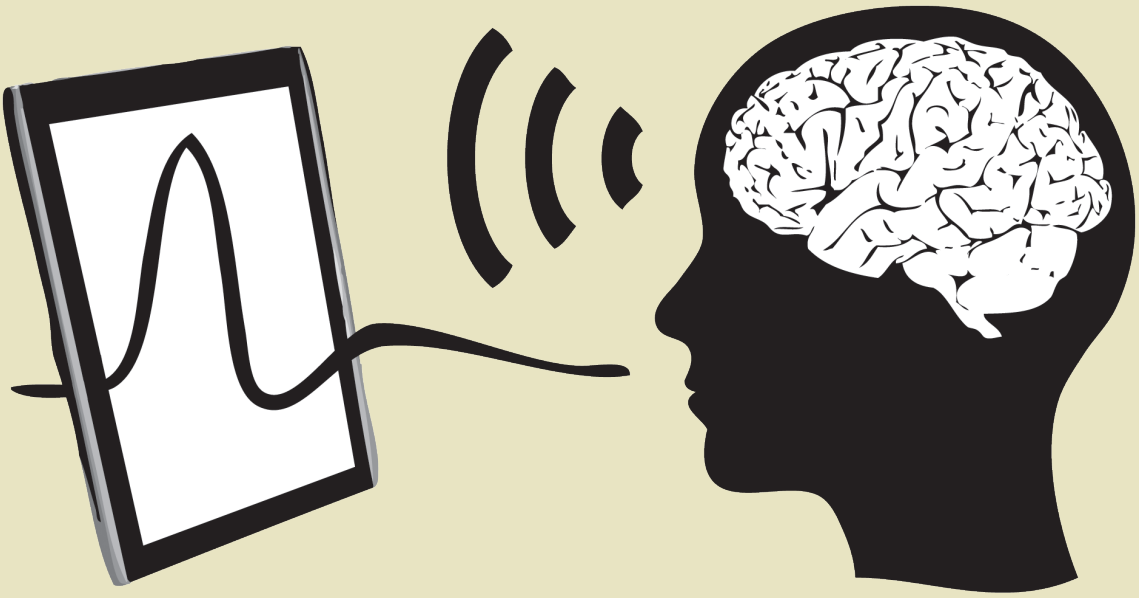


Bruno Miguel Direito Pereira Leitão

Development of classification methods for real-time seizure prediction

Dissertação de Doutoramento na área científica de Ciências e Tecnologias da Informação, especialidade de Inteligência Computacional em Informática Clínica, orientada pelo Professor Doutor António Dourado Correia, co-orientada pelo Professor Doutor Miguel Castelo-Branco, e apresentada ao Departamento de Engenharia Informática da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Setembro 2012

· U C ·

Universidade de Coimbra

This thesis is dedicated

to the memory of my Aunt and Friend,

Piedade Nunes

Abstract

In the last decades, the scientific community has made enormous efforts to understand the basic mechanisms underlying the generation of epileptic seizures. The analysis of the pre-ictal dynamics among different brain regions has been shown as an important source of information towards the understanding of the spatio-temporal mechanisms. This study, partially a contribution to the EPILEPSIAE project, aims the prediction of unforeseeable and uncontrollable epileptic seizures. Ultimately, the successful development of seizure prediction algorithms represents a fundamental step towards the creation of closed-loop intervention systems, which would improve the quality of life of epileptic patients.

The first part of this study aims the development of a patient-specific seizure prediction algorithm based on machine learning with high sensitivity and low false positives rate. The dynamical changes of the brain activity are analyzed using a high dimensional feature set obtained from both scalp and intracranial multichannel electroencephalogram (EEG). The features represent low complexity measures, implementable in real-time scenarios and the classification was performed using cost sensitive support vector machines (SVM). The proposed method was tested in 216 patients of the multicenter EPILEPSIAE database and presented statistical significant results for a small group of patients. We have also analyzed different optimization strategies such as feature selection, feature reduction (classical multidimensional scaling) and post-processing (moving average filter and Kalman filter) in order to improve the results.

We addressed the characterization of the EEG spatio-temporal patterns and the classification of specific brain states. The method proposed, based on the segmentation of topographic maps and on a statistical framework (hidden Markov models), shows promising results for the identification of a pre-ictal stage.

Lastly, we present a novel approach to characterize the pre-ictal period using multi-way models. Using the PARAFAC model, the EEG data is decomposed in rank-one tensors. It is hypothesized that one of the components represents

variations related to the pre-ictal period. Using the high-order data representation, we have also proposed a method to detect the variability of the data using incremental tensor analysis.

The conclusions of this study sustain the hypothesis that epileptic seizures (of a group of patient) are predictable. Concerning the methodologies proposed to analyze the space-time-frequency domain, we hope that the suggested approaches point towards new directions in the field of research of seizure prediction.

Keywords

Computational Intelligence, Classification, Feature extraction, Multiway models, Seizure prediction, Support Vector Machines, Topographic map

Resumo

Nas últimas décadas, a comunidade científica tem vindo a desenvolver grandes esforços no sentido de compreender os mecanismos básicos responsáveis pelo desencadeamento de crises epilépticas. A análise da dinâmica entre diferentes regiões do cérebro humano mostrou que esta pode ser uma importante fonte de informação para explicar a génese das crises.

Enquadrado no projeto europeu EPILEPSIAE, este estudo visa o desenvolvimento de um algoritmo de previsão de crises epilépticas cujo sucesso representaria um passo fundamental na criação de sistemas de intervenção malha fechada e uma melhoria da qualidade de vida dos doentes com epilepsia.

O trabalho de investigação permitiu, numa primeira fase, o desenvolvimento de um algoritmo específico para cada paciente baseado em métodos de aprendizagem computacional com elevada sensibilidade e baixa taxa de falsos alarmes. As variações da dinâmica cerebral de cada paciente são analisadas através de um conjunto de características calculadas através de electroencefalografia (EEG) multicanal de escalpe ou intracraniano. Estas características representam medidas de relativa pouca complexidade, facilmente implementáveis em tempo-real. A classificação é realizada através de máquinas de vectores de suporte com custo assimétrico. A metodologia proposta foi testada num conjunto de 216 pacientes pertencentes a diferentes centros de epilepsia (da base de dados europeia EPILEPSIAE) apresentando resultados de previsão estatisticamente significativos para um reduzido número de pacientes. Posteriormente analisamos o impacto de métodos de seleção de características, redução de características e estratégias de regularização da saída dos classificadores (aplicando os filtros de média móvel e de Kalman) tendo em vista a melhoria dos resultados obtidos.

A caracterização dos padrões espaço-temporais do sinal EEG e a classificação dos estados cerebrais relacionados com a epilepsia foi um dos procedimentos metodológicos usados. Baseado na segmentação dos mapas topográficos e no

modelo estatístico modelo escondido de Markov, o método proposto demonstra resultados promissores na identificação e caracterização do estado pré-ictal.

Por fim, apresentamos uma nova metodologia para a caracterização do período pré-ictal usando tensores de 3ª ordem. Usando o modelo de decomposição PARAFAC, as componentes espaço, tempo e frequência do sinal EEG são decompostas em bases de vectores no espaço tridimensional. Hipotetizamos que uma das bases tridimensionais resultantes representa as variações relacionadas com os processos pré-ictais. Usando a representação tridimensional dos dados (espaço, tempo e características do sinal EEG), propomos um método que permite identificar variações na estrutura de dados.

As conclusões deste estudo sustentam, a nosso ver, a hipótese de que as crises epilépticas (de um determinado número) de pacientes com epilepsia podem ser previstas. Relativamente às metodologias de análise espaço-temporais, expectamos que a abordagens metodológica apresentada aponte para novas direções no estudo de previsão de crises epiléticas.

Palavras-chave

Inteligência computacional, Classificação, Extração de características, Mapas topográficos, Máquinas de vetores de suporte, Previsão de crises, Tensores de 3ª ordem

Acknowledgements

This dissertation is the result of the research carried out at the Center for Informatics and Systems of the University of Coimbra (CISUC) in the last five years. When I started the doctoral program in September 2008, epileptic seizure prediction and electroencephalography signal processing was to a large extent unexplored territory in CISUC. Finishing this thesis has been one of the most challenging endeavors that I have ever faced and was only possible due to the contributions of many people.

First, I would like to thank my supervisor Professor António Dourado who supervised me during my M.Sc and Ph.D. Professor António Dourado efforts and personal commitment was one of the keys for the success of the European project EPILEPSIAE, and conclusion of this dissertation. His methodological contributions, fruitful discussions, and allowing me the freedom to follow personal directions were essential to carry out this study.

To my co-supervisor Prof. Miguel Castelo-Branco, I want to thank him for the fruitful discussions, motivation and especially for allowing me to share his enthusiasm for neurosciences.

I would like to thank the staff of the epilepsy unit of the Hospitals of the University of Coimbra (HUC) and especially Dr. Francisco Sales for the patience and support.

I also would like to express my gratitude to Ph.D. Inv. César Teixeira with whom I shared long days, victories and defeats throughout the last four years. His contributions are far beyond simple discussions about research results and new directions. Thanks also to M.Sc. Ricardo Couceiro for the shared discussions, and, most importantly, for the friendship.

I would like to thank the members of CISUC and IBILI, in particular my colleagues of the Adaptive Computing Group, Rui Costa, Francisco Ventura, Mojtaba Bandarabadi and João Duarte. I have to single out Prof. Bernardete Ribeiro for the fruitful discussions and kind words that helped me to keep going.

It would not make any sense to finish the dissertation without the support of my close family. All the frustrations, worries and anxieties can finally come to an end. Mãe, Pai, Rui, without you I wouldn't have made it.

I am also thankful to all my friends that accompanied me throughout this journey. I have to mention João Bártolo, João Castro and Hugo Tavares for their friendship from day one. To Rui Baltazar, Luís Reis, Gonçalo Ferrão, João Nabais, João Tiago and Tiago Morgado, we finally made it. To João Jorge, Dúlio Passos, Paulo Freitas, Fábio Fernandes, Luís Sousa and Carlos Melo and all my team mates, thank you for your encouragement and support.

I also would like to acknowledge CISUC and the Department of Informatics Engineering for the conditions and resources that allowed me to accomplish this research.

I would like to fully acknowledge the financial support during these four years of research of the *Fundação para a Ciência e Tecnologia* under the fellowship SFRH/BD/47177/2008 and EPILEPSIAE FP7 211713 Grantt.

The last words are dedicated to the love of my life, Joana. Thank you for your unconditional support and for always having the right words, the right smile.

Contents

List of Figures

List of Tables

List of Abbreviations

AED – Anti-epileptic drugs

BOLD – Blood oxygen level dependent

ECG - Electrocardiography

EEG – Electroencephalography

MDS – Multidimensional scaling

mRMR – Minimum redundancy maximum relevance

RFE – Recursive feature elimination

iEEG – Intracranial Electroencephalography

IPSP – Inhibitory postsynaptic potentials

EPSP – Excitatory postsynaptic potentials

fMRI – Functional magnetic resonance imaging

FFT – Fast Fourier transform

AR – Autoregressive

SPC – Seizure prediction characteristic

SOP – Seizure occurrence period

SPH – Seizure prediction horizon

IT – Intervention time

$FPR.h^{-1}$ *– False positive rate per hour*

FP – False positive

FPR_{max} *– maximum false prediction rate*

SVM – Support Vector Machine

IIR – Infinite impulse response

PCA – Principal component analysis

TP – True positive

FN – False Negative

KW – Kruskal-Wallis test

TN – True negative

POI – Point of interest

HMM – Hidden Markov Model

BCI – Brain-computer interface

ICA - Independent component analysis

NMF - Non-negative matrix factorization

PARAFAC – Parallel Factor Analysis

ALS – Alternating least squares

NTF - Non-negative tensor factorization

STF – Space-time-frequency

ITA – Incremental tensor analysis

DTA – Dynamic tensor analysis

NOTATION

Upper-case bold letters denote matrices, for example $\mathbf{X}$; lower-cases bold letters denote column vectors, for example $\mathbf{x}$. Indexing is presented using subscripting, as an example x_i denotes the *ith* element of $\mathbf{x}$ and X_{ij} represents the element in the *ith* row and *jth* column of $\mathbf{X}$.

$\mathbf{x}$	-	*Input series*
$\mathbf{x}'$	-	*First derivative of* $\mathbf{x}$
$\mathbf{x}''$	-	*Second derivative of* $\mathbf{x}$

N	-	*Number of samples*
μ	-	*Mean*
σ^2	-	*Variance*
χ	-	*Skewness*
κ	-	*Kurtosis*
E_a	-	*Averaged energy*
E_{ac}	-	*Accumulated energy*
α_{rp}	-	*Relative power in the alfa frequency band (8 – 15 Hz)*
β_{rp}	-	*Relative power in the beta frequency band (15 – 30 Hz)*
ϑ_{rp}	-	*Relative power in the theta frequency band (4 - 8 Hz)*
δ_{rp}	-	*Relative power in the delta frequency band (< 4 Hz)*
γ_{rp}	-	*Relative power in the gamma frequency band (>30 Hz)*
P	-	*Total (absolute) spectral power*
p_f	-	*Modulus-squared of the FFT coefficient at frequency* f
H_a	-	*Hjörth activity*
H_c	-	*Hjörth complexity*
H_m	-	*Hjörth mobility*

c_{xx} - *Autocorrelation function*

ar_i - *AR model parameters*

m - *Embedding dimension*

τ - *Time delay*

$\vec{s}$ - *Vector in the state space*

S - *Number of state space vectors*

Θ - *Heaviside step-function*

R - *Diameter of hypersphere*

C_s - *Correlation sum*

D_2 - *Correlation dimension*

d - *Distance between two vectors*

L - *Lyapunov exponent*

C_{rt} - *Cross-correlation sum*

ς - *Dynamical similarity index*

T_{pq} - *Dynamical entrainment between channels p and q*

ϕ_p - *Phase of the channel p*

M - *Mean phase coherence*

$P_{poisson}$ - *Probability of occurrence of an event at any sample*

P_{SOP} - *Probability of emitting an alarm within the SOP period*

P_{binom} - *Probability of predicting k out of K events by an unspecific predictor*

$P_{binom,z}$ - *Probability of predicting k out of K events by an unspecific predictor, considering z independent features*

α_p - *Significance level*

N_W - *Number of samples in a moving window*

N_{feat} - *Number of features*

$\mathbf{w}$ - *Weights (Support-vector machines)*

b - *Bias*

y_i - *Target or class at instant i*

N_P - *Number of patterns in the training set*

N_{SV} - *Number of support vectors*

$\Re^{N_{feat}}$ - *Feature space, N_{feat}-dimensional*

a_i - *Lagrangian multipliers*

ξ - *Slack variables*

C	-	*Soft margin cost parameter*
υ	-	*Mapping function between the original space and the feature space*
K	-	*Kernel function*
$\mathbf{D}$	-	*Dissimilarity matrix*
F_{β}	-	*F-measure*
N_T	-	*Number of patterns in the testing set*
cl_i	-	*Observation (i.e. cluster)* i
bs_i	-	*Hidden state* i
q_{ij}	-	*Transition probability*
$\mathbf{Q}$	-	*Matrix of transition probabilities*
e_{ij}	-	*Emission probability*
$\mathbf{E}$	-	*Matrix of emission probabilities*
$\underline{\mathbf{X}}$	-	*Tensor representation*
$\mathbf{X}_{(i)}$	-	*Matricization of the tensor* $\underline{\mathbf{X}}$ *along the* i*th mode or way*
N_d	-	*Number of modes or ways*
E_r	-	*Error of reconstruction*
λ	-	*Forgetting factor*
$\mathbf{U}_i$	-	*Factor matrices for the* i*th mode or way*
$\underline{\mathbf{G}}$	-	*Core tensor representation*

1. Introduction

1.1 Overview

Epilepsy is a brain disorder characterized by recurrent 'spontaneous' interruptions of normal brain function, also known as epileptic seizures (Fisher et al., 2005). Epilepsy has many possible presentations and causes, affecting the whole age range (from neonates to elderly people) (Duncan et al., 2006). In fact, epilepsy is not a single disease, but a variety of syndromes that share the predisposition to generate epileptic seizures. Different risk factors are associated to the disease such as head trauma, central nervous system infections, tumors, and cerebrovascular disease.

An epileptic seizure can be described as a transient clinical event occurring simultaneously with an abnormal neuronal activity in the brain. These clinical events range from subtle manifestations such as sensory abnormalities, to convulsions.

The prevalence of epilepsy is between four and ten per 1000 people per year (Duncan et al., 2006). In Portugal, epilepsy affects around 50,000 people (Ngugi et al., 2011). Although the majority of the patients face a positive prognosis, up to 30% continue to experience seizures despite therapy with anti-epileptic drugs (AED), also known as pharmaco-resistant epilepsy (Kwan and Brodie, 2000). Brain surgery may be adequate in order to reduce or eliminate the occurrence of seizures in about 8% of the cases (Duncan et al., 2006).

The apparent unpredictable nature of the epileptic seizures is one of the most devastating aspects of epilepsy and has a substantial effect on the health and social integration of these patients (Murray, 1993; Schulze-Bonhage et al., 2010). Medically intractable seizures limit the mobility of the patients and increase the possibility of severe injuries. The negative influence of the disease also affects close family members and friends that may experience anxiety and may need to adjust daily routines to manage with the patient's concerns.

The development of intervention devices, designed to reduce the number of seizures, abort seizures or simply to warn patients of upcoming seizures would

represent an enormous step towards the improvement of their quality of life (Morrell, 2006; Schulze-Bonhage et al., 2010).

Two main approaches have been considered for the development of seizure intervention systems: open-loop and closed-loop systems. Open-loop systems operate based on a predetermined schedule (also used for the treatment of tremor and Parkinson's disease), such as in deep brain stimulation therapy (Fisher et al., 2010). First trials provided stimulation via electrodes implanted in the patient's brain. On the opposite, closed-loop systems should intervene in order to reduce seizure likelihood, using AED, brain or vagus nerve stimulation, or to instruct a patient to take preventive measures (Shoeb et al., 2009; Fisher, 2012). The research of closed-loop therapeutic devices would greatly benefit from a reliable seizure prediction method, reducing the possibility of side effects related to the open-loop systems (Litt et al., 2003; Shoeb et al., 2011; Fisher, 2012).

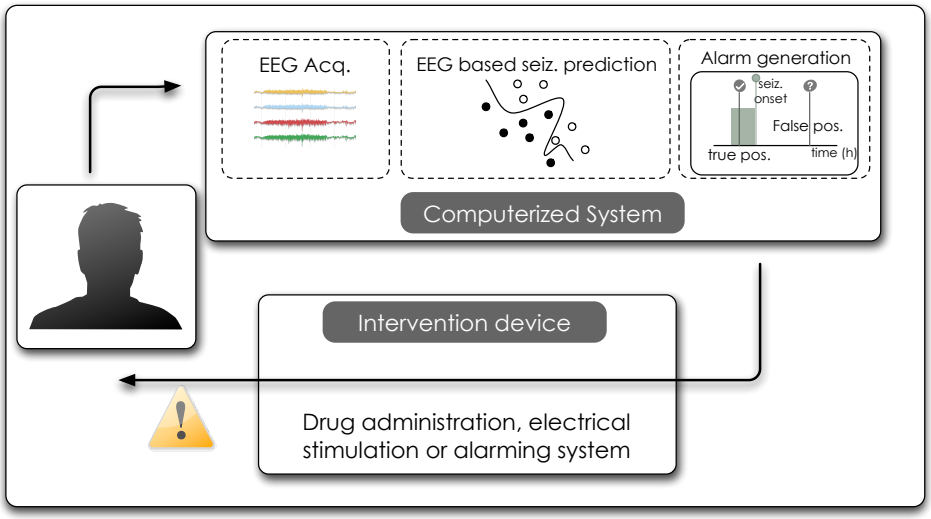


Figure 1.1 – Once seizure prediction performance presents reasonable sensitivity and specificity, then closed-loop intervention devices may play an important role in the improvement of the quality of life of epilepsy patients.

The existence of a 'pre-seizure' state is supported by different clinical findings. These include variations in cerebral blood flow, blood oxygen-level-dependent signal (BOLD) (Federico et al., 2005), and variations in some properties of the electrocardiogram (ECG) (Delamont et al., 1999; Novak et al., 1999) prior to the occurrence of seizures. Such evidences supported the hypothesis that the transition between the normal and seizure activity is gradual (or based on a cascade of changes), which could potentially be detected (Lopes da Silva et al., 2003).

At the cellular (and neuronal networks) level, epileptic manifestations are associated to an abnormal, excessive or synchronous neuronal activity. These findings motivated a number of studies based on connectivity concepts aimed at characterizing the electroencephalography (EEG) of epileptic patients and identify the transition state.

The history of the research of characteristic features for seizure prediction begun in the 1970s with the work of Viglione and colleagues (Viglione and Walsch, 1975). Despite the numerous efforts that have been made since, methodological statistical evaluation demonstrated that none of the proposed prediction algorithms performed above a random predictor (Mormann et al., 2007; Andrzejak et al., 2009; Stacey et al., 2011).

1.2 Main Contributions

The main contributions of this thesis consist in the research of the spatio-temporal dynamics underlying the transition from the inter-ictal (period between seizures) to the ictal phase (period related with the seizure activity). A multidimensional approach, based on a set of several univariate features, is used to characterize the so-called pre-ictal state.

Some of the key contributions that are given in this thesis are:

- Development of a seizure predictor, based on machine learning methods, in a quasi-prospective study. Using a high dimensional feature set based on multichannel EEG data, we have tested the patient-specific seizure predictor on long-term EEG data. This study is part of the first prospective assessment and evaluation of predictors' performance on a large, multi-center, database of patients. The collaborative European project called EPILEPSIAE (Evolving Platform for Improving the Living Expectations of Patients Suffering from IctAL Events)[1], developed a database with 275 pharmaco-resistant epileptic patients. One of the main objectives was to study the ability of machine learning methods to classify accurately pre-ictal patterns, using a low complexity measures for real-time implementation.

