerning electrode selection (focal electrodes or laterality). The subset computed using RFE-SVM presents similar results. The best model obtained by NSGA-II presents very good results in the test dataset (sensitivity of 99,87%, specificity of 62,30% and an accuracy of 66,87%), but failed to obtain similar results using an out-of-sample, validation dataset. This suggests that NSGA-II algorithm was not able to find a model with a good generalization capability.

TABLE II
SUMMARY OF THE RESULTS

Patient	Method	Sensitivity	Specificity	Accuracy
A	mRMR	29,02%	73,02%	69,33%
	RFE	11,02%	88,83%	91,91%
	NSGA-II	0,27%	99,61%	95,78%
B	mRMR	30,51%	71,33%	67,45%
	RFE	40,00%	71,92%	71,68%
	NSGA-II	53,87%	61,00%	59,13%
C	mRMR	68,16%	75,97%	74,96%
	RFE	67,81%	82,90%	80,16%
	NSGA-II	89,54%	84,16%	80,63%

Patient B - The model that presented the best results achieved an accuracy of 71% (Fig. 2 b.1 and b.2), and was obtained using RFE-SVM. NSGA-II found a solution with 7 channels and 9 features, and emphasized the importance of spectral features (spectral edge power, in particular); the model presented a sensitivity of 63%, specificity of 70%, and 68% of accuracy, in the testing set. The performance in the validation dataset was slightly worse, the SVM presented a sensitivity result of 54%, specificity of 61%

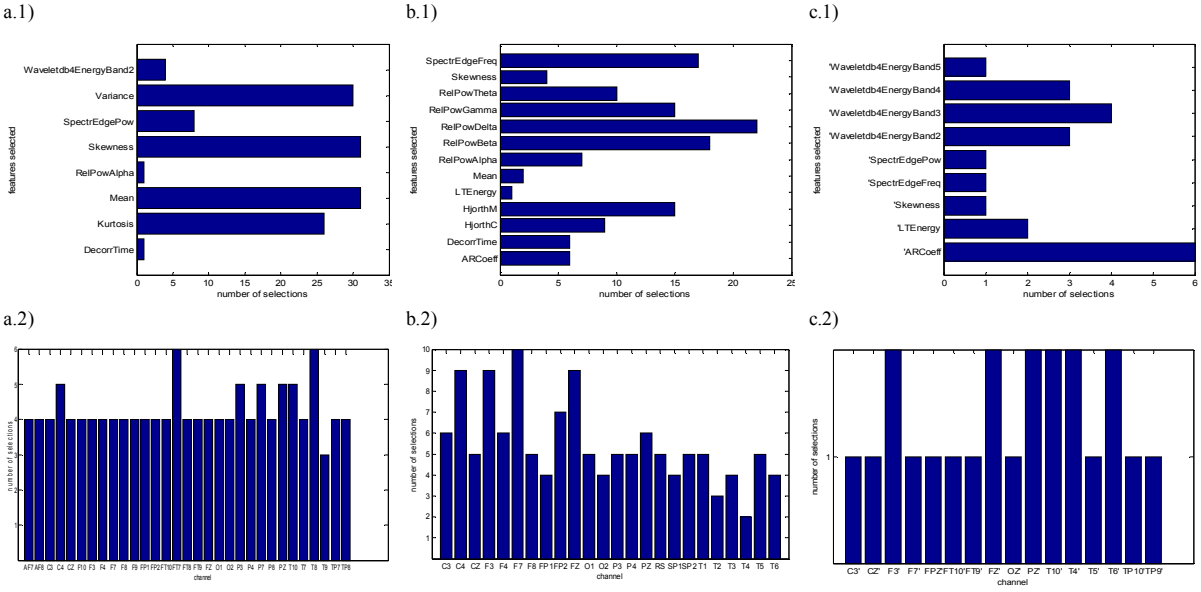

Figure 2 - Overview of the best models obtained for each patient. *a.1)* features selected in patient A dataset using mRMR *a.2)* channels selected in patient A dataset using mRMR *b.1)* features selected in patient B dataset using RFE-SVM *b.2)* channels selected in patient B dataset using RFE-SVM *c.1)* features selected in patient C dataset using NSGA-II *c.2)* channels selected in patient C dataset using NSGA-II

and accuracy of 60%. The classification results obtained using both mRMR and RFE-SVM are quite similar; the features selected highlight the importance of spectral information (especially low frequency bands of the wavelet coefficient analysis). No relevant pattern is noticeable in the electrode selection.