[1] http://www.epilepsiae.eu

- Development of feature selection and feature reduction methods in order to improve the predictor's performance.
- Development of post-processing strategies for effective alarm generation.
- A method to determine the spatio-temporal behavior of brain electrical activity, based on the topographic mapping of time varying relative power of different frequency sub-bands.
- Development of a computational framework based on multi-way models to determine their ability to characterize the epileptic pre-ictal period.
- Development of an iterative tensor analysis based on EEG data to determine spatio-temporal data structure variability in epilepsy patients.

In the body of this thesis each of the above topics is expanded and the new contributions are compared with the ones described in the literature.

1.3 Organization of the Thesis

The thesis is organized in six chapters.

In Chapter 2, we review background material on epileptic seizure prediction. We introduce the basic mechanisms of seizure generation, describe the electroencephalogram and present some examples of seizures. We present some of the most well-known studies in seizure prediction as well as statistical tools used to assess the significance of seizure prediction algorithms. We end the chapter by reviewing the current research in seizure intervention systems.

The third chapter is dedicated to the application of machine learning methods to a large set of 216 patients. We first present basic concepts on classification. In Section 3.2, the architecture of the proposed patient-specific seizure predictor is presented. Section 3.2.6 presents a summary of the results obtained using scalp EEG data, as well as the statistical evaluation. The same architecture is used to develop predictors using intracranial EEG (iEEG) data (Section 3.3).

Still in the third chapter, the possibility of using feature selection techniques to improve the classification results is analyzed. Two different strategies are used, minimum redundancy maximum relevance (mRMR) and recursive feature elimination (RFE). Additionally, the post-processing of the classification output using the Kalman filter and firing power method is also examined (Section 3.6).

These methods contribute to a lower complexity algorithm, a key feature for real-time implementation in portable devices.

We also discuss the existence of circadian rhythms and its relationship with the output of the predictors.

In Chapter 4, a method to identify the different states of the epileptic brain is presented. The methodology proposed is based on the topographic mapping of time varying relative power of delta, theta, alpha, beta and gamma frequency sub-bands, estimated from EEG (Section 4.3). The method reconstructs the trajectory of points of interest identified in the topographic mappings (Section 4.3.2). In Section 4.3.5, we introduce the statistical Markov framework used to model the states and the transition between states. The results and discussion are presented in Section 4.4 and Section 4.5, respectively.

In the fifth chapter, we focus on analyzing the ability of multi-way models (also known as tensors) to characterize the epileptic pre-ictal period. A space, time, and frequency representation of the data is proposed (Section 5.2). In Section 5.3, an incremental tensor analysis is proposed to identify data structure variability and temporal structure of the reconstruction error.

In Chapter 6, the main conclusions of this work are outlined, as well as some future work.

1.4 List of Publications

1.4.1 Book Chapter

[B.1] Direito, B., Costa, R., Feldwitch-Dentrup, H., Valderrama, M., Nikolopoulos, S., Schelter, B., Jachan, M., Teixeira, C. A., Aires, L., Timmer, J., Le Van Quyen, M. and Dourado, A. (2011), EPILAB: A MATLAB platform for multifeature and multialgorithm seizure prediction, *in*: Osorio, I., Zaveri, H. P., Frei, M. G., Arthurs, S., eds., 'Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics', CRC Press, Boca Raton, Florida, pp. 489-500

1.4.2 Articles in Journal

[J.1] Direito, B., Teixeira, C. A., Ribeiro, B., Castelo-Branco, M., Sales, F. and Dourado, A. (2012), 'Modeling epileptic brain states using EEG spectral analysis and topographic mapping', *Journal of Neuroscience Methods* 210, 220–229.

[J.2] Teixeira, C. A., Direito, B., Feldwitch-Drentrup, H., Valderrama, M., Costa, R. P., Alvarado-Rojas, C., Nikolopoulos, S., Le Van Quyen, M., Timmer, J., Schelter, B. and Dourado, A. (2011), 'EPILAB: A software package for studies on the prediction of epileptic seizures', *Journal of Neuroscience Methods* 200, 257–271.

1.4.3 Articles in Conference Proceedings

[C.1] Direito, B., Teixeira, C. A., Ribeiro, B., Castelo-branco, M. and Dourado, A. (2012), Space Time Frequency (STF) Code Tensor for the Characterization of the Epileptic Preictal Stage, *in*: 'Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society', pp. 621-624.

[C.2] Teixeira, C. A., Direito, B., Bandarabadi, M. and Dourado, A. (2012), Output regularization of SVM seizure predictors: Kalman Filter versus the "Firing Power" method, *in*: 'Proceedings of the 34th Annual International Conference of the IEEE Engineering in Medicine and Biology Society', pp. 6530-6533.

[C.3] Direito, B., Ventura, F., Teixeira, C. A. and Dourado, A. (2011), Optimized feature subsets for epileptic seizure prediction studies, *in*: 'Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society', pp. 1636–1639.

[C.4] Direito, B., Teixeira, C., Dourado, A., 2011. On the benefits of classical multidimensional scaling in Epileptic seizure prediction studies, *in*: 'Proceedings of the 1st Portuguese Biomedical Engineering Meeting', pp. 1-4.

[C.5] Direito, B., Duarte, J., Teixeira, C. A., Schelter, B., Le Van Quyen, M., Sales, F., Schulze-Bonhage, A. and Dourado, A. (2011), Feature selection in high dimensional EEG features spaces for epileptic seizure prediction, *in*: 'Proceedings of the 18th International Federation of Automatic Control World Congress', pp. 6206-6211.

[C.6] Teixeira, C. A., Direito, B., Costa, R. P., Valderrama, M., Feldwitch-Drentrup, H., Nikolopoulos, S., Le Van Quyen, M., Schelter, B. and Dourado, A. (2010), A computational Environment for Long-Term Multi-Feature and Multi-Algorithm Seizure Prediction, *in*: 'Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society', pp. 6341-6344.

[C.7] Ventura, A., Franco, J., Ramos, J., Direito, B. and Dourado, A. (2009), Epileptic Seizure Prediction and the Dimensionality Reduction Problem, *in*: 'Proceedings of the 19th International Conference on Artificial Neural Networks', pp. 1-9.

[C.8] Direito, B., Dourado, A., Vieira, M. and Sales, F. (2008), Combining Energy and Wavelet Transform for Epileptic Seizure Prediction in an Advanced Computational System, *in*: "Proceedings of the International Conference on Biomedical Engineering and Informatics", pp. 380–385.

[C.9] Direito, B., Dourado, A., Vieira, M. and Sales, F. (2008), Classification of Epileptic EEG Data Using Multidimensional Scaling, *in*: 'Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Engineering', pp. 551–555.

[C.10] Direito, B., Dourado, A., Sales, F., and Vieira, M. (2008), An Application for Electroencephalogram Mining for Epileptic Seizure Prediction, *in*: 'IEEE International Conference on Data Mining', pp. 87–101.

[C.11] Costa, R. P., Oliveira, P., Rodrigues, G., Direito, B. and Dourado, A. (2008), Epileptic Seizure Classification Using Neural Networks With 14 Features, *in*: 'Proceedings of the 12th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems', pp. 281-288.

[C.12] Dourado, A., Martins, R., Duarte, J., and Direito, B. (2008), Towards Personalized Neural Networks for Epileptic Seizure Prediction, *in*: 'Proceedings of the 18th International Conference on Artificial Neural Networks', pp. 479–487.

1.4.4 Posters in Conference Proceedings

[P.1] Direito, B., Teixeira, C. A., Bandarabadi, M. and Dourado, A. (2011), Modeling epileptic brain states using EEG spectral analysis and topographic mapping, *in*: '5th International Workshop on Seizure Prediction'.

[P.2] Direito, B., Dourado, A., Teixeira, C. A., Aires, L., Costa, R. P., Schelter, B. and Le Van Quyen, M. (2009). EPILAB: A Matlab® Platform for Multi-Feature and Multi-Algorithm Seizure Prediction Studies, *in*: '4th International Workshop on Seizure Prediction'.

[P.3] Teixeira, C. A., Costa, R. P., Direito, B., Sales, F. and Dourado, A. (2009), Low-complex TD-RBF and TD-SVM seizures predictors based on EEG energy and ECG entropy, *in*: '4th International Workshop on Seizure Prediction'.

[P.4] Direito, B., Martins, R., Costa, R. P., Dourado, A., Sales, F. and Vieira, M. (2008), Computational Intelligence Algorithms for Seizure Prediction, *in*: '8th European Congress of Epileptology'.

2. The Electroencephalogram and Epilepsy

This chapter is dedicated to the theoretical framework of the thesis. It is based on four sections. The first presents an overview of the neurophysiological and neuroanatomical concepts necessary to understand epilepsy and epileptic seizures generation. In the second section we introduce the electroencephalography (EEG) as a measure of electrophysiological activity of the brain. Basic notions of EEG and some standards are presented for an easier understanding of signal processing methods used in this thesis. In the same sense, the intracranial electroencephalography (iEEG) is also introduced. The third section presents a summary of the most noticeable studies on epileptic seizures predictability. A detailed description of the most relevant measures extracted from EEG is given, including their mathematical formulations. It is intended to give the reader a formal description of these methods along with the results and challenges documented by the authors. The weaknesses met by the scientific community are summarized, and frameworks to assess the performance of prediction algorithms are presented. The last section summarizes the existent seizure intervention devices and the two main approaches in their development: open-loop and closed-loop systems.

2.1 Epilepsy and Epileptic Seizures

Epilepsy is conceptualized as a neurologic condition characterized by an enduring predisposition to generate epileptic seizures (Fisher et al., 2005). Despite of the variability of the causes underlying epilepsy, its seizures are clinical manifestations caused by an abnormal and excessive synchronous activity of neuronal networks in the brain, particularly in the cerebral cortex. The clinical consequences, also known as ictal semiology, depend strongly on where the abnormal neural networks behaviors occur and the properties of the abnormalities.

2.1.1 Brain Structures and Basic Mechanisms of Seizure Generation

Large part of the brain is composed by the cerebrum, which is divided almost equally into two hemispheres, left and right. The outer portion of the cerebrum, the cerebral cortex, is a folded structure with a total surface area that ranges from 1600 to 4000 cm^2 (Nunez and Srinivasan, 2006). Cortical neurons are the elementary units of this structure and are intensely interconnected. Typically neurons possess a cell body, dendrites and an axon. The composition of the cerebral cortex is based on a high density of neuron cell bodies, dendrites and some axons that explain its grayish appearance, reason why the cerebral cortex is also known as gray matter.

Just below the gray matter is another type of tissue known as white matter. This region is composed primarily by nerve fibers (axons of the neurons). The numerous interconnections between different cortical regions existent in the white matter allow the communication between distant neuronal networks.

The connection between neurons is performed by specialized structures, the synapses. A single neuron may be covered by 10^4 synaptic connections that transmit inputs from (and to) other neurons (Nunez and Srinivasan, 2006). The synaptic inputs to a neuron can trigger two opposite effects:

- neurotransmitters that cause hyperpolarization of the neuron produce inhibitory postsynaptic potentials (IPSPs), making it harder to the target neuron to fire an action potential,
- neurotransmitters that lead to depolarization produce excitatory postsynaptic potentials (EPSPs), lowering the threshold for firing action potential.

The balance between EPSPs and IPSPs determines the firing of action potentials (Figure 2.1).

The excitability of individual neurons and neuronal networks is governed by different factors, and can be divided into those that act inside the neuron (neuronal or intrinsic) or in the cellular environment (extraneuronal or extrinsic). The former includes the distribution of voltage and ligand-gated channels (that determine the properties of the action potential), biochemical modification of receptors, activation of second-messenger systems and genetic abnormalities that may originate conformational alterations of ion channels. Extrinsic factors such as changes in

extracellular ion concentration, conformational alterations of the synaptic contacts, and changes in metabolism by glial cells also influence the excitability of neurons.

The network organization of neurons has a major impact in the individual excitability. Moreover, changes in the function of a single cell within a path or network can significantly affect both neighboring and distant neurons due to (local and broad) dense interconnections.

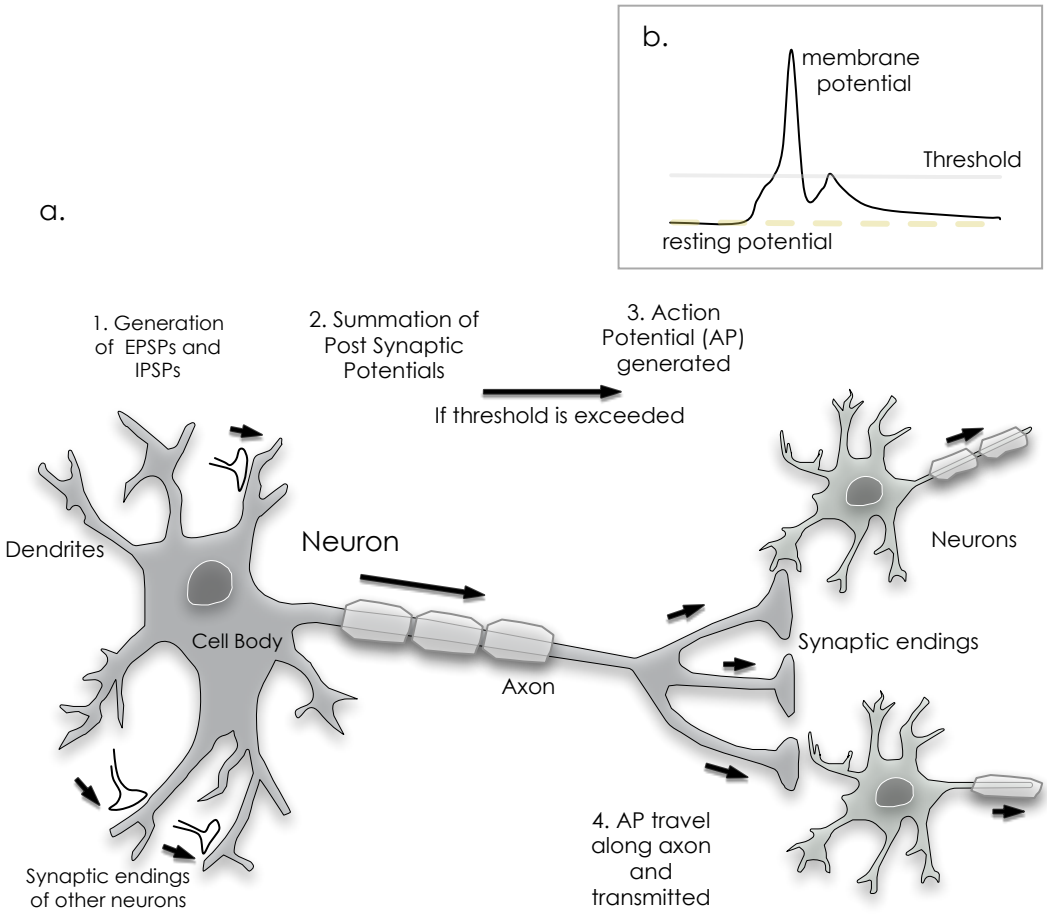


Figure 2.1 - Synaptic transmission of an electric signal (a). An action potential is generated if the summation of post synaptic potentials exceeds a threshold (b). The nerve impulses (action potentials) travel along the axon and the membrane depolarizes. The arrival of a nerve impulse at the terminus of the neuron signals the synaptic transmission, i.e., the release of the neurotransmitters at the synaptic cleft (Cooper, 2000).

Epilepsy can result from many different pathologic processes, however the underlying consequence involves an upset in the balance between excitation and inhibition. The hypersynchronous discharges that occur during a seizure may begin in a very small region and then spread to neighboring regions. This synchronous activity from a sufficient number of neurons results in a so-called spike discharge

on the EEG. The propagation of this pattern occurs due to the recruitment of surrounding neurons leading to a spread of the seizure activity into contiguous areas via local connections, and to more distant areas via long association pathways.

2.1.2 Epileptic seizures

Seizures are categorized according to their location and spread. If the seizure has a generalized onset, simultaneously involving widespread cortical regions over both hemispheres and subcortical regions, it is identified as a *generalized seizure*, while if the seizure presents a localized onset arising in a limited brain region, cortical or subcortical, with variable degree of spread, it is known as a *focal seizure* (Engel, 2006).

The clinical manifestations of focal seizures are usually associated to the region of the brain where the seizure happened. A focal seizure originating in the temporal lobe may involve feelings of euphoria, fear, and déjà vu, due to the presence of cerebral areas related to the processing of memories and emotional reactions (hippocampus and amygdala). These seizures usually start with a small region in one hemisphere, but may become more complex as they spread to other regions worsening the clinical manifestation. In the worst case, they may become generalized (also known as *secondarily generalized seizures*).

Generalized seizures do not have a recognizable focus onset; abnormal electrical activity appears to simultaneously affect different regions of the brain. Clinical manifestations of generalized seizures usually include loss of consciousness and motor disturbances. These seizures may be hypomotor, i.e. with decreased motor activity (e.g. absence seizures), may present whole-body rigidity and abnormal vigorous jerking (e.g. tonic-clonic seizures), or include loss of muscle tone (e.g. atonic) (Engel, 2006).

Li and colleagues (Li et al., 2002) presented a method based on the quantification of movement patterns to assist clinicians in seizure classification. This distinction is important for guiding diagnostic evaluation and prescribing appropriate therapy.

2.1.3 Epileptogenic zone

Surgical resection of brain tissue may be considered as an alternative for pharmaco-resistant epilepsy patients. Pre-surgical planning of the surgery is useful

to review some general definitions in epilepsy. Six different areas are considered during the pre-surgical evaluation (Carreño and Lüders, 2001).

The *irritative zone* is considered the area of the cerebral cortex that generates inter-ictal spikes. The *seizure-onset zone* is the area of the cortex from which clinical seizure are initiated. The *symptomatogenic zone* is the area of cortex, which generates the initial seizure symptoms. A structural lesion that is related to the generation of seizure is identified as *epileptogenic lesion*. The *functional deficit zone* is the area of the cortex with abnormal functioning in the inter-ictal period. Finally, pre-surgical evaluation also involves the analysis of the *eloquent cortex* (area of the cortex necessary for the normal behavior of cortical functions such as motor activities or speech), to minimize the side effects of the surgical resection.

Lüders et al. (2006) proposed a new definition for *epileptogenic zone*, "the minimum amount of cortex that must be resected (inactivated or completely disconnected) to produce seizure freedom", that was previously defined as "the minimum area of the cortex necessary and sufficient for initiating seizures".

2.2 Electroencephalography

The signal recorded with the EEG exam is a representation of voltage fluctuations obtained from electric potentials of summated excitatory and inhibitory postsynaptic potentials. Synchronous activity involving a large number of neurons (six cm^2 of cortex) is considered necessary for detection with scalp EEG. The parallel arrangement of pyramidal neurons in the cortex (that usually display similar orientation and polarity) and the summation of their postsynaptic potentials give rise to a 'detectable' cortical dipole layer that dominates the EEG signal.

The potentials recorded with EEG can be described in terms of its spatial distribution (frontal, posterior, lateral and bilateral) and spectral properties. Regarding the spectral properties of the signal, EEG potentials can be classified into two types: oscillations and transients.

The sequence of potentials recorded can be normal (and categorized according to the frequency range into delta, theta, alpha, beta, and gamma bands) or abnormal (seizures). The classical approach suggests the classification of an EEG wave into five frequency sub-bands: delta (0.5 - 4 Hz), theta (4 - 7.5 Hz),

Alpha (8 – 13 Hz), Beta (13 – 30 Hz) and Gamma (> 30 Hz) (Nunez and Srinivasan, 2006).

EEG transients are usually less interesting, although they are important for clinical diagnoses purposes. Abnormal EEG sharp transients often lead to diagnosis of epilepsy and are referred to as inter-ictal epileptiform potentials (Lopes da Silva, 2008). *Spikes*, *polyspikes*, *spike and wave complex*, *sharp waves* and *sharp and slow waves discharges* are known transients related to epilepsy with particular morphological characteristics.

However, transients due to artifacts from non-cerebral electric potentials may sometimes confound the analysis. Possible artifacts include eye blinking, cardiac impulses and muscle activity. EEG recordings also contain physical artifacts. For instance, scalp EEG signal is acquired by sensors placed in the scalp of the patient. Sudden movements or the detachment of electrodes may cause abrupt variations in the recordings. Interferences due to the power line are also quite common.

In epilepsy studies, the analysis of EEG recordings focuses on (especially in pre-surgical evaluation) two cortical areas: the *irritative zone* (area of the cortex that generates inter-ictal spikes), and *seizure onset zone* (area where clinical seizure are initiated) (Lopes da Silva, 2008).

2.2.1 Scalp Electroencephalogram

The scalp EEG is a non-invasive measure of the electrical potentials generated in the brain based on the placement of recording electrodes on the scalp. It provides a large-scale and robust measure of dynamic behavior of the brain function (Nunez and Srinivasan, 2006).

There is a long-established convention for the placement of the electrodes for routine scalp EEG recordings known as *international 10-20 system* (Schomer and Lopes da Silva, 2010). According to this convention the electrodes on the left hemisphere are assigned with odd numbers, electrodes on the right assigned with even numbers, and midline electrodes (electrodes located in the line defined by nasion, vertex and inion) assigned with suffix "Z". The prefix of the designation of the electrodes is defined according to the region where the electrodes are placed. For example, electrodes in the frontal region have the prefix "F", while "C", "T", "P", and "O" indicate central, temporal, parietal and occipital regions respectively (Figure 2.2).

In addition to the 21 electrodes defined by the 10-20 system, intermediate 10% electrodes positions may also be used.

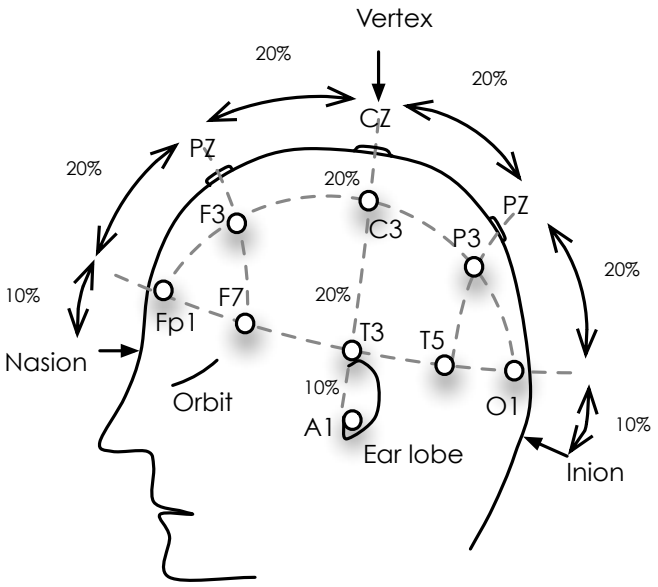


Figure 2.2 – EEG electrode placement according to the international 10-20 system.

EEG recordings are displayed, for an easier interpretation, organized into channels. These may be arranged into bipolar, referential (or monopolar), or average montages. In the first method, the difference between two adjacent electrodes is measured; in the referential method, the potential of each electrode is compared to a certain electrode considered as reference and the last (average montages) consists in the difference between each electrode and the average of all electrodes. Consider, for instance, a bipolar montage. The Fp1-F7 EEG channel measures the voltage differences within the frontal lobe and left hemisphere, summarizing the neuronal activity occurring in that region of the brain.

The EEG of a patient with epilepsy can be broadly classified as ictal and inter-ictal. Seizures, also known as the ictal period, usually present a sudden redistribution of spectral energy on a set of EEG channels (Grewal and Gotman, 2005). This phenomenon is caused by the abnormal hypersynchronous activity of neural networks. The spectral and amplitude components that present these alterations usually vary among patients. The set of channels affected by this variation also changes according to the cerebral site where the abnormal behavior occurs.

As an example, Figure 2.3 illustrates a seizure. The seizure, classified as complex partial (focal seizure with loss of consciousness), begins at second 6973 and consists on rhythmic theta waves.

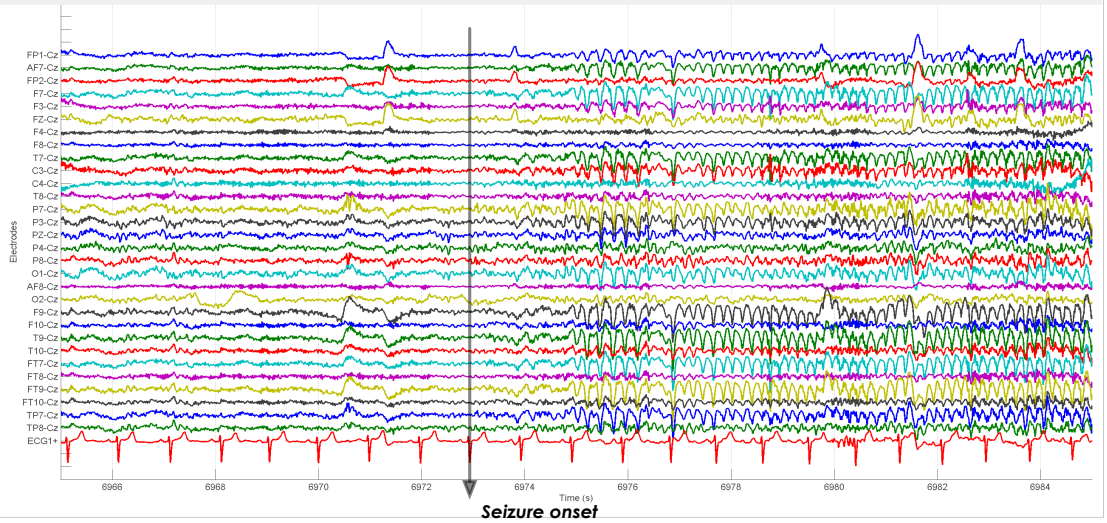


Figure 2.3 - Example of a seizure within the scalp EEG. The EEG signal is presented with a referential montage with Cz reference, high pass filter set at 0.5 Hz and a low pass filter set at 40 Hz. The seizure begins at 6973 seconds and involves rhythmic theta waves.

The period between seizures, also known as inter-ictal, has served as the cornerstone of epilepsy evaluation. The assessment of inter-ictal EEG may exhibit abnormal rhythmic activity (or discharges), which can confirm a diagnosis of epilepsy. The need to actually record a seizure is governed by whether it is of extreme importance to identify specific properties (particularly for planning epilepsy surgery). The discharges, which usually involve a set of channels, are called "epileptiform" features. *Spike* and *spike-wave complex* may be recorded even if the patient does not present any clinical symptom.

However, due to the anatomy of the brain, distribution and orientation of the neuronal networks, and skull, the electrophysiological sources may suffer distortions and seriously affect the recordable scalp EEG signals. This effect (also known as the *volume conduction effect*) transforms the signal and, in certain cases, the signal recorded is a composition of different (extremely mixed) sources (Nunez and Srinivasan, 2006). Different strategies were proposed to solve the mixing problem but lie outside the scope of this thesis.

2.2.2 Intracranial Electroencephalogram

Intracranial recordings are able to obtain local information with more detail. Similarly to scalp EEG, intracranial EEG (iEEG) provides a spatial and temporal picture of the electrical sources present in the brain, however the signals represent smaller areas. The electrodes are placed directly on the brain cortex or on deep structures of the brain (stereoelectroencepalography or sEEG) and record the summation of smaller neural networks, therefore avoiding some of the artifacts that contaminate scalp EEG (Figure 2.4). For example, the space averaging due to head volume conduction and source mixture imposed by the skull is, in part, avoided using iEEG.

The implanted intracranial electrodes provide a sparse spatial coverage (it is only possible to implant a limited number of electrodes). In practice, intracranial recordings provide different type of analysis compared to the one obtained using scalp recordings (Nunez and Srinivasan, 2006).

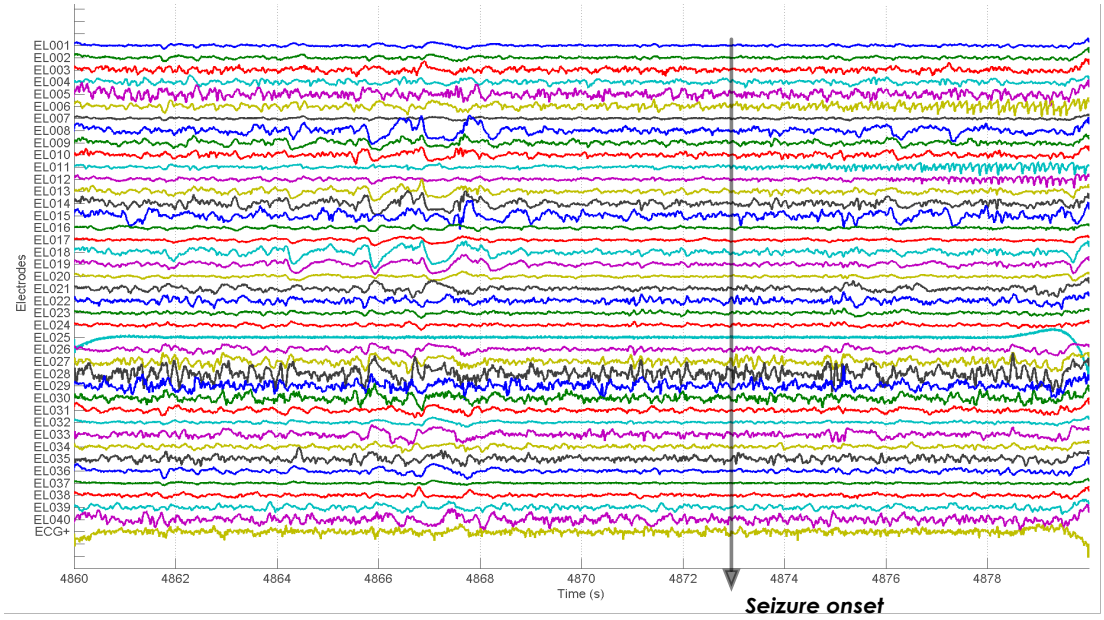


Figure 2.4 – Example of a seizure recorded using intracranial EEG. The seizure begins at second 4873 and involves rhythmic alpha waves in electrodes EL006 and EL011. The pattern propagated to EL012.

2.3 Measures Used For Epileptic Seizure Prediction

Two different models have been proposed to explain the transition from the inter-ictal to the ictal state. One model considers a sudden and abrupt transition

between inter-ictal state (characterized by a normal, apparently random, steady-state activity) and the ictal state (described as a paroxysmal occurrence of a synchronous activity). The second model proposes a rather large period between the inter-ictal and ictal states; in this case, critical parameters gradually change with time (Lopes da Silva et al., 2003). The second model for the transition between the inter-ictal and the ictal state, as explained by Lopes da Silva et al. (2003), is based on rather large random fluctuations of the underlying dynamical system (in a intermediary state, i.e., the pre-ictal state), possibly originating changes detectable in the EEG.

This hypothesis is further confirmed by the existence of several physiological phenomena minutes to hours before seizures. Highly significant BOLD variations were detected before seizure onset based on the analysis of functional magnetic resonance imaging (fMRI). The changes were not restricted to the focal region suggesting the spatio-temporal complexity of the pre-ictal state (Federico et al., 2005). It was also reported that cardiac autonomic activity varied before and after the occurrence of seizures (Delamont et al., 1999; Novak et al., 1999). Furthermore, warning symptoms, also known as prodome and aura, were found in approximately 50% of the patients preceding epileptic seizures (Rajna et al. 1997).

Based on these findings, several studies presented measures to characterize physiological signals in order to describe the pre-ictal state. The EEG signal has been the main source of information for understanding the brain dynamics leading to a seizure, especially due to its temporal resolution, ability to identify electrophysiological abnormalities, exploration of network activity, etc.

2.3.1 Measures characterizing EEG time series

Most of the methods that have been proposed thus far use a moving time window analysis in which some descriptor is computed from a segment of EEG data with pre-defined length. The length of the window assumes values that range from few seconds to several minutes. Different intervals between consecutive windows have also been considered.

The measures are usually classified into two classes, *univariate* if the measure characterizes a single EEG channel, and bivariate (or *multivariate*) if the measure characterizes the relation between two (or more) channels.

2.3.1.1 *Univariate measures*

Linear univariate measures are used to characterize amplitude and frequency information contained in the EEG time series. The characterization of the state (a description of the system at a given time) and dynamics (rules that describe the evolution of the system over time) of one channel time series involves the use of nonlinear univariate measures, for example those derived from the chaos theory.

Statistical Moments

Different studies used statistical moments characterizing amplitude distribution of an EEG time series. Aarabi et al. (2009) reported a decrease of the variance and an increase of kurtosis in the pre-ictal state in comparison with the inter-ictal state. Other authors used these measures as part of a wider set of measures to characterize the EEG time series and discriminate the different brain states (Mormann et al., 2005; Smart et al., 2007; Direito et al., 2011b).

Let us consider $\mathbf{x}$ the input vector (representing the EEG signal) and N the number of samples. Table 2.1 summarizes the first four statistical moments.

Table 2.1 – Statistical moments of time series

Order	**Description**	**Expression**
First	Mean - quantitative measure on the position of the amplitude of the samples	$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$ (2.1)
Second	Variance - quantitative measures on the dispersion of the amplitude of the samples around the mean	$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2$ (2.2)
Third	Skewness - measures the degree of asymmetries of the amplitude distribution of the samples	$\chi = \frac{\frac{1}{(N-1)}\sum_{i=1}^{N}(x_i - \mu)^3}{\sigma^3}$ (2.3)
Fourth	Kurtosis - measures the peakedness of the amplitude distribution and its difference to a Gaussian distribution	$\kappa = (\frac{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^4}{(N-1)\sigma^4}) - 3$ (2.4)

Accumulated Energy

The accumulated energy was proposed by Litt et al. (2001). Their findings identified bursts of energy in the epileptic focus during the inter-ictal period. The authors hypothesized the intensification of these rhythmic discharges before the seizures, leading to an increase in the accumulated energy (signaling upcoming seizures).

The average energy (Esteller et al., 2005) is computed by averaging all successive values of energy calculated in a window, as in (2.5).

$$E_a = \frac{1}{N}\sum_{i=1}^{N} x_i^2 \tag{2.5}$$

A second averaging is performed to compute the accumulated energy. The authors used a 10 windows averaging and then added to a running sum of accumulated energy. The accumulated energy in the *kth* epoch is described as (2.6).

$$E_{ac_k} = \frac{1}{10}\sum_{j=1}^{10} E_a + E_{ac_{k-1}} \tag{2.6}$$

Relative Spectral power in bands

Different physiological (and pathological) processes are related to certain frequency bands. The power spectrum quantifies the signal power associated to different frequency ranges. Mormann et al. (2005) reported a decrease of power in the delta frequency band in the pre-ictal period in comparison to the inter-ictal period. This phenomenon is also accompanied by an increase in the remaining bands.

Five main spectral bands are defined in classical EEG analysis, delta δ (0.5 - 4 Hz), theta ϑ (4 - 7.5 Hz), alpha α (8 – 13 Hz), Beta β (13 – 30 Hz), and Gamma γ (>30 Hz). Let us consider the Fast Fourier Transform (FFT) applied to an input time series $\mathbf{x}$. The relative power can be defined as the average of the squared coefficients of the FFT in the frequency range defined (Blum, 1998). For example, the delta band can be described as (2.7),

$$\delta_{rp} = \frac{1}{P}\sum_{f=0.5\,Hz}^{4\,Hz} p_f \tag{2.7}$$

where p_f is the modulus-squared of the FFT coefficient at frequency f, and P is the total spectral power.

Wavelet Coefficients

The wavelet transform is an alternative to the conventional Fast Fourier Transform. It decomposes the signal in different scales (or resolution levels) into elementary waves called wavelets. These are obtained using a prototype wavelet, also known as mother wavelet, by dilations and contractions. The analysis of the wavelets combines information about the temporal and frequency content of a given signal (Daubechies, 1992; Akin, 2002).

Gigola et al. (2004), and Direito et al. (2008) reported that wavelet-based measures could present variations prior to epileptic seizures. Ouyang et al. (2007) used the wavelet coefficients to determine a nonlinear similarity index between non pre-ictal and pre-ictal data. The authors concluded that the similarity index decreased during the pre-ictal stage.

Hjörth parameters

Hjörth (1970) defined the normalized slope descriptors of activity, mobility and complexity. The parameters provide a dynamic temporal measure of a time series. The activity, mobility and complexity can be considered as measures of mean power, root-mean-square frequency, and root-mean-square frequency spread. Let us consider $\sigma^2(\mathbf{x})$ the variance of the input signal $\mathbf{x}$, and $\mathbf{x}'$ and $\mathbf{x}''$ the first and second derivatives of $\mathbf{x}$. Mathematically, the three Hjörth parameters are defined as (2.8), (2.9) and (2.10) (Barlow, 1993).

$$H_a = \sigma^2(\mathbf{x}) \tag{2.8}$$

$$H_m = \sqrt{\frac{\sigma^2(\mathbf{x}')}{\sigma^2(\mathbf{x})}} \tag{2.9}$$

$$H_c = \sqrt{\frac{\sigma^2(\mathbf{x}'') \times \sigma^2(\mathbf{x})}{(\sigma^2(\mathbf{x}'))^2}} \tag{2.10}$$

Alarcon et al. (1995) reported that variations of Hjörth mobility and complexity characterized seizure onset.

Decorrelation time

An important characteristic of a system's dynamic is the autocorrelation time that is computed using the autocorrelation function. The autocorrelation function of time series **x** can be defined as (2.11).

$$c_{xx_k} = \frac{\sum_{i=1}^{N-k} x_i x_{i+k}}{(N-1)\sigma^2} \tag{2.11}$$

In summary c_{xx_k} compares each point of the time series with another point located k samples away. If uncorrelated, c_{xx_k} is close to zero. On the opposite, if similar (i.e. correlated), the coefficient c_{xx_k} is close to one. The first zero crossing of the autocorrelation function is an estimate of the decorrelation time and tells us about the typical time scales of data variability.

Lai et al. (2002) examined fluctuations of the autocorrelation function in the pre-ictal state compared to the inter-ictal state. A slower decay of the c_{xx} function before the seizures suggests an increase of the periodicity of the signal.

Linear Modeling

An autoregressive (AR) model assumes that the output of the model in each instant t depends on a weighted sum of a certain number of p previous output values (this number is the order of the model) plus a constant term $const$, and a noise as defined in (2.12). The AR model can be interpreted as a predictor (Pardey et al., 1996).

$$Y_t = const + \sum_{i=1}^{p} ar_i Y_{t-1} + \varepsilon_t \tag{2.12}$$

In equation (2.12), Y_t is the predicted value for instant t estimated at instant $t-1$, ar_i are the parameters of the model, $const$ is a constant, and ε_t is noise (usually considered white).

Assuming the stationarity of the signal, it is possible to hypothesize that a quantification of the error between the actual value and the predicted value (also called the forward prediction error) of an AR model is related to the predictability of the behavior of the signal.

The coefficients can be determined by fitting the data, using for example the least squares criterion. Several authors reported that specific patterns of AR parameters may be associated to the pre-ictal state (Rogowski et al., 1981; Rajdev et al., 2010). More recently, AR models were used with machine learning methods for the classification of the pre-ictal segments (Chisci et al., 2010). A multivariate AR model has also been used by Jouny et al. (2005).

Correlation Sum

There is an ongoing debate about the formal meaning of chaoticity measures when applied to noisy, nonstationary time series obtained from the epileptic brain (Lai et al., 2003; Sackellares, 2008). Nevertheless, according to Iasemidis and Sackellares (1996), the intermittency of seizures is unexplainable using concepts of linear dynamics; linear state transitions would only be possible caused by external inputs not always present prior to seizures.

According to the multistable dynamics proposed by Lopes da Silva and colleagues (Lopes da Silva et al., 2003; Suffczynski et al., 2004) for the characterization of the biological neural networks involved in epilepsy, at least two states are possible: an inter-ictal state characterized by random activity and an ictal (paroxysmal) state with a more defined structure.

In order to identify the changes between states, several authors proposed nonlinear measures to characterize properties of the phase space portrait of the system (Iasemidis and Sackellares, 1996). The reconstruction of a state trajectory in the phase plane for the scalar time series $\mathbf{x}$ is obtained by the method of delayed coordinates where in each instant t, the state $\vec{s}_t$ in the trajectory is defined by (2.13) (Takens, 1981).

$$\vec{s}_t = (x_t, x_{t-\tau}, ..., x_{t-(m-1)\tau}) \tag{2.13}$$

The method involves grouping the samples of $\mathbf{x}$ to form vectors as in (2.13). Two parameters must be defined, the embedded dimension m and the time delay τ (Hilborn, 2000). A popular method for choosing an appropriate embedding dimension was presented by Cao (1997). The method estimates the distance between two reconstructed vectors in a $m+1$ embedding dimension and the distance of the same two vectors in a m embedding dimension. According to Cao (1997), m is an appropriate embedding dimension "if two points close in the m-

dimensional reconstructed space will be still close in the $m+1$ -dimensional reconstructed space".

Considering a time series with N samples, each of the $N-(m-1)\tau$ reconstructed vectors represents an instantaneous state of the system m -dimensional space, and the sequential plot of the states defines the state space trajectory, which is an illustration of the system's dynamics. The assumption is that, if the appropriate values are selected for m and τ , the dynamics presented in the original state space are preserved (Hilborn, 2000).

The correlation dimension represents an estimate of the dimensionality of the state space and may be used to characterize the complexity of the signals. The correlation dimension is based on the correlation sum (2.14). The correlation sum quantifies the probability that two vectors of the state space trajectory lie within a hypothetical hypersphere of radius $\frac{R}{2}$. Consider the number of trajectory points equal to S and Θ the Heaviside step-function (Hilborn, 2000; Lai et al., 2002). Then the correlation sum for a hypersphere with radius $\frac{R}{2}$ may be defined as (2.14).

$$C_s(R) = \frac{1}{S}\sum_{j=1}^{S}\left[\frac{1}{(S-1)}\sum_{k=1,k\neq j}^{S}\Theta(R-\left|\vec{s}_j-\vec{s}_k\right|)\right] \tag{2.14}$$

According to Hilborn (2000), the correlation dimension is defined using the correlation sum as (2.15).

$$D_2 = \lim_{R\to 0}\frac{ln(C_s(R))}{ln(R)} \tag{2.15}$$

In other words, the correlation dimension can be estimated by the local slope of the correlation sum logarithm (Lai et al., 2002).

Lehnertz and Elger (1998), Martinerie et al. (1998), Lehnertz et al. (2003), and Maiwald et al. (2004) reported variations of correlation dimension values before seizures.

Lyapunov exponents

In chaos theory, one basic indicator of the predictability of the system is the sensitive dependence on initial state conditions. The exponential divergence or

convergence of nearby trajectories in state space reflects the chaoticity of the system and can be quantified by the Lyapunov exponents. The estimation of the Lyapunov exponents from noisy time series is severely affected by the state space reconstruction parameters (Wolf et al., 1985). After the appropriate estimation of these parameters (Cao, 1997), it is possible to compute the local Lyapunov exponents and measure the growth rate of tangent vectors to a given orbit.