Patient C - For this patient, the best solution was computed using NSGA-II (a complete description is presented in Fig. 2 c.1 and c.2). The values of specificity, sensitivity and accuracy were all above 80%, with specificity near 90%, using only 22 features (notice that there is some emphasis in frontal-temporal electrode sites). The features selected suggest that spectral information is determinant for a good classification.

mRMR and RFE-SVM also achieved good classification results; the subset computed using RFE-SVM is based on spectral features, namely, relative power in the frequency sub-bands delta, theta, alpha, and the subset obtained using mRMR has important contribution from statistical moments and features that characterize low frequency sub-bands.

IV. Conclusion

In this paper we analyzed three different feature selection methods in seizure prediction studies. The objective was to present possible methodologies to identify optimal combination feature-electrode for seizure prediction.

The dimensional reduction performed using feature selection represented a significant improvement in the performance of the predictors; the reduction of the high computational cost associated to high dimensional feature spaces can also be an important asset for real-time implementation of the predictors.

The resulting subsets revealed specific patterns for each patient, confirming the need for individually tailored algorithms, and appropriate combination between electrodes and features.

The results obtained for Patient C, the best computed in this work, suggest that it is possible to select an optimal feature subset based on a reduced set of features and channels. The electrode selection was not confined to the focal electrodes indicating the importance of areas outside the ictal region. The features selected highlight the relevance of spectral information; different sub-band are represented in the subset. The autoregressive predictive error (ARcoeff.) is also selected in the best model.

Patients A and B did not present so encouraging results. Nevertheless, the best models achieved accuracies close to 70%. The analysis of the subsets allowed us to conclude that, similarly to patient C, electrode selection was not restricted to the ictal area.

The large number of electrodes that were selected in the feature subsets may represent an obstacle towards the development of clinical devices. Future work includes applying these methods restricting the number of channels.

NSGA-II computational requirements can represent an obstacle for the development of predictors; RFE-SVM and mRMR are, considering computational arguments, more viable solutions.

References

[1] B. Litt, R. Esteller, J. Echauz, M. D'Alessandro, R. Shor, T. Henry, P. Pennell, C. Epstein, R. Bakay, M. Dichter., and G. Vachtsevanos, "Epileptic seizures may begin hours in advance of clinical onset: a report of five patients.," *Neuron*, vol. 30, Apr. 2001, pp. 51-64.

[2] F. Mormann, R.G. Andrzejak, C.E. Elger, and K. Lehnertz, "Seizure prediction: the long and winding road.," *Brain : a journal of neurology*, vol. 130, Feb. 2007, pp. 314-33.

[3] P. Ataee, A. Yazdani, S. Setarehdan, and H.A. Noubari, "Manifold learning applied on EEG signal of the epileptic patients for detection of normal and pre-seizure States.," *Conference Proceedings of the International Conference of IEEE Engineering in Medicine and Biology Society*, vol. 2007, 2007, pp. 5489-5492.

[4] P.W. Mirowski, Y. LeCun, D. Madhavan, and R. Kuzniecky, "Comparing SVM and convolutional networks for epileptic seizure prediction from intracranial EEG," *2008 IEEE Workshop on Machine Learning for Signal Processing*, Oct. 2008, pp. 244-249.

[5] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, 2003, pp. 1157-1182.

[6] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echauz, and B. Litt, "Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients.," *IEEE Transactions on Biomedical Engineering*, vol. 50, 2003, pp. 603-615.

[7] M. D'Alessandro, G. Vachtsevanos, R. Esteller, J. Echauz, S. Cranstoun, G. Worrell, L. Parish, and B. Litt, "A multi-feature and multi-channel univariate selection process for seizure prediction.," *Clinical neurophysiology*, vol. 116, Mar. 2005, pp. 506-16.

[8] M. Ihle, H. Feldwitch-Drentrup, C.A. Teixeira, A. Witon, B. Schelter, J. Timmer, and A. Schulze-Bonhage, "EPILEPSIAE - A common database for research on seizure prediction," *Computer Methods and Programs in Biomedicine*, 2010.

[9] C.A. Teixeira, B. Direito, R.P. Costa, M. Valderrama, H. Feldwitch-Drentrup, S. Nikolopoulos, M. Le Van Quyen, B. Schelter, and A. Dourado, "A computational Environment for Long-Term Multi-Feature and Multi-Algorithm Seizure Prediction," *32nd Annual International IEEE EMBS Conference*, 2010.

[10] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data.," *Journal of bioinformatics and computational biology*, vol. 3, Apr. 2005, pp. 185-205.

[11] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene Selection for Cancer Classification using Support Vector Machines," *Machine Learning*, vol. 46, 2002, pp. 389-422.

[12] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Transactions on Evolutionary Computation*, vol. 6, 2002, pp. 182-197.

[13] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," *Computer*, 2001, pp. 1-30. G. O. Young, "Synthetic structure of industrial plastics (Book style with paper title and editor)," in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.