Let us consider d_o the distance between two near neighbor trajectories in the state space in the instant zero and d_t the distance between the trajectories t instants after. The rate of divergence L is defined by (2.16) and is an estimate of the short term largest Lyapunov exponent (Hilborn, 2000).

$$d_t = d_0 e^{Lt} \tag{2.16}$$

The existence of a decrease of the Lyapunov exponents several minutes before the epileptic seizures is documented in several studies (Iasemidis et al., 1990; Iasemidis et al., 1999; Drongelen et al., 2003; Pardalos et al., 2004; Świderski et al., 2005; Świderski et al., 2007). This indicates that the predictability of the brain dynamics tends to increase as the seizure approaches.

Dynamical similarity index

The dynamical similarity index was proposed by Le Van Quyen and colleagues (Le Van Quyen et al., 1999; Le Van Quyen et al., 2000; Le Van Quyen et al., 2003) to measure the dynamical similarity between different segments of the EEG. The idea is to quantify the differences between the dynamics in a normal (or reference), inter-ictal segment of EEG and in a moving window possibly passing through pre-ictal stages.

First, one computes the nonlinear characteristics of a reference EEG window using time delay embedding. The embedding is performed by taking the sequence of time intervals between consecutive negative-to-positive crossings of a determined threshold. The dynamics are reconstructed using the method of delays, considering a high embedding dimension (each state vector of the reference, inter-ictal window is denoted as $\vec{s}_r$). The reconstruction is then projected into a reduced state space considering $\tilde{m} < m$ principal components to remove noise (obtained using the singular value decomposition). A transformation matrix $\mathbf{A}$ is estimated to project each vector $\vec{s}_r$ of the reference m-dimensional state space into a transformed $\tilde{m}$-dimensional space, i.e. $\tilde{\vec{s}}_r = \mathbf{A}\vec{s}_r$.

A similar procedure is used in a running window. Next, each vector of the running window is projected into the $\tilde{m}$ -dimensional space using the transformation matrix $\mathbf{A}$ previously estimated.

To compute the dynamical similarity index we first need to define the cross correlation integral between two dynamics. Assuming that S_r and S_t are the total number of vectors estimated within the reference (inter-ictal) and running window respectively, we can compute the cross correlation using (2.17).

$$C_{rt} = \frac{1}{S_r S_t} \sum_{i=1}^{S_r} \sum_{j=1}^{S_t} \Theta(R - \left| \vec{s}_{t_j} - \vec{s}_{r_i} \right|) \tag{2.17}$$

To obtain a normalized index, we also need to compute the auto-correlation sum as defined in (2.14) (for an easier interpretation, we define the autocorrelation within the reference state trajectory as C_r, and C_t as the autocorrelation within the testing window).

The dynamical similarity is then defined as (2.18).

$$\varsigma_t = \frac{C_{rt}}{\sqrt{C_r C_t}} \tag{2.18}$$

The index returns the probability of finding points estimated in the running window in a neighborhood R of the points estimated in the reference inter-ictal window (Le Van Quyen et al., 2003).

2.3.1.2 *Multivariate measures*

Multivariate measures are used to characterize the interactions between different regions of the brain. Several studies showed that the pre-ictal phenomenon is a spatio-temporal complex mechanism (Federico et al., 2005; Hegde et al., 2007). Multivariate measures are usually based on synchronization indexes and quantify the interactions present in the EEG recordings of two different EEG channels.

Dynamical entrainment

Dynamical entrainment is a measure proposed by Iasemidis et al. (2001) to characterize the 'entrainment' between two brain regions. The main idea is to quantify the differences between the nonlinear behaviors of two electrode sites. Let us consider two channels p and q. The authors suggest the Lyapunov exponent (2.16) to measure the uncertainty about the future state of the system (Iasemidis

et al., 2004). For each channel, a time series is created based on the Lyapunov exponent estimated using a sliding window. Next, the statistical difference between these time series referring to the two channels p and q is computed and denoted here as T_{pq}, i.e. dynamical entrainment. It is defined by (2.19).

$$T_{pq} = \frac{\mu_{p-q}}{\frac{\sigma_{p-q}}{\sqrt{N_W}}} \tag{2.19}$$

$$\mu_{p-q} = \frac{1}{N_W} \sum_{k=1}^{N_W} (L_{p_k} - L_{q_k}) \tag{2.20}$$

In equation (2.19), N_W represents the number of differences within a sliding window, μ_{p-q} denotes the mean of the differences between the Lyapunov exponents estimated for each channel (2.20), and σ_{p-q} the corresponding standard deviation. A high value of T_{pq} corresponds to a low entrainment and vice versa.

This measure is also associated to a "dynamical resetting of the epileptic brain" (Iasemidis et al., 2004). The results presented high sensitivity with relative low false alarms rate (Iasemidis et al., 2001).

Mean phase coherence

Synchrony is another important measure for the characterization of the interaction between two time series. Let us consider two different channels p and q. According to Pikovsky et al. (2001), synchronization may be described in terms of phase locking. In a synchronous state between channels, the phase difference is bounded, and we can estimate a constant phase shift between the phases of the synchronized channels, i.e. $\phi_{p_t} - \phi_{q_t} = \text{constant}$ where ϕ denotes the phase of the channel.

The mean phase coherence is one of the measures used to quantify the phase synchronization between different electrode sites, and is interesting because it allows interpreting the results in terms of basic neurophysiological processes such as postsynaptic potentials, etc. (it is often analyzed as a coupling measure and as a measure of functional connectivity between brain structures). It can be defined by (2.21),

$$\mathbf{M} = \left| \frac{1}{N} \sum_{t=1}^{N} e^{i(\phi_{p_t} \phi_{q_t})} \right| \tag{2.21}$$

where N is the number of samples in the window considered and i is the imaginary unit number. High values of $\mathbf{M}$ indicate a high degree of phase synchronization (for instance, two synchronized signals present $\mathbf{M} = 1$, also known as "in phase"). The analysis of long-term EEG recordings identified an increase in the synchronization measure several hours before seizures (Mormann, 2000).

2.3.2 Predictability

The first studies on seizure prediction, based on a restricted number of patients and limited data (in particular inter-ictal), presented optimistic results. The analysis of the results showed that these were able to distinguish inter-ictal from pre-ictal data. Several authors presented studies, based on small trials, and referred promising results using different measures (an overview of the most important studies is presented in Table 2.2). Elger and Lehnertz (1998) found long lasting and marked changes, up to 25 minutes before the epileptic seizure onset considering the profile of the correlation dimension. Iasemidis et al. (2001) were able to select the most entrained brain regions, according to the dynamical entrainment; 90% of the seizures studied presented significant changes (20 to 40 minutes in advance of the seizure onset). Litt et al. (2001) reported that accumulated energy increased in the 50 minutes before the seizure onset. Accumulated energy based on wavelet coefficients was proposed by Gigola et al. (2004) and the results obtained allowed the authors to conclude that epileptic seizures could be forecasted 70 minutes before their occurrence.

2.3.2.1 *Machine Learning applied to seizure prediction*

Recently, machine learning methods were proposed to develop patient-specific algorithms and exploit the consistency of different features in seizure prediction. In these approaches, machine learning methods were applied to classify high dimensional feature vectors describing the temporal (and spatial) content of EEG data. Mirowski et al. (2009) computed synchronization bivariate measures (using the Freiburg iEEG database [2]) and created high dimensional vectors based on different channel pairs over consecutive time points. The authors compared

[2] Epilepsy Center of the University Hospital of Freiburg, Germany - https://epilepsy.uni-freiburg.de/freiburg-seizure-prediction-project/eeg-database

different machine learning algorithms to discriminate the inter-ictal and pre-ictal states. The best combination *feature-machine learning method* achieved sensitivity of 71% without any false positives on a set of 15 patients out of the 21 patients used in the study. The study considered a pre-ictal period of, at least, 50 minutes in the training of the machine learning models. Chisci et al. (2010) presented a solution based on autoregressive modeling of the EEG time series and support vector machines (SVM) for binary classification of pre-ictal and inter-ictal states. The authors applied a post-processing step using a Kalman filter (Kalman, 1960) in order to decrease the number of false positives. According to the authors, the method proposed achieved a sensitivity of 100% with very low number of false positives due to the regularization behavior of the Kalman filter. Park et al. (2011) described a patient-specific algorithm for seizure prediction on multiple features of spectral power from EEG data and SVM. The authors tested the algorithm in 18 patients with promising results. The methods correctly predicted 97.5% of the seizures and presented a false alarm rate of 0.27 per hour. Both studies were tested considering the Freiburg iEEG database.

Table 2.2 – Some representative previous studies on seizure prediction. (n.s. – not specified)

Author	**Year**	**Measure**	**Number of Patients (number of seizures)**	**Sensitivity (%)**	**False Positive rate (per hour)**	**Anticipation time (assumed pre-ictal period) minutes**	**Total EEG - Inter-ictal controls - hours**
Viglione and Walsh	1975	Linear analysis	n.s.	n.s.	n.s.	n.s.	n.s.
Iasemidis et al.	1990	Lyapunov exponents	n.s	n.s.	n.s.	n.s	n.s
Lehnertz and Elger	1998	Correlation dimension	16 (16)	94	0	12 (30)	21 (16.9)
Le Van Quyen et al.	2001	Similarity index	13 (23)	83	n.s.	6 (20)	15 (0)
Iasemidis et al.	2001	Dynamical entrainment	5 (58)	91	n.s.	49 (variable)	266 (53.9)
Litt et al.	2001	Accumulated energy	5 (30)	90	0.12	19 (180)	>312 (50)
Mormann et al.	2003	Phase synch.	18(32)	81	0	4-221 (240)	117 (49)

D'Alessandro et al.	2003	Feature selection	4 (46)	63	0.28	3 (10)	n.s. (160)
Maiwald et al.	2004	Accumulated energy	21 (88)	30	0.15	n.s. (32)	588 (509)
Gigola et al.	2004	Wavelet Transform accumulated energy	4 (13)	92	0	n.s. (70)	177 (140)
Esteller et al.	2005	Accumulated energy	4 (42)	71	0.11	85 (180)	294 (>168)
Mormann et al.	2005	30 different measures	5 (51)	n.s.	n.s.	n.s. (5-240)	311 (>107)
Mirowski et al.	2009	Synch. measures	21 (88)	100	0	60 (60)	588 (509)
Chisci et al.	2010	AR model	9 (n.s.)	100	n.s.	92 (15)	>588 (>509)
Park et al.	2011	Spectral power measures	18 (80)	97.5	0.27	n.s. (30)	492.4 (433.2)

2.3.3 Weaknesses and caveats

Some concerns, following the suggestions of Osorio et al. (1998), were raised due to the limited number of patients and lack of inter-ictal data of most reported studies; the prediction methods should be evaluated based on both sensitivity (ability of the methods to correctly predict upcoming seizures) and false prediction rate (number of alarms raised that did not precede seizures).

Later analyzes could not reproduce the previous optimistic findings in a prospective fashion. Winterhalder et al. (2003) examined the reproducibility of the results presented by the dynamical similarity index (Le Van Quyen et al., 2003) while proposing an assessment criteria (seizure prediction characteristic). The promising results obtained using accumulated energy could not be reproduced in follow-up studies; Harrison et al. (2005a) reported "no consistent increases in broadband energy prior to seizures". Non linear measures such as Lyapunov exponents also failed to confirm the results presented in previous studies. A study by Lai et al. (2003) showed major obstacles that may seriously affect the predictability power of Lyapunov exponents computed from EEG data. Later, Harrison et al. (2005b) demonstrated "that the correlation dimension and integral have no predictive power for epileptic seizures".

In summary, the optimistic findings based on small, selected datasets, usually based on pre-ictal data (without taking into account the unbalance between pre-ictal and inter-ictal) could not be reproduced on unselected, long-term EEG data.

The improvement of results is conditioned by a better understanding of the causes behind false positives and false negatives, especially in the inter-ictal period (usually associated to high false positive rate) and the challenges imposed by confounding variables. For example, the signal-to-noise ratio is usually very low (particularly in scalp recordings) and thus analytical techniques based on amplitude or power may reflect artifacts related to noise. Studies using continuous multi-day recordings revealed that daily rhythms (the circadian fluctuations) might affect the measures that were proposed to characterize epileptic seizures. Relevant aspects of the epileptic process may, in some cases, contribute marginally (Kuhnert et al., 2010). Another important aspect is the influence of AED. The AED blood level may be an important factor in the behavior of characterizing measures of the EEG data (Stacey et al., 2011).

The understanding of these confounding variables is an important step towards the improvement of prediction algorithms (Mormann et al., 2007).

In 2005, a set of five patients with continuous multi-day EEG recording was the subject of a series of studies for the First International Collaborative Workshop on Seizure Prediction (Lehnertz and Litt, 2005; Mormann et al., 2005). The aim of these studies was to test and compare the different approaches suggested. The results obtained showed poor predictive performance of univariate measures (Jouny et al., 2005; Mormann et al., 2005). In particular, non-linear measures did not present higher predictive performance than linear measures. Iasemidis et al. (2005) and Mormann et al. (2005) reported that multivariate measures presented better performance results that reflected changes in the time scale of hours prior to the seizures. However, the evaluation scheme used to obtain these results considered each channel separately.

One key finding was that statistical significant electrographic changes detectable by certain dynamical measures occurred before seizures (Mormann et al., 2005). The authors concluded that this evidence of a "pre-ictal state" could be used for therapeutic devices if appropriate algorithms could be developed. Nevertheless, a system working with 100% sensitive and without generating false positives is considered unrealistic in the present state of knowledge.

2.3.3.1 *Assessing the performance of a prediction algorithm*

Another important consensus obtained in the First International Collaborative Workshop on Seizure Prediction (Lehnertz and Litt, 2005) was the need for a common framework aimed at assessing the performance of prediction algorithms. Before addressing the relevance of the performance in terms of applicability in clinical devices, the methods must (at a minimum) demonstrate that they perform better than chance in a prospective fashion. The importance of statistically sound validation of the methods was also tackled by, for example, Mormann et al. (2007), and Snyder et al. (2008). To date, mainly two different approaches were suggested to assess seizure prediction methods: based on Monte Carlo simulations (Andrzejak et al., 2003; Mormann et al., 2005), or based on the comparison with analytical results derived from naïve prediction schemes (Winterhalder et al., 2003; Schelter et al., 2006a).

Andrzejak et al. (2003) introduced the concept of seizure time surrogates. First, one constructs an ensemble of seizure time surrogates where the original seizure-onset times are shuffled and placed in times randomly selected from inter-ictal periods (Figure 2.5). In other words, the seizure time surrogates are artificial seizure-onset time series, where some constrains related to properties of the original seizure-onset times are imposed (e.g. number of seizures, distribution intervals between consecutive seizures, etc.). A determined algorithm is then tested on both seizure time surrogates and original seizure-onset time series. If the predictor (considering the original time series) outperforms the seizure time surrogates (for a number of independent realizations), it is possible to reject the null hypothesis, i.e. "the transition from inter-ictal to the ictal state is an abrupt phenomenon, an intermediate pre-ictal state does not exist" (Andrzejak et al., 2003). The null hypothesis considers that the predictive power of the seizure predictor is not better than the predictive power expected by a random distribution of the seizure-onset times and "can be rejected at a level of significance determined by the number of surrogates used" (Andrzejak et al., 2003).

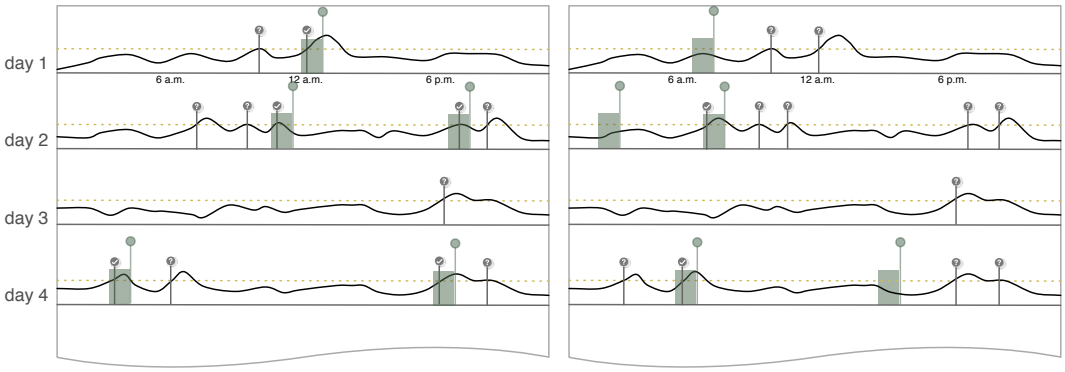


Figure 2.5 – Schematic representation of the seizure time surrogates method used for statistical validation of the predictors. On the left, the seizure onset times and the alarms raised by a determined method whenever the measure exceeds a threshold (dotted line). On the right, artificial seizure onset times generated by randomly shuffling the original inter-seizure intervals.

Winterhalder et al. (2003) and Schelter et al. (2006a) suggested a different approach based on the concept of seizure prediction characteristic (SPC) to evaluate and compare the performance of seizure prediction methods. According to the authors, an ideal seizure prediction method would indicate the exact point in time when a seizure would occur. However this is an unrealistic view of the real behavior of predictive models, so the authors suggest the seizure occurrence period (SOP) to deal with the uncertainty. SOP is defined as "the period during which the seizure is to be expected" (Winterhalder et al., 2003). The authors also considered the seizure prediction horizon (SPH) that represents the time window for a clinical device to potentiate a therapeutic intervention or a behavioral adjustment (also known as intervention time, IT). An alarm is considered a correct prediction when the seizure onset occurs in the SOP (Figure 2.6).

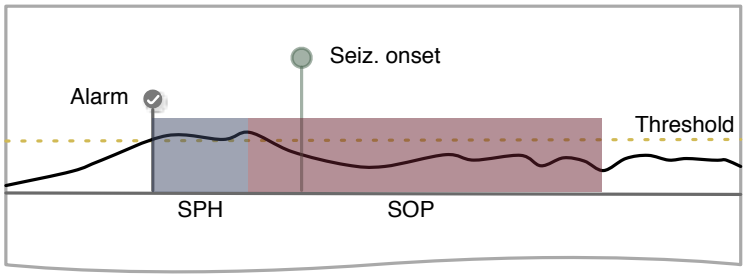


Figure 2.6 – Schematic representation of the method Seizure Prediction Characteristic and definition of a correct prediction according to its parameters. The alarm is generated when the measure (black solid line) exceeds a certain predefined threshold (dashed yellow line). The seizure onset time must occur after the end of the SPH and before the end of the SOP. Alarms outside this period are false predictions.

The sensitivity of the prediction algorithm is calculated as the fraction of correct predictions to all seizures, and the false prediction rate is the number of false prediction in a given time interval (FPR.h^{-1}). Due to the dependency between the sensitivity and specificity (or FPR.h^{-1}) the performance of the method should always be evaluated taking into account both measures (Osorio et al., 1998).

Another important aspect is that the sensitivity of the seizure prediction method should be higher than an unspecific one for a determined specificity value (random or periodical alarms). Let us consider the random predictor that is generating alarms according to a determined FPR_{max} (i.e. the false prediction rate equals the upper bound defined by a FPR_{max}).

The probability of emitting an alarm (false positives or FP) at any sample of the time series with N samples is defined by $P_{poisson} = \frac{FP}{N}$ (Schelter et al., 2006a).

If we consider a time interval with the duration of SOP (i.e. Δt =SOP), the probability of emitting an alarm within that period is defined by (2.22).

$$prob_{\Delta t} = 1 - (1 - P_{poisson})^{N(\Delta t)} \tag{2.22}$$

This probability can be approximated by (2.23) (Schelter et al., 2006a).

$$P_{SOP} = 1 - e^{-FPR_{max} \times SOP} \tag{2.23}$$

In summary, (2.22) and (2.23) represent the probability of a random predictor raising at least one alarm during the SOP.

Schelter et al. (2006a) extended the analysis for the use of multiple testing. According to the authors, the probability of randomly predicting k of K seizures is governed by a binomial distribution. The probability of predicting k seizures taking into account a binomial distribution with probability P_{SOP} may be described as (2.24).

$$P_{binom}\{k; K; P_{SOP}\} = \sum_{j \geq k} \binom{K}{j} p^j (1 - P_{SOP})^{K-j} \tag{2.24}$$

If the prediction method proposed uses multiple features or channels, this should also be considered, i.e. the number of independent extracted features should also be taken into account in the definition of the unspecific random predictor. Let

us consider z the number of independent features extracted. The probability defined in (2.24) should be corrected to (2.25).

$$P_{binom,z}\{k;K;P_{SOP}\} = 1 - \left(\sum_{j<k} \binom{K}{j} p^j (1-P_{SOP})^{K-j} \right)^z \tag{2.25}$$

Considering a determined significance level α_p, a statistical test may be derived from the probability computed in (2.25) and can be defined as (2.26).

$$critical\ value_{rand,z} = max_k (P_{binom,z}\{k;K;P_{SOP}\} > \alpha_p) \times 100\% \tag{2.26}$$

If at least one of the independent features considered presents better results than the ones obtained by the random predictor, we can consider the existence of predictive information contained in the EEG data (Schelter et al., 2006a).

2.3.4 Present and future challenges

Attempts for testing seizure prediction algorithms in a prospective manner, using validation schemes based on seizure times surrogates or random predictors were previously reported but the sensitivity and false prediction rate obtained were unacceptable for clinical applications (Mormann et al., 2007). The Chaovalitwongse et al. (2005) study, for example, was considered inconclusive because the parameters used in the validation set were optimized using the original onset times; therefore the study was not strictly prospective. Iasemidis et al. (2005) presented a long-term prospective seizure prediction study based on the data of two patients. The results showed promising sensitivity results however were associated to a low specificity. However, parameter selection was performed in the validation set.

Mormann et al. (2007) proposed some guidelines to assure the methodological quality of studies: studies should be tested on unselected continuous long-term recordings in order to comprise different physiological states of the patients; studies should report both sensitivity and specificity; the results should be tested using statistical validation schemes based on Monte Carlo simulations (surrogate time series) or random prediction models to prove that the algorithms perform (at least) above chance level; prediction algorithm can be optimized using training data (in-sample), but should be test on independent testing data (out-of-sample).

One important step towards the quality of new studies is the quality and size of the database. Continuous long-term recordings fulfilling certain qualitative criteria would allow research groups to assess the validity of the methods, would ease the transition to a pseudo-prospective approach.

Despite the improvements made in recent years, no algorithm has yet shown the performance necessary for clinical application. Advances in electrode development, signal acquisition, data processing, computational analysis, and statistical evaluation are fundamental to improve the understanding on the underlying mechanisms of seizure generation (Stacey et al., 2011). A better understanding of the complex spatio-temporal interactions between different brain regions may be crucial for the development of improved methods for seizure prediction.

2.4 Seizure Prediction and Seizure Intervention Devices

Devices related to epilepsy might not cure epilepsy but represent an important asset in the control and monitoring of the clinical manifestations of uncontrolled seizures (Fisher, 2012).

For example, alerting systems play a major role in the decrease of severe injuries caused by epileptic seizures by notifying the parents or relatives of epilepsy patients of the occurrence of particular symptoms. Different approaches are used in seizure alerting systems, such as accelerometer placed in the bed and/or patient (Carlson et al., 2009), movement detection using video setups (Cunha et al., 2012), EEG monitoring (Gotman, 1982), sound monitoring, etc.

A different strategy is followed in seizure intervention systems; the main idea is to control epileptic seizures. There are two main approaches: open-loop and closed-loop systems.

Open-loop systems intervene based on a schedule, ignoring the possibility or probability of seizure occurrence. Electrical deep brain stimulation has been used and demonstrated potential to reduce the frequency of epileptic seizures (Fisher et al., 2010). Alternatively, some authors proposed the electrical stimulation of the vagus nerve reporting the decrease in the frequency of seizures and "positive effects in terms of seizure severity" (Shahwan et al., 2009). Other experimental approaches

have been proposed such as the cooling of brain cortex (blocking synaptic transmission) (Rothman et al., 2005), etc.

On the opposite, closed-loop systems should only intervene as the seizure probability increases, either using AED, electrical stimulation of the brain or vagus nerve, or generating alarms (Litt et al., 2001; Stacey and Litt, 2008; Fisher, 2012). However, the success of such therapeutic devices depend of the research of a reliable seizure prediction method, and would be an important tool to reduce the possibility of side effects related to the open-loop systems (Litt et al., 2003; Shoeb et al., 2011; Fisher, 2012), high false positive rates and low sensitivities (Schulze-Bonhage et al., 2010). The ability of delivering the therapy at the appropriate times is extremely helpful compared to open-loop strategies due to the possible reduction of medication, electrical stimulation, etc., and to the increase of the efficacy of the therapies.

Some of the properties desirable for an intervention device are the possibility to implant the device (size, magnetic and electric fields associated, etc.), low power consumptions (concerns related to the battery, heating of surrounding tissues and neuronal damage (Raghunathan et al., 2011)) and *convenient-to-carry* (Litt et al., 2001).

2.4.1 Commercial and research efforts

Different companies have proposed devices (and continuing to develop) using open-loop and closed-loop strategies (using either seizure detection and seizure prediction).

Neuropace[3] is the company responsible for the development of a responsive neurostimulator (RNS) that is programmable, battery powered, microprocessor-controlled. The device is implanted in the cranium and connected to leads placed near the patient's seizure focus.

Cyberonics[4] is developing a vagus nerve stimulator (VNS) for the treatment of refractory epilepsy. Shoeb et al. (2009) has also presented a study performed at the Massachusetts General Hospital based on the pairing of a VNS device and a seizure detection method.

[3] http://www.neuropace.com

[4] http://us.cyberonics.com

Neurovista[5] is performing an extensive clinical study in Australia based on iEEG. The Seizure Advisory System (SAS) is based on electrodes located in the brain surface that are connected by wires to the storage part placed on the chest of the patients. Another part of the setup is an external processing device that communicates wirelessly with the storage unit. The release of preliminary results has been very limited as well as the access to the processing performed to predict seizures.

EPILEPSIAE is an European consortium financed by the European Union FP7 project, with three major goals: development of *Brainatics*, a transportable alarming device, development of seizure prediction algorithms (Teixeira et al., 2011) and creation of a large, multicenter database (Klatt et al., 2012).

2.5 Summary and Discussion

In this chapter, we presented an overview of the theoretical aspects related to the prediction of epileptic seizures. First, biological and electrophysiological concepts related to the epileptic seizure generation were presented. A summary of the most noticeable seizure prediction studies was presented in Section 2.3, as well as the difficulties faced to reach clinically sustainable results. Finally, Section 2.4 presents some of the most important efforts towards the development of seizure intervention devices, and highlights the importance of the research of closed-loop intervention systems.

According to Stacey and Litt (2008) the efficacy of closed-loop intervention devices based on seizure detection algorithms has two main drawbacks: uncertain accuracy of the methods and late intervention of the devices, i.e. intervention once the seizure has started. An indispensible step towards the development of new closed-loop intervention devices is the research of a reliable seizure prediction method.

Despite of the various efforts developed in recent years, which resulted in many seizure prediction studies, the results are still unsatisfactory in terms of clinical applicability.

The next chapters present different approaches to seizure prediction. Chapter 3 presents an extensive study based on the application of machine learning

[5] http://www.neurovista.com

methods to a large database and taking into account the main guidelines presented in this chapter. Chapters 4 and 5 suggest novel approaches to the characterization of the pre-ictal period with the aim of improving seizure predictors.

3. Epilepsy and Pattern Recognition

The development of seizure prediction algorithms able to achieve clinically sustainable results would allow the transition from open-loop to closed-loop strategies for seizure management, where continuous monitoring and successful alarm generation would reduce the side effects associated to open-loop mechanisms.

In this chapter several techniques of machine learning are applied to a large set of 216 epileptic patients from the EPILEPSIAE database for seizure prediction. Classification is made using support vector machines (SVM) in a multiclass structure. Pre-processing for classification such as feature selection, feature reduction is addressed by appropriate techniques.

Firstly, the main concepts of multiclass classification and pattern recognition are introduced. The algorithm conceptualizes data extracted from EEG as feature vectors and solves a classification problem in order to generate alarms, following the main guidelines presented in Chapter 2. Optimization methods are proposed to further improve the results obtained. The results are discussed taking into account the most noticeable studies in machine learning and seizure prediction. Most of the work in this chapter has been previously published. Specifically, Section 3.4 presents the work published in C.3 and C.5, Section 3.5 corresponds to publication C.4, and Section 3.6 follows publication C.2.

3.1 Overview of Pattern Recognition and Machine Learning

The mechanisms leading to epileptic seizures are still largely unknown. Therefore, the ability to develop appropriate model-driven approaches has been a challenging task (Lopes da Silva et al., 2003; Suffczynski et al., 2004, Wendling et al., 2005). Machine learning methods approaches classification and prediction as the construction of systems able to learn from experimental data.

Several authors proposed machine learning methods as a tool to solve detection and prediction of epileptic seizures. To a certain extent, seizure detection has been successfully achieved. Shoeb (2009) proposed a SVM-based binary classification approach (seizure, no-seizure) to detect seizures in a patient-specific setting with very promising results. Seizure prediction has also been tackled using machine learning but the results are still far from clinical usability (for a review see Section 2.3.2.1).

3.1.1 Basic concepts

The similarity between objects represents the fundamental concept in pattern recognition. The degree of similarity between two objects represents the proximity between the values of the characteristics used in their description. Let us consider *apples*. To conceptualize (and to distinguish) apples, we use certain features such as shape, taste, texture, weight, etc. (Marques de Sá, 2001).

In pattern recognition, to determine the *class membership* of different objects, an identical analysis is made, i.e. objects are grouped into classes according to the proximity of the values of certain properties of the objects (also known as *features* or *attributes*). These should be as distinctive as possible, i.e., should maximize the differences between objects of different *classes*. The success of classification depends on two main aspects: the discriminant capability of the *features* that are extracted from each *object* (or *pattern*), and the intelligence of the *classifier* (the set of rules used to determine the class of each pattern).

3.1.2 The design of patterns

According to the main guidelines presented by Mormann et al. (2007), the design and testing of seizure prediction methods should be performed considering long-term EEG data. EEG data are based on dynamic processes evolving in specific regions of the brain. For instance, in the seizure onset zone, the EEG data may reflect a sudden change in the spectral energy distribution (Grewal and Gotman, 2005). As the seizure evolves, the characteristics of the signal continue to change.

To capture these dynamics, a moving observational time window with N_W samples is used to create time-profiles of measures that characterize certain properties of the data. These measures are the features, and each sample of the features over time represents a pattern. Each pattern can then be defined as a collection of features (univariate and multivariate) that describe a variety of

characteristics extracted from the EEG data expected to have a discriminant capabilities for the aim of the study. In summary, the *feature vector* encodes the data structure of an *epoch* (a specific time interval), capturing spatial information (encoded in the channels) and information derived from the selected features (characterizing the temporal dynamics, frequency content, etc.).

Figure 3.1 presents the general method used to design patterns. In the example illustrated, we consider an EEG epoch (the observation window) with N_W samples and compute the spectral power in five frequency bands based on the FFT coefficients. To simultaneously capture spectral and spatial information, we concatenate the five spectral features extracted from each EEG channel to form a feature vector. The example considers non-overlapping windows, so a new feature vector is designed t seconds (i.e. N_W samples) after the first window.

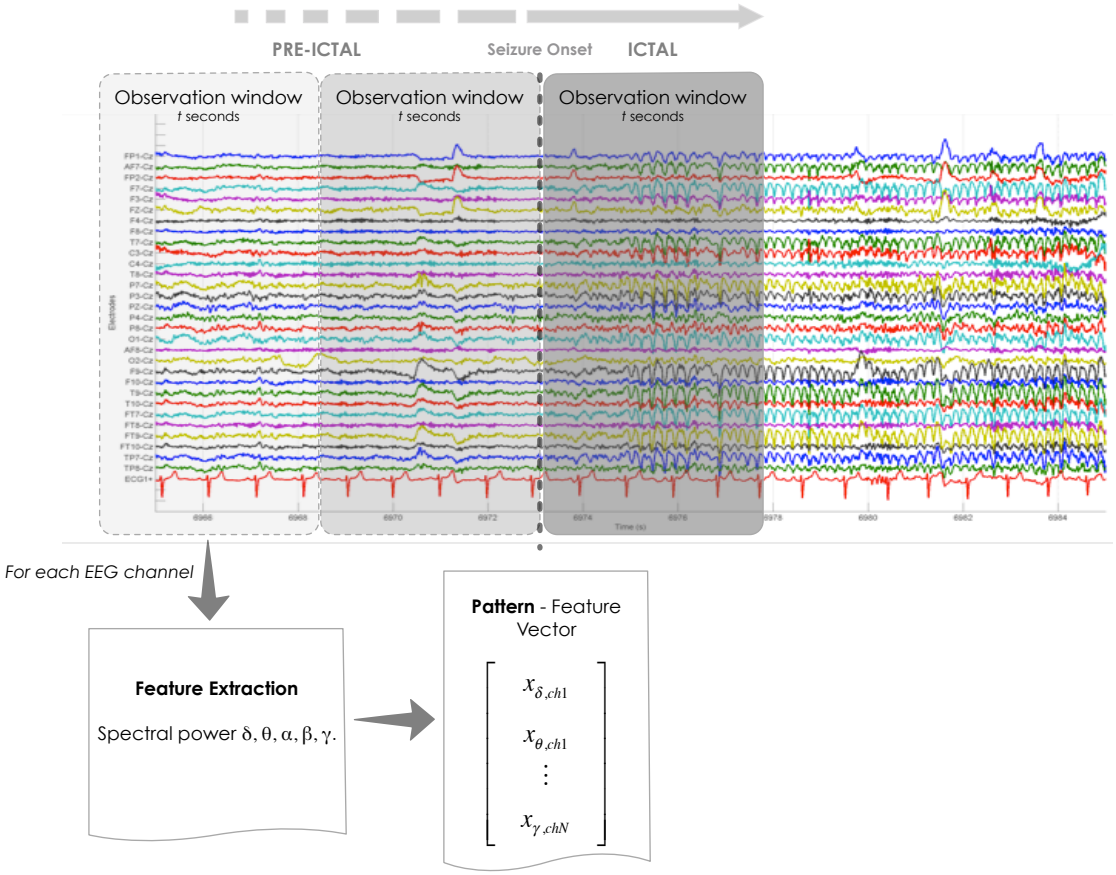


Figure 3.1 – The design of patterns. A set of measures (in the case spectral power) is computed using an observation window with N_W samples. The values (measures in each channel) are concatenated into a feature vector at time t. The dimension of the feature vector, N_{feat}, is the product of the number of features by the number of channels

3.1.3 Classification

A classifier can be considered as a set of rules used to determine class membership of feature vectors with N_{feat} components represented in the feature space $\Re^{N_{feat}}$. The main goal is to divide the feature space $\Re^{N_{feat}}$ into regions assigned to the different classes (Marques de Sá, 2001).

Machine learning is an area of artificial intelligence dedicated to the design and development of algorithms able to learn from known data and predict unknown data by constructing, for example, statistical-based models. These models can be interpreted as a map between a multivariate input and the output.

The inputs are the feature vectors, constructed using the descriptors of the EEG data. The output is a membership discriminating among four different classes: *ictal* (patterns occurring during the seizure), *pre-ictal* (patterns before the seizures), *post-ictal* (patterns after the end of the seizure), *inter-ictal* (the rest of the time).

One part of the data, the training set, is used for the supervised construction of the classifiers i.e. supervised training, leading to the definition of decision boundaries in the feature space $\Re^{N_{feat}}$. Once the decision boundaries are defined, the classifier determines the class membership on the rest of the data composing the testing set. Although the classifier is optimized for the training set, its performance is measured in the testing set, i.e. its generalization capability in unknown data.

3.1.3.1 *Support-Vector machines*

Support Vector Machine (SVM) is a machine learning algorithm used for multi-dimensional classification characterized by good generalization capability, few number of parameters to tune and dependence on the data (Burges, 1998; Bishop, 2006). The decision boundary estimated from known data, can be used to increase the understanding of the data (e.g. interpret the most important variables (Chang and Lin, 2008; Dhar and Cherkassky, 2010), etc.).

SVM generalization performance (i.e. error rates on test sets) has shown to be better, or at a minimum comparable, to other machine learning methods such as artificial neural networks, random forest, etc. (Burges, 1998; Meyer et al., 2003).

The algorithm is based on the structural risk minimization proposed by Vapnik and Chervonenkis (1974). According to the statistical learning theory, the

generalization error bound is minimized by maximizing the *margin* (see Figure 3.2); i.e., the shortest distance between the decision function (also known as the separating *hyperplane*) and the closest data points of each class (Burges, 1998).

Let us consider the SVM in its simplest form, i.e. determining the decision boundary of a two-class classification problem.

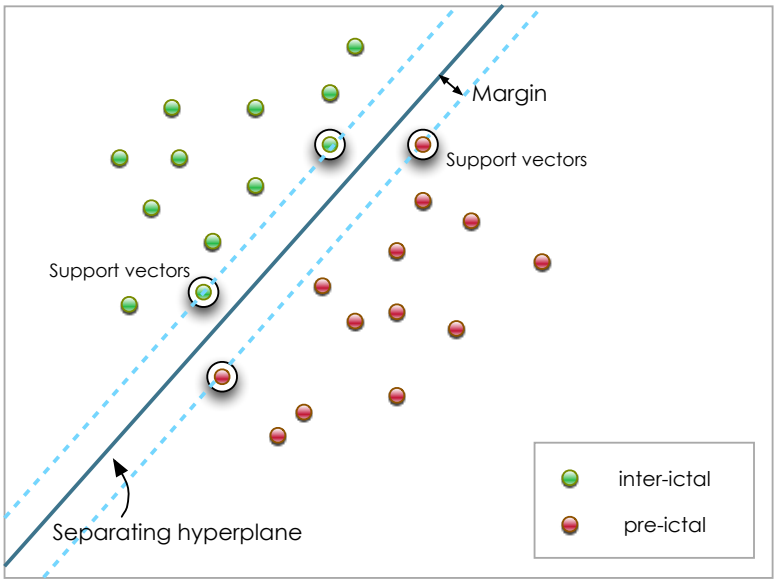


Figure 3.2 – Linear separating hyperplane for a separable two-dimensional case. Support vectors are circled and are placed on the boundary of the class, i.e. they support the boundary of the class. The margin is the distance between the dashed lines (boundaries of the classes) and the separating hyperplane and is maximized by the placement of the hyperplane.

The points $\mathbf{x}$ in (3.1) are used to define the separating hyperplane,

$$\mathbf{w}^T \cdot \mathbf{x} + b = 0 \tag{3.1}$$

where b is a bias and $\mathbf{w}$ is a vector of weights perpendicular to the separating hyperplane. Thus, the decision function D for classifying any given point $\mathbf{x}_i$ is (3.2),

$$D(\mathbf{x}_i) = sign(\mathbf{w}^T \cdot \mathbf{x}_i + b) \tag{3.2}$$

or, using inequalities (3.3).

$$\begin{aligned} \mathbf{w}^T \cdot \mathbf{x}_i + b \geq 0 \quad & if \ y_i = +1 \\ \mathbf{w}^T \cdot \mathbf{x}_i + b < 0 \quad & if \ y_i = -1 \end{aligned} \tag{3.3}$$

The decision function is invariant under rescaling allowing to divide the weights $\mathbf{w}$ and bias b by any scalar without changing the hyperplane. By using the appropriate scalar, we can define the shortest distance of a point to the hyperplane (margin) as $\frac{1}{||\mathbf{w}||}$. The points closer to the hyperplane may be defined as (3.4).

$$y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) = 1 \;\; if \; \mathbf{x}_i \; is \; a \; support \; vector \tag{3.4}$$

The hyperplanes satisfying these conditions are known as *canonical hyperplanes*, and the feature vectors as *support vectors* (Figure 3.3).

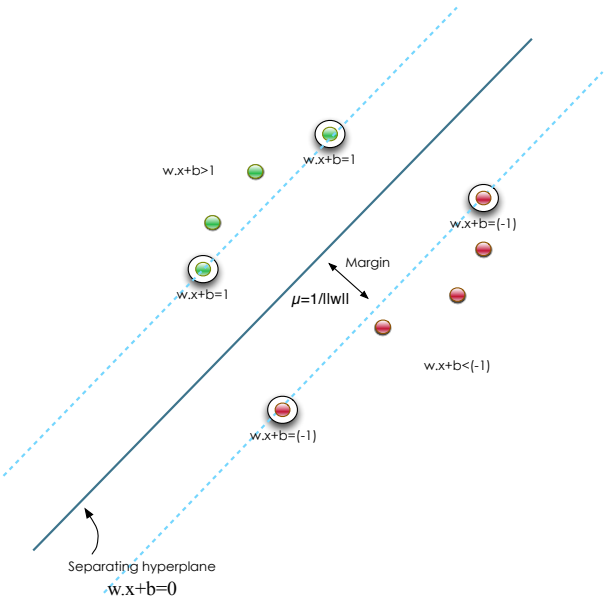


Figure 3.3 – Distance of the support vectors to the separating hyperplane. Points farther apart do not influence the decision function.

To maximize the margin it is necessary to maximize $\left|\left|\frac{1}{\mathbf{w}}\right|\right|$, which is equivalent to minimize the denominator $||\mathbf{w}||$. The SVM approach to solve the minimization problem can be expressed as the constrained optimization problem (3.5),

$$min\,\Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} \tag{3.5}$$

subject to (3.6).

$$y_i(\mathbf{w}^T \cdot \mathbf{x}_i + b) \geq 1 \tag{3.6}$$

The quadratic function $\Phi(\mathbf{w})$ and linear constraint (3.6) define a convex problem, thus assuring the global minimization of $||\mathbf{w}||$. The Lagrangian method can be used to solve the problem.

Let us consider N_P the number of patterns in the training set. According to Vapnik (1999) the solution can de defined as (3.7),

$$L_{lagrange}(\mathbf{w},b,a) = \frac{1}{2}(\mathbf{w}^T\mathbf{w}) - \sum_{i=1}^{N_P} a_i \left\{ y_i \left[\mathbf{w}^T \mathbf{x}_i + b \right] - 1 \right\} \tag{3.7}$$

where a_i are the lagrange multipliers. To find the optimum solution $\mathbf{w}_0, b_0, a_0$ with respect to both $\mathbf{w}$ and b, we will minimize (3.7) solving the stationary equations (3.8) and (3.9),

$$\frac{\partial L_{lagrange}(\mathbf{w}_0, b_0, a_0)}{\partial b_0} = 0 \tag{3.8}$$

$$\frac{\partial L_{lagrange}(\mathbf{w}_0, b_0, a_0)}{\partial \mathbf{w}_0} = 0 \tag{3.9}$$

leading to (3.10) and (3.11).

$$\sum_{i=1}^{N_P} a_{0_i} y_i = 0 \quad a_{0_i} \geq 0, i = 1,...,N_P \tag{3.10}$$

$$\mathbf{w}_0 = \sum_{i=1}^{N_P} a_{0_i} y_i \mathbf{x}_i \quad a_{0_i} \geq 0, i = 1,...,N_P \tag{3.11}$$

The solution must also satisfy the Kuhn-Tucker conditions (3.12).

$$a_i \left\{ y_i \left[\mathbf{w}^T \mathbf{x}_i + b \right] - 1 \right\} = 0, \quad i = 1,...,N_P \tag{3.12}$$

Only some training patterns have nonzero coefficients a_{0_i}, the support vectors. The number of support vectors N_{SV} is upper bounded by N_P.

Replacing (3.11) into the Lagragian (3.7) while taking into account the Kuhn-Tucker conditions, defines the dual of our optimization problem (3.13).

$$L_{lagrange\_dual}(a) = \sum_{i=1}^{N_P} a_i - \frac{1}{2} \sum_{i,j=1}^{N_P} y_i y_j a_i a_j \mathbf{x}_i^T \mathbf{x}_j \tag{3.13}$$

Considering the parameters estimated, it is possible to define the decision function for new points $\mathbf{x}_{new}$ as (3.14).

$$D(\mathbf{x}_{new}) = \sum_{i=1}^{N_P} y_i a_i \mathbf{x}_i^T \mathbf{x}_{new} + b_0 \tag{3.14}$$

The fact that the sum uses all the points (N_P) is not relevant since non-boundary points have Lagrange multipliers a_i equal to zero.

Linearly non-separable data

Most of the real problems in classification are not linearly separable as in Figure 3.3; instead the separating hyperplane may have different configurations.

Soft margin

When the classes are non-separable because of possible outliers, the optimal hyperplane must consider the ideal separable situation avoiding the particularities of the outliers (that decrease the generalization capability of the decision boundary).

Cortes and Vapnik (1995) proposed a formulation for the SVM introducing the soft margin in order to reduce the effect associated with outliers and noise. These authors introduced additional non-negative slack variables $\xi_i \geq 0$ and a penalty parameter C, defining a new optimization problem (3.15),

$$min\, \Phi(\xi) = \frac{1}{2} \left\| \mathbf{w} \right\|^2 + C \sum_{i=1}^{N_P} \xi_i \tag{3.15}$$

subject to (3.16).

$$y_i \left(\left(\mathbf{w}^T \mathbf{x}_i \right) + b \right) \geq 1 - \xi_i, \quad i = 1,...,N_P \tag{3.16}$$

The main idea is to maximize the margin, and simultaneously penalize erroneous data (patterns far from the ideal separable situation) by minimizing the sum of errors $\sum_{i=1}^{N_P} \xi_i$. It is possible to determine the decision function using the same formalism presented in (3.7) to this situation (Vapnik, 1999).

The constant C defines the tradeoff between the two terms in (3.15). A large value of C increases the influence of the sum of errors, leading to a small tolerance to misclassification error (tendency to smaller margins). On the opposite, small C constant decreases the influence of the errors and the margin tends to expand. In practice, the optimal value of C must be found by trial and error using a validation set and it cannot be readily related to the characteristics of the data set.

Kernel approach

To tackle the problem of linearly non-separable data, Cortes and Vapnik (1995) proposed transform the input vector space into a higher-dimensional feature space Z where linear separability is improved. The transformation is done through some nonlinear mapping chosen *a priori*. This would allow constructing an optimal separating hyperplane.

According to Vapnik (1999), the estimation of the optimal separating hyperplane in Z uses the inner product of two vectors $\upsilon(\mathbf{x}_1)$ and $\upsilon(\mathbf{x}_2)$, that corresponds to the mapping into the feature space of vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ (3.17),

$$\mathbf{x}_i \in \Re^{N_{feat}} \rightarrow \upsilon(\mathbf{x}_i) \in \Re^Z \tag{3.17}$$

where υ represents the mapping function. It is possible to estimate the inner product of the vectors in the Z feature space as a function of two variables in original input space using a kernel function.

Let us consider K a kernel function in the original input space. Computing $K(\mathbf{x}_i, \mathbf{x}_j) = \upsilon(\mathbf{x}_i) \cdot \upsilon(\mathbf{x}_j)$, we avoid computing the image of each vector in the Z feature space. We can rewrite equation (3.14) considering the kernel as (3.18).

$$D(\mathbf{x}_{new}) = sign(\sum_{i=1}^{N_P} y_i a_i K(\mathbf{x}_i, \mathbf{x}_{new}) + b) \tag{3.18}$$

Multiclass classification

Despite being originally designed for binary classification, combinations of SVMs have been successfully applied in multi-class classification problems. Hsu and Lin (2002) reported that *one-against-one* is the most suitable method for practical use.

The method constructs $k(k-1)/2$ classifiers, where k is the number of classes. Each classifier is trained on data from only two classes. The final decision

is made according to the voting approach *max wins*, in which the label predicted for an input vector is the one that is voted the most among all classifiers.

3.2 Patient-specific Seizure Predictor

A patient-specific seizure predictor is developed through a succession of processing stages. Figure 3.4 illustrates the main parts of the method. In the following section, we detail each step of the predictor.

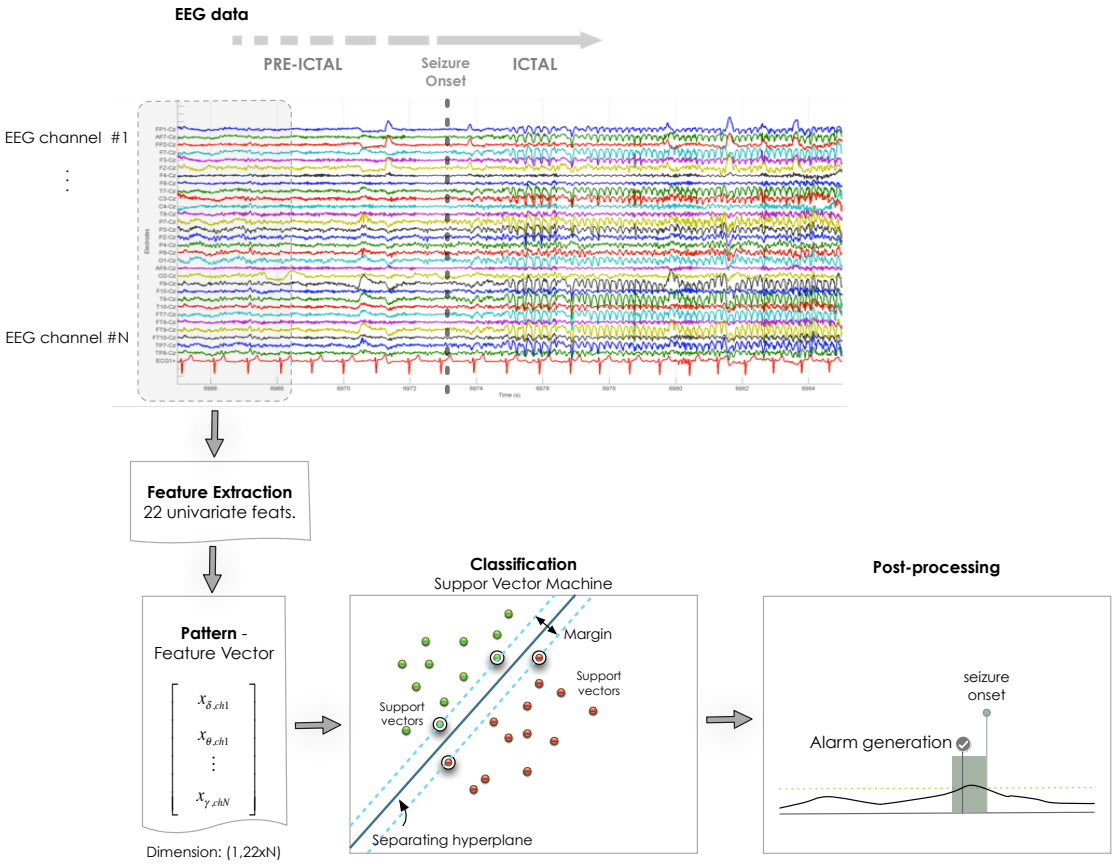


Figure 3.4 – Schematic representation of the predictor proposed. The first step extracts 22 univariate features from each channel. After the construction of feature vectors, we apply a SVM to learn a decision function on the training set. The testing set is classified using the model trained. A post-processing step is necessary to generate alarms based on the classifier output.

3.2.1 Dataset description

We have trained and tested our predictors on the EPILEPSIAE database [6] (Ihle et al., 2012; Klatt et al., 2012). This database contains data of 275 epileptic

[6] More information available at www.epilepsiae.eu

patients, including high quality, long-term EEG scalp (217 patients) and intracranial (58 patients) recordings. The data represents the joint effort of different epilepsy centers (Coimbra, Paris and Freiburg) on behalf of the European project EPILEPSIAE, and were based on a common annotation and acquisition framework for an efficient organization and analysis.

Seizure onset times and epileptiform activity were identified by certified epileptologists from the different centers. The sampling rate of the data varied from 250Hz up to 2500Hz.

The study in this chapter included 185 scalp recording patients (out of the 217 available) that were successfully trained using the strategy and parameters proposed in our approach.

Some of the recordings presented limitations that did not allowed us to consider them in this study. Numerical difficulties were observed (some related to inconsistencies in the feature files), not allowing the appropriate training of the classifiers. Time discontinuities in the recording, saturation of electrodes, and missing data in certain pre-ictal periods are some of the artifacts that limited the inclusion of all the available recordings. The solution to these problems would be to analyze the deficiencies of each dataset and select only the suitable EEG (and features) segments. However, this would influence the strategy of the selection of training and testing sets. Nevertheless, we believe that the number of patients presented is significant for the objective of the study.

Approximately 63% of the patients in this group presented focal temporal seizures (Figure 3.5).

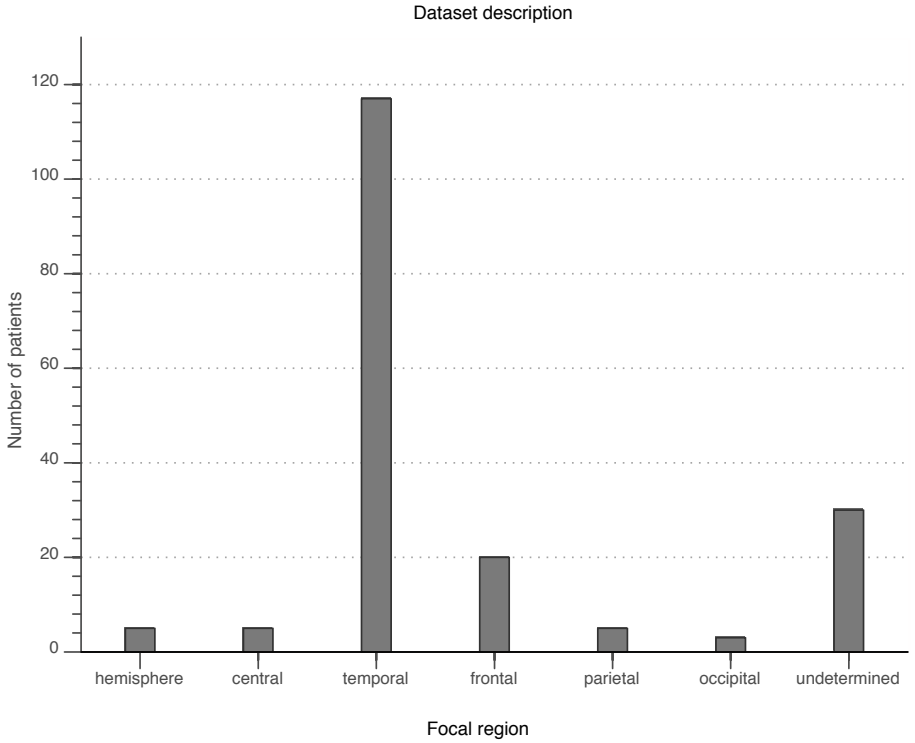


Figure 3.5 – Dataset analysis. 63% of the patients presented temporal epilepsy. Frontal epilepsy represented approximately 11% of the patients.

For the purpose of prediction, we only consider seizures that were identified as clinical seizures using the standardized annotation protocol. The seizure onset time is defined as the first electrographic change related to upcoming seizures. If not identifiable using the EEG data, the seizure onset time is defined as the clinical onset of the seizures (using video-EEG recordings).

3.2.2 Feature extraction

The prediction algorithm starts with the extraction of different univariate measures from the original EEG data.

The features were computed using a sliding window. In the literature the length of the window usually varies from 10 seconds to 40 seconds (Mormann et al. 2007). In this thesis, we assumed a window length of five seconds without overlap, i.e. consecutive five seconds windows were considered.

Different arguments were considered for the selection of the window length. Short segments provide a higher temporal resolution at the cost of a lower spectral resolution. On the opposite, longer segments increase frequency resolution while the temporal resolution decreases. Another important argument is the assumption of stationarity of the segment as "stationarity is a requirement necessary to allow the

metric algorithms of the nonlinear dynamics to be used in the analysis of EEG" (Blanco et al., 1995). According to the same authors, statistical tests revealed that the stationarity window ranges from several seconds to minutes depending on the problem being studies. We believe that a five seconds window represents a good trade-off.

The other parameter is the displacement between consecutive windows. We considered a non-overlapping window, so the shift of the window is five seconds. For the purpose of seizure prediction, it was considered that a five seconds interval is suitable to represent variations of the EEG data associated to seizure generation.

To minimize the impact of preprocessing in the information contained in the EEG data we restricted the preprocessing to filtering. To remove the power line artifacts at 50Hz, a notch filter was applied to the raw data. The filter consisted of a 4^{th} order Butterworth infinite impulse response (IIR) filter used to remove the frequency band between 48Hz and 52Hz.

A brief introduction to the 22 features used in the chapter is presented next, allowing the reader to have a deeper understanding of their relations with the raw EEG signal and to better analyze the fundamentals of the classifiers and results. These univariate measures represent computational efficient features, with potential for online implementation (Teixeira et al., 2011). Table 3.1 summarizes the set of features used and their computational times obtained in a computer running LINUX (Ubuntu) with a Intel® Core 2 Duo 2.4 GHz processor with 4 GB of RAM.

3.2.2.1 *Auto-Regressive (AR) modeling predictive error*

The prediction error is the difference between the value predicted by an autoregressive model and the actual value (Pardey et al., 1996). To make a prediction, we used an AR model (see equation (2.12)) with order 10 using the Burg maximum entropy method (Ulrich, 1972). The mean square prediction error is used as feature.

3.2.2.2 *Decorrelation time*

The decorrelation time is determined using the autocorrelation function (see equation (2.11)). It corresponds to the first zero crossing of the autocorrelation function (Box and Jenkins, 1990) and is an indicator of data periodicity.

3.2.2.3 *Energy*

The energy is computed by the average of the squared values of the EEG time series.

3.2.2.4 *Hjörth*

Hjörth (1970) proposed three parameters to quantify the dynamical properties of a given signal. The Hjörth mobility and complexity (see equations (2.9) and (2.10), respectively) reflect some properties of spectral information of the signal and have been related to significant variations in the pre-ictal period (Mormann et al., 2005).

3.2.2.5 *Spectral power*

Five main spectral bands are defined in classical EEG analysis. Variations in the spectral power of EEG signal have been related to epilepsy. The redistribution of spectral energy on the seizure onset is well documented (Shoeb, 2009). However, Mormann et al. (2005) reported that variations might also occur prior to seizure onset.

We compute five features corresponding to the main spectral bands: delta δ (0.5 - 4 Hz), theta ϑ (4 - 8 Hz), alpha α (8 – 15 Hz), beta β (15 – 30 Hz), and gamma γ (30 Hz – Nyquist frequency). First, we compute the FFT of the EEG signal within the five seconds window. The normalized spectral power in a determined frequency band corresponds to the average of the squared FFT coefficients in that frequency band divided by the total spectral power (3.19).

$$
\begin{aligned}
\delta_{rp} &= \frac{1}{P}\sum_{f=0.5\,Hz}^{4\,Hz} p_f \\
\theta_{rp} &= \frac{1}{P}\sum_{f=4\,Hz}^{8\,Hz} p_f \\
\alpha_{rp} &= \frac{1}{P}\sum_{f=8\,Hz}^{15\,Hz} p_f \\
\beta_{rp} &= \frac{1}{P}\sum_{f=15\,Hz}^{30\,Hz} p_f \\
\gamma_{rp} &= \frac{1}{P}\sum_{f=30\,Hz}^{NyquistFreq.} p_f
\end{aligned}
\tag{3.19}
$$

where p_f is the squared of the FFT coefficient at frequency f, and P is the total spectral power in all bands.

3.2.2.6 *Spectral edge frequency and power*

Most of the spectral power is comprised in the 0.5 Hz – 40 Hz band. We define spectral edge frequency as the frequency below which 50% of the total power of the signal is located. The spectral edge power is the value of the power existing below the spectral edge frequency (Drummond et al., 1991).

3.2.2.7 *Statistics*

The first four statistical moments are used to describe the data within each EEG segment. We use the formulation presented in Chapter 2 (see equations (2.1), (2.2), (2.3), and (2.4)).

3.2.2.8 *Energy of wavelet coefficients*

An alternative to the short-time Fourier transform to evaluate the time-frequency content of a signal is the wavelet transform. The wavelet coefficients quantify the correlation between the signal and different translations and scales of the mother wavelet (the mother wavelet used was the daubechies-4) (Semmlow, 2004). The wavelet transform was used to decompose the signal in six decomposition levels (corresponding to different frequency bands). The features correspond to the average of the squared wavelet coefficients of each level.

Table 3.1 – Set of the 22 univariate features used in this study. The computation time indicated was estimated using a five seconds window for a single channel (the values present the quotient $\textit{window length (5 s)} \,/\, \textit{computation time}$), in computer running LINUX (Ubuntu) with a Intel® Core 2 Duo 2.4 GHz processor with 4 GB of RAM. The results show that the features are computable in real time (adapted from Teixeira et al., 2011).

Feature		**Computation time (times faster than window length)**
AR modeling predictive error		1000
Decorrelation time		1162.8
Energy		6250.0
Hjörth	Mobility Complexity	357.1

Spectral power	Delta band (0.5 – 4Hz) Theta band (4 – 8Hz) Alpha band (8 – 15 Hz) Beta band (15 – 30 Hz) Gamma band (30 Hz – Nyquist freq.)	384.6
Spectral edge	Power Frequency	609.8
Statistics	1st moment – Mean 2nd moment – Variance 3rd moment – Skewness 4th moment - Kurtosis	943.4
Energy of wavelet coefficients	Six decomposition levels	192.3

Channel selection

It has been reported that, despite the general interest in the development of closed-loop seizure intervention devices based on seizure prediction algorithms, only part of the patients would consider wearing an ambulatory prediction device (Schulze-Bonhage et al., 2010). Even fewer patients would accept wearing a high number of scalp EEG electrodes. The discomfort and discrimination motivated by a complete array of electrodes is an important factor for these results. In the EPILEPSIAE project six electrodes were considered a good tradeoff between information and patient comfort. Other studies also follow this tradeoff for seizure prediction (Schelter et al., 2006a; Park et al., 2011; Mirowski et al., 2009).

Three choices of the six electrodes were decided. The first choice of electrodes makes a discretization of the patient scalp in order to get a general view. It includes electrodes F7, FZ, F8, T5, PZ and T6 (nomenclature according to the international 10-20 system) and aims to analyze the brain activity of the central, frontal and temporal areas (this choice is identified as *10-20*). The second choice is based on the seizure onset zone, the *focal* region. Three electrodes are selected within (or as close as possible) the focal region and the other three electrodes are selected over regions not involved in seizure generation according to trained epileptologists' opinion. The third choice of electrodes corresponds to a *random* selection of six electrodes among the electrodes available for each patient.

The 22 features extracted from each one of the six channels are then concatenated to form a feature vector with 132 (22 times 6) elements that encodes information from each feature and from the spatial relations between channels.

Finally, the feature vector is assigned with one of the four labels: *pre-ictal*, *post-ictal*, *ictal*, and *inter-ictal*,

There is no clinical definition of pre-ictal interval. Mormann et al. (2007) concluded that electrophysiological changes might develop minutes to hours before the actual seizure onset. This claim suggests that the pre-ictal period is probably patient-specific. The approach followed in the EPILEPSIAE project and adopted in this thesis consists on the consideration of different intervals: 10, 20, 30 and 40 minutes.

The duration of the post-ictal period is also an open issue. Fisher and Engel (2010) defined the post-ictal state as an "abnormal condition occurring between the end of an epileptic seizure and return to baseline condition", but also concluded that the existence of boundaries is not always clear.

According to Kaibara and Blume (1988) only a fraction of the patients revert immediately to an inter-ictal EEG behavior; a post-ictal change is apparent in 69% of the patient analyzed. The duration of the post-ictal period varied between 406 seconds for multiple effects (more than one phenomenon) and 79 seconds for single changes.

To accommodate these changes, we selected the post-ictal state as the 10 minutes period between the end of a seizure and the returning to the normal, inter-ictal behavior.

3.2.3 Dimension reduction for an intuitive visualization of the separability of the classes

Dimension reduction of the feature space is an important tool in pattern recognition. The complexity and high dimensionality of the feature sets difficults the task of classification algorithms and their convergence to a global minimum. Multivariate analysis methods such as dimension reduction try to find transformations to reduce dimensionality and ease the interpretation of the datasets (Semmlow, 2004).

Different approaches have been proposed to reduce the complexity of the feature sets, either selecting features according to a given principle or projecting the data into a new, reduced space where human intuitive visualization of data is possible, as for example in a two- or three-dimensional space.

Pearson (1901) proposed the pioneer principal component analysis (PCA) as a data exploratory tool. The main idea is to project each feature vector into another basis, to obtain a more meaningful representation of the data. The PCA finds, by a linear transformation, an orthonormal basis where each axis represents the directions in which the data show higher residual variance. The direction showing higher variance is represented in the first axis. Then iteratively it finds directions orthogonal to all previous directions, along which the remaining variance is the highest. The assumption is that the signal-to-noise ratio is higher in the first axes (higher data dispersion) and that the orthogonal constraint minimizes redundancy among axes.

3.2.3.1 *Multidimensional Scaling*

Multidimensional scaling (MDS) (Kruskal, 1964) represents an alternative exploratory approach to data analysis. The method is based on the idea of preserving, in different dimensions, the dissimilarity between patterns, mapping the original patterns into a new (lower-dimensional) basis preserving as much as possible their dissimilarities in the original $\Re^{N_{feat}}$ space. The choice of the lower-dimensional $\Re^{N_{mds}}$ space is usually lower than four to allow the dataset visualization.

Let us consider the dissimilarity matrix $\mathbf{D}$, containing the pairwise distances in the original high dimensional space (3.20),

$$\mathbf{D} = \begin{bmatrix} d_{1,1} & \cdots & d_{1,N} \\ \vdots & \ddots & \vdots \\ d_{N,1} & \cdots & d_{N,N} \end{bmatrix} \tag{3.20}$$

where $d_{i,j}$ represents the distance between the *ith* and *jth* points, and N is the total number of points. The idea is to create a new configuration based on a criterion by which it is possible to evaluate how well this configuration represents the data, the Kruskal stress function (Kruskal, 1964) (3.21). Assume that $\hat{d}_{i,j}$ is the distance between the *ith* and *jth* points in the reduced space and $d_{i,j}$ the distance between these same points in the original high-dimensional space. Then a measure of goodness of fit of the new configuration would be (3.21).

$$stress = \sqrt{\frac{\sum_{i,j=1}^{N_P} (d_{i,j} - \hat{d}_{i,j})^2}{\sum_{i,j=1}^{N_P} d_{i,j}^2}} \tag{3.21}$$

Zero stress would represent a perfect match between the original dissimilarities and their equivalent in lower dimensional configuration, which in practice never occurs.

The best configuration is obtained minimizing the function (3.21). To solve the minimization problem different approaches from multivariate optimization theory may be used.

We applied MDS to a set of patterns composed by univariate measures to study the structure of the data before the seizure onset. According to different authors, the pre-ictal period may vary between minutes and a few hours (Mormann et al., 2007). To support the selection of a pre-ictal period we investigated the separability of the patterns according to the time interval until the seizure onset using MDS.

To account for the variability among patients, the method was applied for different patients considering the features set described in Section 3.2.1.

The data was labeled in order to determine the temporal evolution towards the seizure onset. Several time intervals were established: (1) inter-ictal, (2) 40 to 30 minutes to the seizure onset, (3) 30 to 20 minutes to the seizure onset, (4) 20 to 10 minutes before the seizure onset, (5) 10 minutes before the seizure onset, (6) the ictal and (7) post-ictal (the 10 minutes after the seizure offset).

The visual inspection in the three-dimensional space obtained using MDS revealed interesting characteristics.

The example presented in Figure 3.6 corresponds to a 10-hour segment of scalp EEG that includes one seizure, in which each feature vector has the dimension of (N_{feat} =) 594 elements (27 channels times 22 features).

We observed that the three-dimensional data of the pre-ictal and inter-ictal patterns tend to become separated 40 minutes before the seizures.

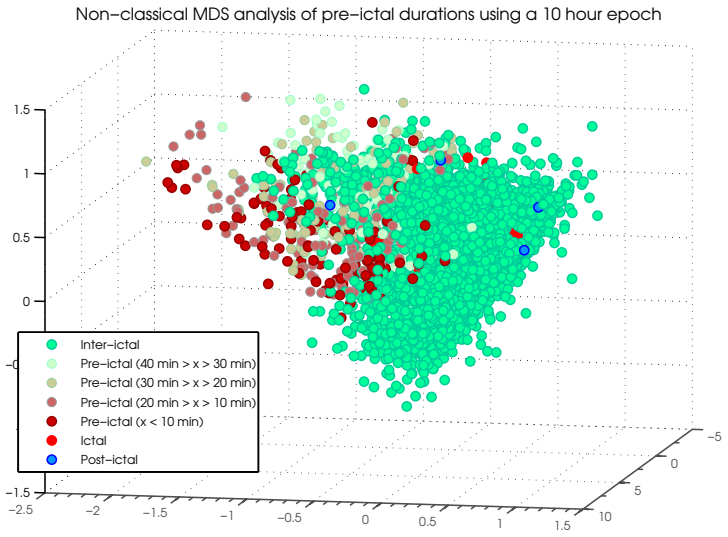


Figure 3.6 – Representative example of MDS over a 10 hour segment of scalp EEG data containing inter-ictal, pre-ictal (divided into four periods of 10 minutes), ictal and post-ictal data; the first three dimensions are presented. Pre-ictal data closer to the seizure onset tends to separate from the rest of the inter-ictal data. Light green representing the interval ranging from 40 minutes to 30 minutes to the seizure onset shows the first indications of some data variability.

Similarly, we analyzed iEEG to observe the structure of the data in the expected pre-ictal period. The example presented in Figure 3.7 shows a 10-hour segment of iEEG data also with one seizure, and each feature vector has ($N_{feat}=$) 1056 elements (48 channels and 22 features).

The three-dimensional representation seems to suggest that the patterns labeled as 10 minutes before the seizure onset (5), are different compared to the rest of the labels.

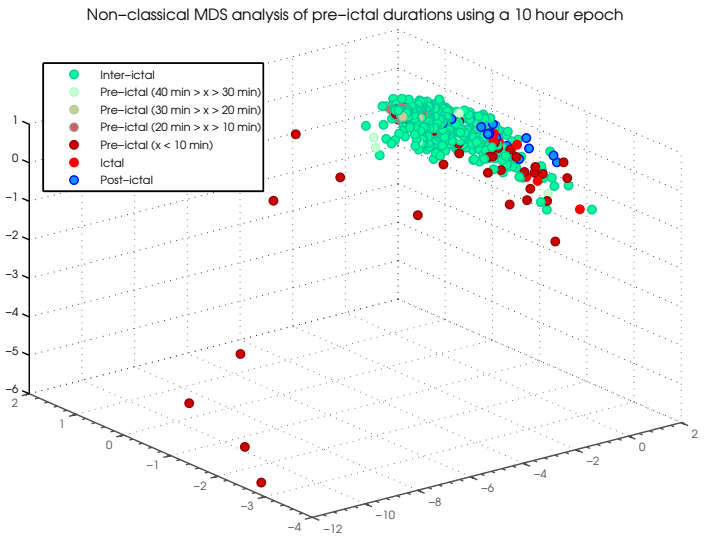


Figure 3.7 – Representative example of MDS over a 10 hour segment of intracranial EEG containing inter-ictal, pre-ictal, ictal and post-ictal data; the first three dimensions are presented. Pre-ictal data closer to the seizure onset (10 minutes) tends to separate from the rest of the inter-ictal and pre-ictal data, while the rest of the data is agglomerated.

This intuitive observation about the duration of the pre-ictal period was used to support the set of possibilities for the seizure prediction scheme proposed throughout this work.

3.2.4 Classification

Given the number of choices of electrode subsets (three) and the different pre-ictal periods considered (four), the total number of datasets created for each patient was 12.

3.2.4.1 *Training the SVM model*

The training set consists in a segment of data that encloses three (when available) or two seizures (Figure 3.8). In some cases, the total number of seizures of the patient's dataset is three, therefore the training set is restricted to two seizures and the other is part of the testing set.

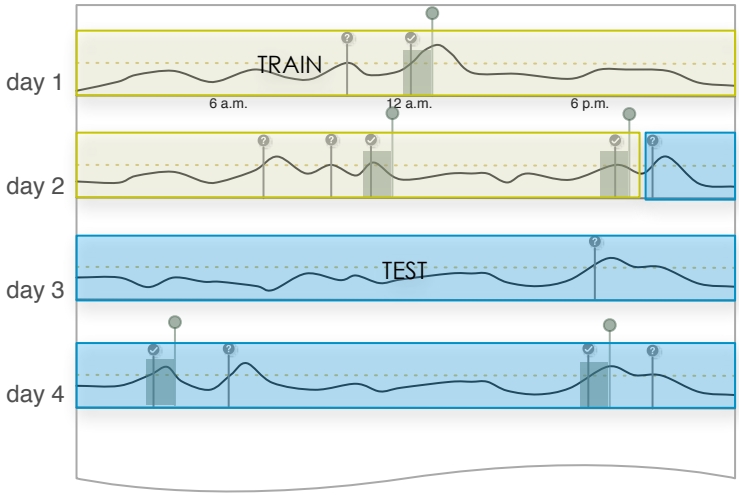


Figure 3.8 – Schematic representation of the method used to divide the datasets. The data was divided in two parts: the training set consists of the segment of data that encloses the first two or three seizures (represented as the green unsigned circles) and the testing set consists of the rest of the data. The training method used (three-fold cross validation) further divides the training set into three folds used as training and validation set iteratively. The check symbols represent true alarms and the question marks represent false alarms.

Classification target

The classification target is defined according to the pre-ictal period defined. Samples from the inter-ictal period are assigned the number '1', samples from the pre-ictal period are assigned the number '2', samples from the ictal period are labeled with number '3' and the post-ictal samples are labeled as '4'. In the ictal stage, we include a 10 seconds period before the seizure onset (this period is equivalent to the intervention time (Winterhalder et al., 2003) necessary to potentiate a therapeutic intervention).

A multiclass SVM is trained to classify accordingly to the suggested target.

Unbalanced datasets

Unfortunately for classification, the data presents a structural problem. The training population is unbalanced; the ictal, post-ictal and pre-ictal classes have relatively small portions compared to the inter-ictal class. The inter-ictal class may represent more than 90% of the training data. A model predicting only inter-ictal labels would present an overall accuracy of 90%, which is clearly a misleading measure.

One of the approaches suggested to overcome the class unbalance is to undersample the class with larger proportions (Chawla et al., 2004; Qiao and Liu,

2009). Thus, to reduce the effect of the class unbalance, we randomly select samples from the inter-ictal class (matching the number of samples of the inter-ictal class with the sum of the rest of the classes).

Cross-validation

According to Guyon and Elisseeff (2003), for an appropriate training scheme, the training set should be further divided into training and validation sets to perform cross-validation analysis, in order to prevent the over-training that limits the generalization capability. We used the cross-validation approach to select the optimal SVM parameters based on a grid-search approach.

Cross-validation consists of removing one *fold* (one fraction of the data) constructing the predictor based on the remaining data (or folds), and then validating on the removed fold. The process is repeated for each fold. The classification results of the different folds are averaged and we assume that the best results obtained indicate the optimal SVM parameters. Then, using the parameters selected, the whole training set is used to build the SVM model. The number of folds to use in the cross-validation analysis is an open problem, but due to the nature of the training set we used a three-fold cross-validation scheme (most of the training sets contain three seizures).

The final step for selecting the best model is the definition of asymmetric soft margins in the SVM. The idea is to impose a different weight for each class so that the classes with fewer points cannot be easily ignored. A linear learning rate is used to optimize the results through a hill-climbing algorithm (the starting point is the one obtained from the cross-validation).

Model optimization

The SVM is trained using the libSVM software package (Chang and Lin, 2011). The framework allows the implementation of cross-validation procedures as well as asymmetric costs for each SVM. We used the linear kernel, which is computationally lighter than non-linear kernels (such as radial basis functions) while presenting comparable performance in relatively high dimensional spaces (Song et al., 2011).

Firstly, we optimized the cost C (see equation (3.15)) (that represents the tradeoff between the classification margin and non-separable patterns) using the *3-fold* cross-validation over an array of possible values $C \in [2^1, 2^4, 2^7, 2^{10}, 2^{13}, 2^{16}]$. Secondly, we refine the search around the best candidate, C_0 using the array of

values $C \in C_0 \times [2^{-1.5}, 2^{-0.5}, 2^{0.5}, 2^{1.5}]$. Finally, we use weighted learning to differentiate the classes. We settled a lower $C_{\text{inter-ictal}}$ value for the inter-ictal class optimizing the value with a learning rate of 0.1. The method is illustrated in Figure 3.9.

The selection of the best model in training was made using a point-by-point analysis (due to the random selection of 'inter-ictal' patterns). To characterize the model output in the training set we estimated the *F-measure*. The best model is selected accordingly to the averaged *F-measure* score obtained in the cross validation scheme.

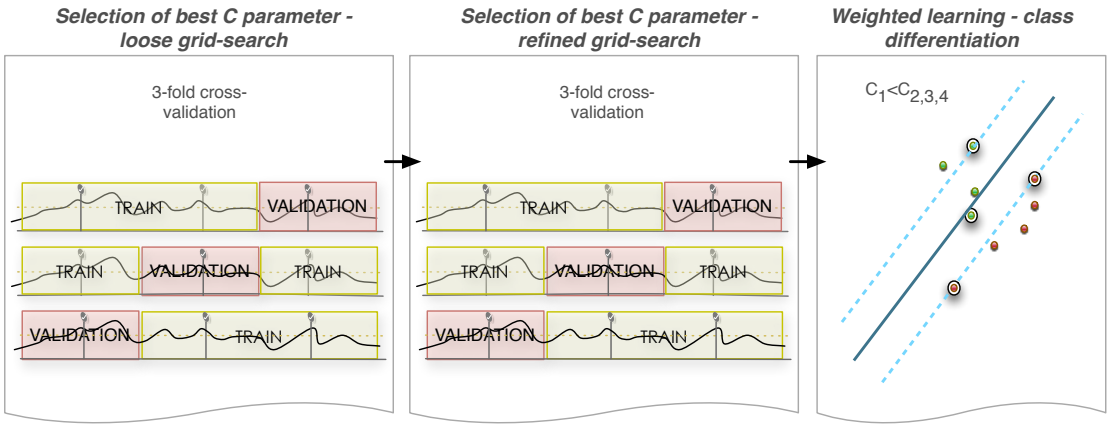


Figure 3.9 – Overview of the training scheme. Firstly we use a loose grid-search and a 3-fold cross validation to select the best candidate C_0. Then, we use a refined grid-search C_0 around to select the best cost. Finally, we use a hill-climbing method to determine an optimal lower $C_{inter\text{-}ictal}$. The best model in each step is determined based on classification results in terms of the *F-measure*.

The *F-measure* F_β is defined as (3.22),

$$F_\beta = \frac{\left(1+\beta^2\right)\times TP}{\left(\left[1+\beta^2\right]\times TP + \beta^2 \times FP + FN\right)} \tag{3.22}$$

where $\beta = 0.5$, TP are the true positives, FP are the false positives and FN are the false negatives. The *F-measure* is especially important in the evaluation of biased datasets (Barandela et al., 2003).

At the end of the training stage, we have 12 models per patient, one for each dataset.

3.2.4.2 *Testing the model*

Each sample is evaluated using the trained SVM model and is labeled as one of the four possible classes, pre-ictal, ictal, post-ictal and inter-ictal.

In theory, an alarm could be raised whenever the output of the model is pre-ictal. However, this behavior does not happen in practice (Chisci et al., 2010; Park et al., 2011; Temko et al., 2011). According to Temko et al. (2011) the metrics used to assess the performance of algorithms "can be divided into epoch-based and event-based". In the first approach, every output of the model is considered individually and due to the number of isolated false positives (possibly not related to a general trend but to outliers) the approach may be application irrelevant. The epoch-based metrics may not transform into a good alarm generator method. Event-based metrics are thought to translate the behavior of the model designed for a particular application. Two scores are usually considered: the sensitivity (ability to detect correctly a seizure) and false positive rate per unit of time (alarms generated outside the pre-ictal period considered).

A post-processing step consists of transforming the output of the model to an appropriate alarm generator method. The approach *firing power* proposed by Teixeira et al. (2011) uses a sliding window with size equal to the pre-ictal period considered and is an event-based technique conceived for seizure prediction (Figure 3.10). The output of the classifier is transformed in a binary vector with labels 'pre-ictal' (1) and 'non pre-ictal' (0). The number of the pre-ictal elements is computed and an alarm is raised whenever the number of points classified as pre-ictal exceeds 50% in the sliding window. A value of zero would represent that no samples were identified as 'pre-ictal' during the analyzed window.

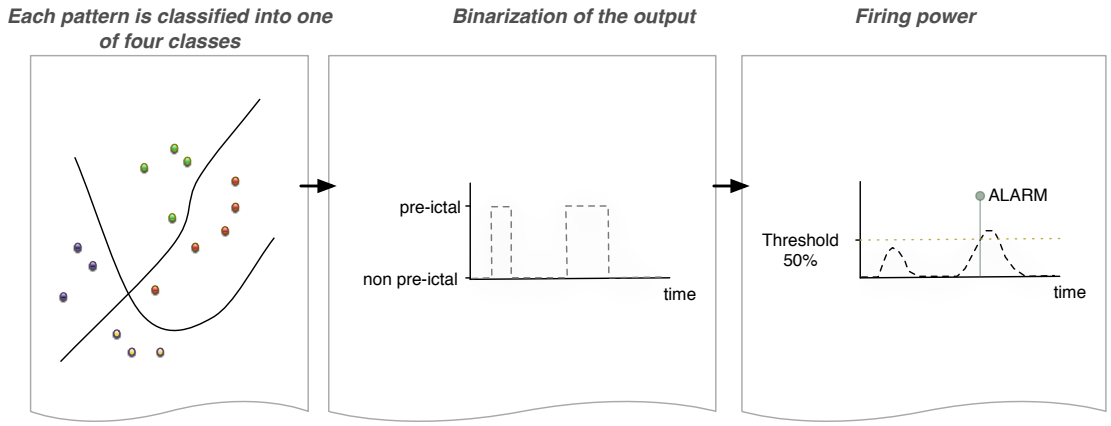


Figure 3.10 - Overview of the post-processing steps. Firstly we classify the data into one of four classes, pre-ictal, ictal, post-ictal and inter-ictal. The second step is the binarization of the classification output. Finally we compute the firing power.

3.2.5 Performance evaluation

An important aspect in seizure prediction studies is the ability to compare different approaches. A common framework gives an important contribute to validate new approaches. In this work we followed the seizure prediction characteristic (SPC) suggested by Winterhalder et al. (2003) that was based on clinical, behavioral and statistical criteria (Section 2.3.3.1).

SPC terminology

To characterize the output of the predictor according to the SPC, we need to define the seizure occurrence period (SOP) and the seizure prediction horizon (SPH). SOP represents a time period during which the seizure eventually would happen, which is equivalent to our definition of pre-ictal period (we consider four different periods, 10, 20, 30 and 40 minutes).

The SPH is a time window (between an alarm and the seizure onset) necessary to render a therapeutic intervention or to warn the patient. A correct prediction has to take both periods into account. SPH is defined as a 10 seconds window (we interpret the period of 10 seconds before the seizure onset as part of the ictal period to accommodate the SPH).

The sensitivity of the predictor is the fraction of correct predictions divided by the total number of seizure. It is defined as (3.23).

$$Sensitivity = \frac{seizures\ correctly\ predicted}{total\ number\ of\ seizures} \tag{3.23}$$

In practice false predictions are impossible to avoid. To quantify their occurrence, the specificity is represented as the rate of false positives per hour. The score is obtained dividing the total number of false positives by the total inter-ictal duration present in the testing set subtracting the time under false warning. It is defined as (3.24).

$$FPR.h^{-1} = \frac{false\ positives}{interictal\ \ duration\ -\ \ false\ \ positives\ \times\ SOP} \tag{3.24}$$

Statistical evaluation

To statistically evaluate the results, we used the analytical random predictor method (Schelter et al., 2006a; Feldwitch-Drentrup et al., 2010). It quantifies the

sensitivity of the analytical random predictor, given a specified SOP, a false positive rate, and the number of degrees of freedom.

We followed the principles discussed in Schelter et al. (2006a). According to these authors, the unspecific random predictor does not uses EEG data and therefore "only the number of independent extracted features influences the critical value". In this work, we interpret the models' output as the independent variables. Furthermore, we assume the upper critical value, i.e. the complete independence of the features.

Let us consider K the number of seizures in the testing set, k the number of predicted seizures, z the number of independent features (i.e. number of models tested) and P_{SOP} defined according to (2.23). The probability $P_{binom,z}$ of predicting k seizures by means of at least one of the 12 models is given (3.25).

$$P_{binom,z=12}\{k;K;P_{SOP}\} = 1 - \left(\sum_{j<k} \binom{K}{j} p^j \left(1 - P_{SOP}\right)^{K-j} \right)^{12} \tag{3.25}$$

The statistical test to determine the critical sensitivity of the random predictor is defined by (2.26). The statistical significance α_p used in this study was 5%. Hence, the statistical test is defined as (3.26).

$$critical\ value_{rand,z=12} = max_k (P_{binom,z=12}\{k;K;P_{SOP}\} > \alpha_p) \times 100\% \tag{3.26}$$

According to Schelter et al. (2006a), if the sensitivity of the predictor is superior to the upper critical sensitivity of an unspecific random predictor, it is concluded that the predictor is significant.

Lastly, one has to take into account that statistical significant results can be obtained for individual patients by chance. A binomial test including all the patients is also performed to test the number of statistically significant cases. The level of statistical significance used in the statistical tests was 5%, i.e. to conclude that the study is statistically significant and the results are above chance level, the *p*-value of the test has to be inferior to 0.05.

3.2.6 Results

We performed the analysis of the 12 models per patient and compared their performances against the performance of a random predictor. Appendix A presents the performance of the best model for each patient.

In certain cases, two (or more models) present similar sensitivities and FPR.h^{-1}, consequently presenting identical *p*-values (3.26). To determine the best model, we consider the best performance in terms of epoch-based metrics, in this case the *F-measure*. In other words, in the case of a tie considering event-based metrics, we considered epoch-based metrics to determine the best model (Temko et al., 2011).

The testing sets involve 932 seizures. The total number of correctly predicted seizures was 386 (41.42%) and the average false prediction rate was 0.21 h^{-1}.

3.2.6.1 *Significance*

We found that the performance for 19 patients (10.27%) was above the upper critical sensitivity presented by a random predictor. The results were determined statistically significant according to a binomial test at the group level (5% significance level). The *p*-value obtained was 0.0025.

To study the influence of the different variables in the results (localization and lateralization, SOP and selected set of electrodes), we performed a statistical evaluation of the performance measures. First we made a preliminary evaluation to test if the data comes from a standard normal distribution: the one-sample Kolmogorov-Smirnov goodness of fit test (the method considers the null hypothesis that the data originated from a normal distribution) (Massey, 1951). Since none of the metrics evaluated presented a normal distribution (the null hypothesis was rejected), the Kruskal-Wallis (KW) test (a non-parametric alternative to ANOVA) was used to compare the samples from the different groups (Montgomery and Ruger, 2003). The test compares the medians of the groups, and estimates a *p*-value for the null hypothesis that all the data is extracted from the same distribution. The significance level was set as 5%.

The statistical analysis is used in three measures: alarm sensitivity, FPR.h^{-1} (event-based metrics), and *F-measure* (epoch-based metric).

Localization and lateralization

The patients analyzed in this study presented different types of epilepsy. Seven labels were defined considering the localization of the epileptic focal region: hemisphere, central, temporal, frontal, parietal, occipital and undetermined (insufficient information about the patients to determine a label). Figure 3.11 presents a summary of the results in terms of alarm sensitivity and false positive rate grouped according to the type of epilepsy. Higher sensitivities are obtained

when the epileptic focus is located in the parietal (only five patients meet these conditions) and frontal lobes (20 patients). Lower $FPR.h^{-1}$ are associated to occipital (five patients) and central (only one patient) lobe epilepsies.

From the 19 statistically significant patients, 12 presented temporal lobe epilepsies.

According to the statistical analysis, none of the metrics analyzed presented significant differences among the locations considered.

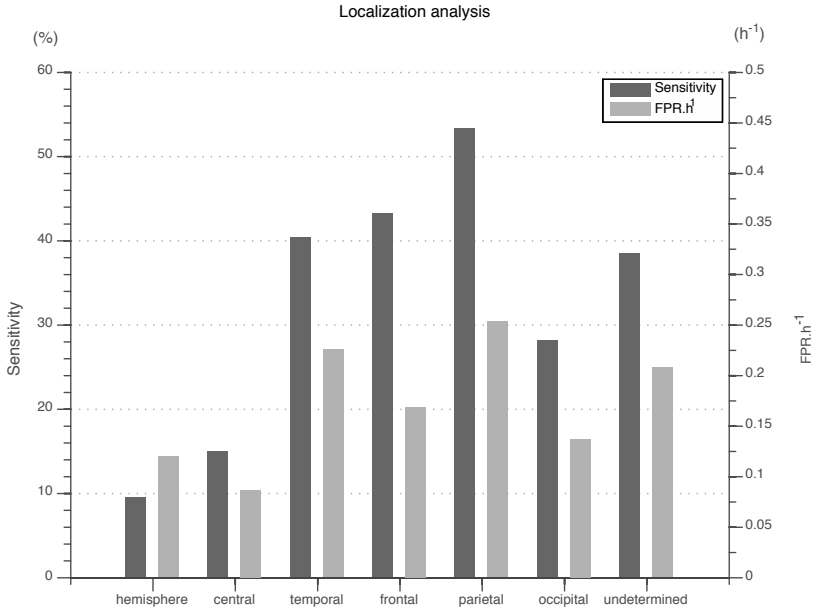


Figure 3.11 – Sensitivity and $FPR.h^{-1}$ grouped according to the type of epilepsy. The best sensitivity is obtained for patients suffering from parietal lobe epilepsy and the lower $FPR.h^{-1}$ are presented when the focal origin lies in the central region.

Figure 3.12 presents the results concerning the lateralization of the epileptic focus. The bilateral configuration presents the highest sensitivity but simultaneously a high false positive rate. Right and left lateralized epilepsies present similar sensitivities; however the right side presents a lower number of false positives.

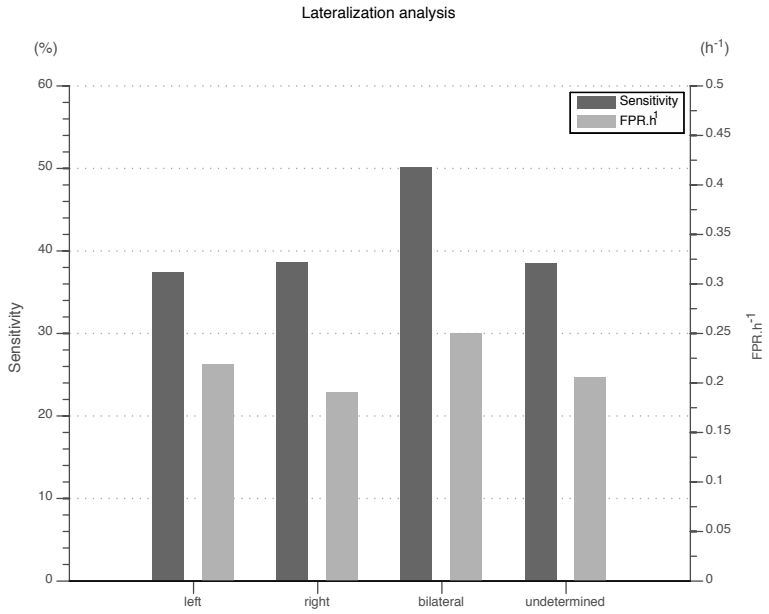


Figure 3.12 – Performance measures grouped according to the lateralization.

We evaluated possible differences among the three performance measures grouped according to the lateralization (*F-measure*, alarm sensitivity and false positive rate). The KW test (at the significance level of 5%) determined that the *F-measure* between groups is significantly different.

SOP

We used four different pre-ictal periods (SOP). The best models considered a SOP of 10 minutes for 35 patients, 20 minutes for 37 patients, 30 minutes for 51 patients, and a SOP of 40 minutes was found in 62 patients. The SOP of the best models tends to indicate that longer pre-ictal ranges represent a better fit to the predictor.

From the 19 patients that presented results statistically significant, three datasets were based on 10 minutes SOP, four on 20 minutes, six on the 30 minutes and finally six on the 40 minutes.

The analysis of the average values of the sensitivity and false positive rate according to the SOP points toward a familiar notion in machine learning, the trade-off between sensitivity and specificity. Higher sensitivities obtained for shorter SOP are accompanied by high false positive rates. The longer the pre-ictal period considered the lower the sensitivities and false positive rates of the models (Figure 3.13).

The average pre-ictal period among the 185 scalp patients analyzed was approximately 27 minutes.

We evaluated possible statistical significant differences among the three performance measures (*F-measure*, alarm sensitivity and false positive rate) grouped according to the SOP. Two measures presented significant differences based on the KW test: the *F-measure* and FPR.h^{-1}.

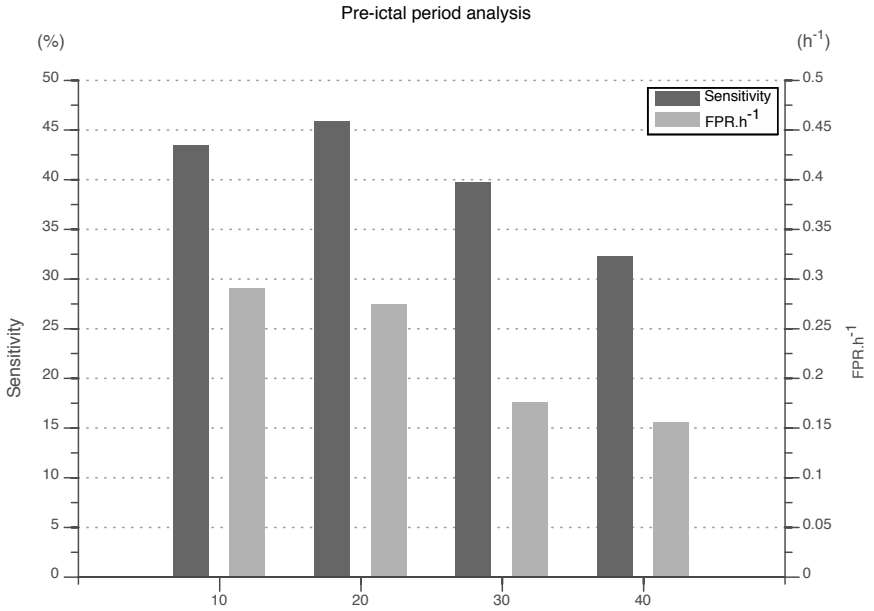


Figure 3.13 – Performance of the best models grouped by SOP duration. The SPH was determined as 10 seconds for every model. Shorter pre-ictal periods present higher sensitivity, however accompanied by a high FPR.h^{-1}.

Electrode selection

The selection of different sets of electrodes did not present any relevant change in performance. The best model for 66 patients was based on the *10-20* choice (scalp discretization), 51 based on the *focal* (three focal electrodes and three extra focal electrodes) and 68 patients based on randomly selected electrodes. From the 19 patients that presented results statistically significant, nine datasets were based on *random* electrode selection, seven based on *focal* electrodes, and three were based on the discretization of the *10-20* system.

Figure 3.14 summarizes the influence of electrode selection on the best models. The figure presents the average sensitivity and false positive rate grouped by the electrode selection (we considered the best models for each patient).

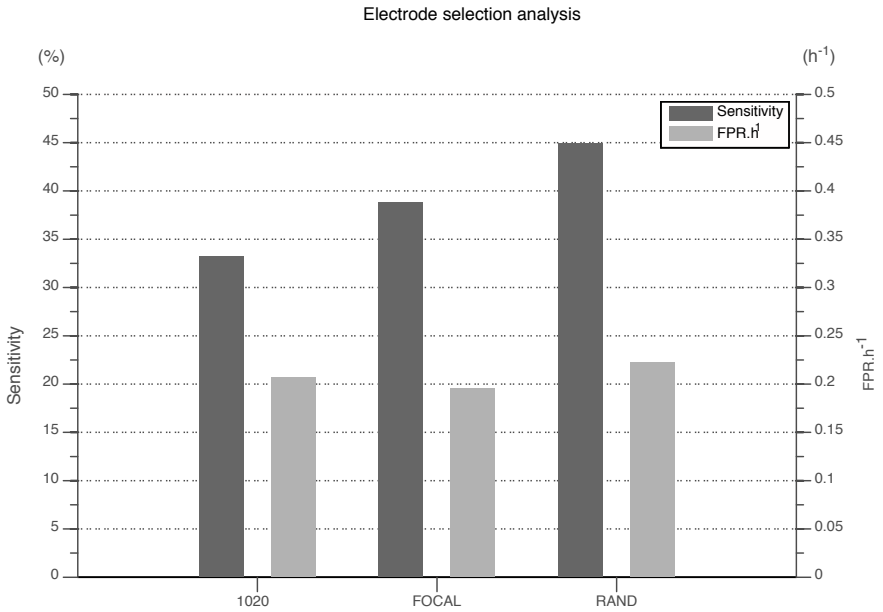


Figure 3.14 – Performance of the best models in terms of sensitivity and FPR.h^{-1}. The random selection presents the best sensitivity values. The FPR.h^{-1} does not seem to be influenced by the different electrodes selected.

The statistical analysis did not exhibited significant differences among the different electrode combination.

3.3 Invasive Recordings

The current approaches for seizure intervention devices (i.e. closed-loop model) such as the ones based on deep brain electrical stimulation and vagus nerve stimulation use invasive EEG recordings to measure brain activity. Intracranial recordings present the advantage of higher signal-to-noise ratio (and minimized volume conduction effect) and better spatial resolution. The analysis and comparison of the results obtained using intracranial recordings and scalp recordings, is an important aspect of seizure prediction.

For the patients with intracranial recordings, only two choices of six electrodes were used: *random* and the *focal* (three electrodes over the focal region and three from non focal regions). Considering four different SOP (10, 20, 30 and 40 minutes), the total number of models is eight per patient.

The feature extraction, classification, post-processing and statistical evaluation used are similar to the ones presented in Section 3.2.

3.3.1 Dataset description

The study using iEEG was performed on 31 patients from the EPILEPSIAE database (Ihle et al., 2012; Klatt et al., 2012), that were successfully trained using the strategy and parameters proposed in our approach (see Section 3.2.1).

Approximately 42% of the patients in this group presented focal temporal seizures (Figure 3.15).

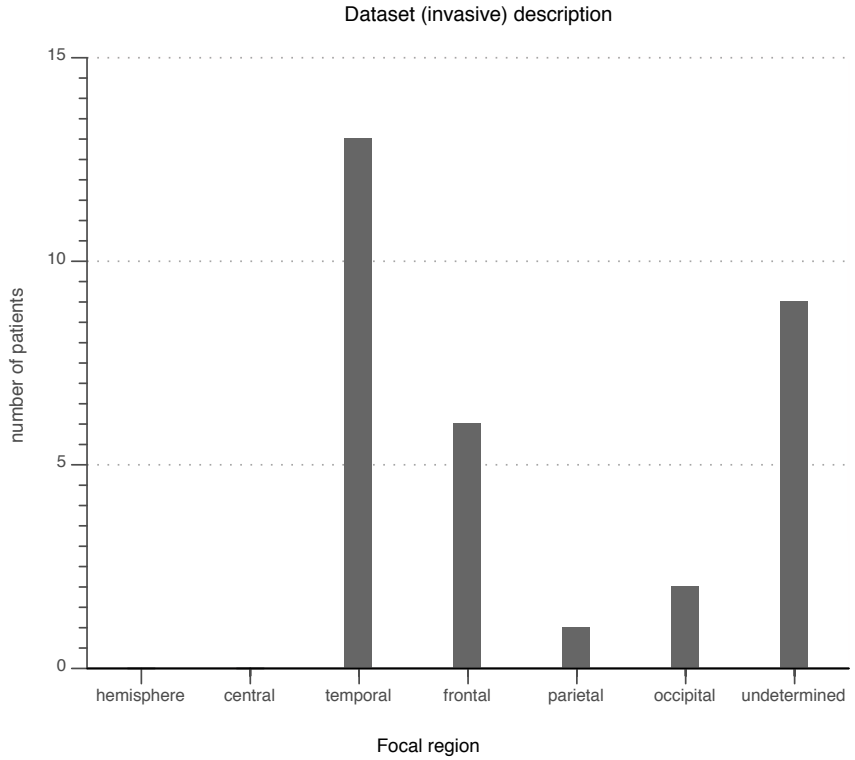


Figure 3.15 – Invasive EEG: 13 patients presented temporal epilepsy, 6 patients were diagnosed with frontal epilepsy.

3.3.2 Results

For each patient, we assessed the eight models performing the statistical evaluation based on the random predictor (3.25). The results are detailed in Appendix B.

The testing sets involve 274 seizures. The total number of correctly predicted seizures was 78 (28.47%) and the average false prediction rate was 0.23 h^{-1}.

3.3.2.1 *Significance*

The results obtained for five patients were above the upper critical sensitivity presented by a random predictor (16.13%). We applied a binomial test at the

group level (5% significance level) and obtained a *p*-value of 0.0179 thus allowing us to conclude that the results are statistically significant.

We performed statistical analysis similar to the one used Section 3.2.2.1: we evaluated the influence of the localization and lateralization, SOP and electrode selection according to alarm sensitivity, FPR.h^{-1}, and *F-measure*.

The preliminary normality test (one-sample Kolmogorov-Smirnov goodness of fit) allowed us to conclude that the data do not represent a normal distribution. Hence, to evaluate if any of the variables influence the performance values we used the KW test.

Localization and lateralization

From the five patients with results above chance level, two were diagnosed with temporal lobe epilepsy and two with frontal lobe epilepsy (one undetermined). Figure 3.16 summarizes the average results of the best models (grouped by the foal region). The dataset considered temporal, frontal, parietal (only one patient with sensitivity zero) and occipital lobe epilepsies. Based on the statistical analysis (KW test at the significance level of 5%), no significant differences were found.

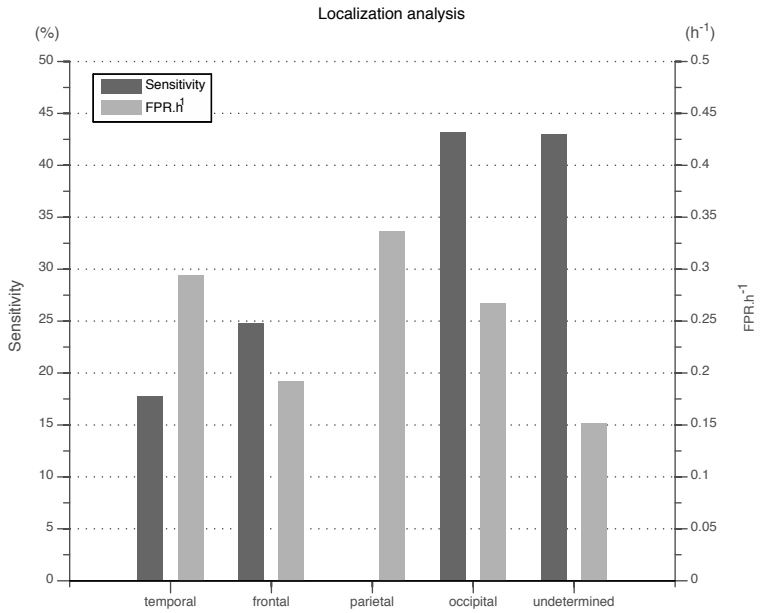


Figure 3.16 – Results obtained grouped by the location of the epileptic focus. The temporal and parietal lobe epilepsy patients present a low sensitivity while occipital lobe epilepsy patients (two patients) present a sensitivity of approximately 50%.

Concerning the lateralization of the focal region, right and left lateralization present similar results (Figure 3.17). Statistical analysis confirmed that no

significant differences exist between right and left hemispheres in terms of the performance of the predictors.

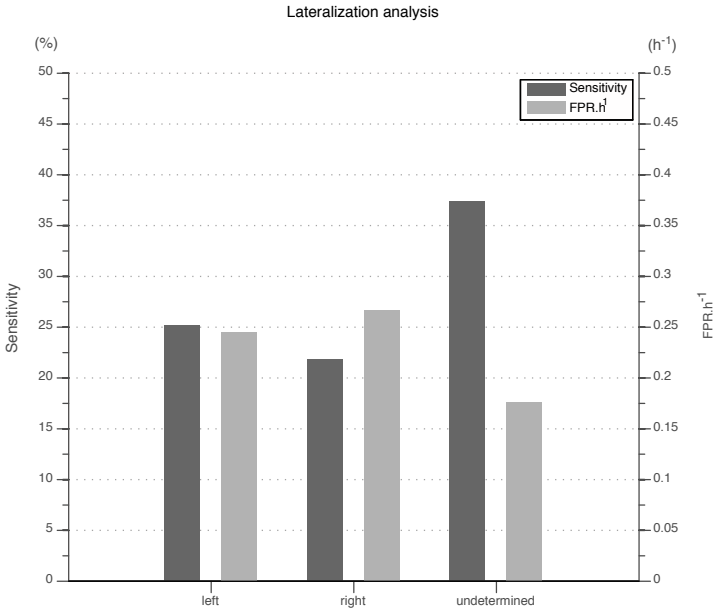


Figure 3.17 – Results grouped according to lateralization of the epileptic focus.

SOP

Similarly to the analysis performed using scalp recordings, we evaluated the influence of different SOP. Fourteen of the 31 patients analyzed had the best models based on the 40 minutes SOP (four based on the 10 minutes SOP, eight based on the 20 minutes and five based on the 30 minutes). The patients with results above the performance of a random predictor are distributed; one for the 10, 30 and 40 minutes SOP and two for the 20 minutes SOP.

The average results displayed in Figure 3.18, do not present a clear pattern (slight decrease of the average false positive rate associated to longer pre-ictal ranges). The average pre-ictal period is approximately 29 minutes.

The SOP variable seems to not influence the performance according to the statistical analysis.

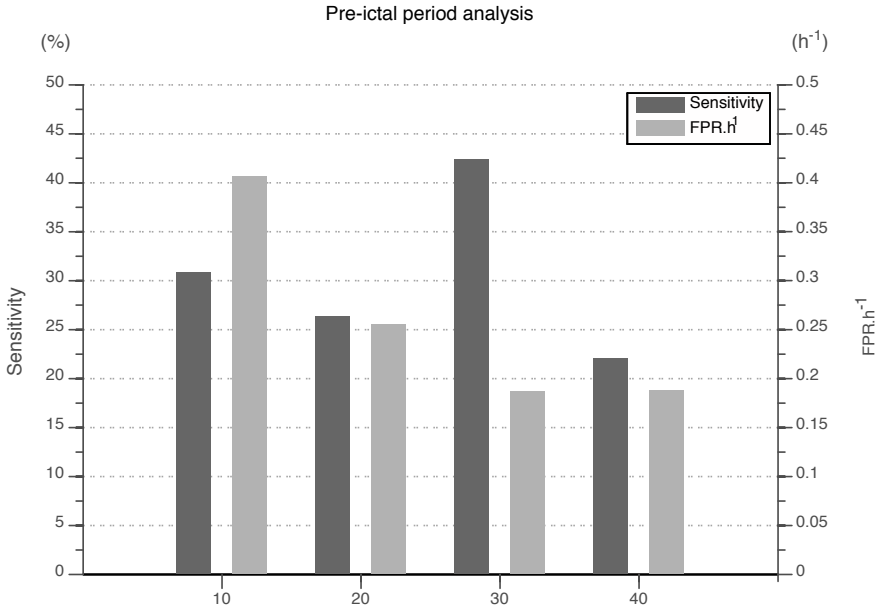


Figure 3.18 – Results based on the SOP considered.

Electrode selection

As mentioned before, only two sets of six electrodes were selected in invasive EEG. Among the best models for each patient, 19 were based on the *random* electrode selection and 12 based on the *focal* selection. Three out of the five cases presenting results above chance level were based on the *focal* selection and two based on the *random*. Figure 3.19 presents the average values grouped according to the electrode selection. No particular pattern was found.

The statistical analysis does not show significant changes among the different electrode selections.

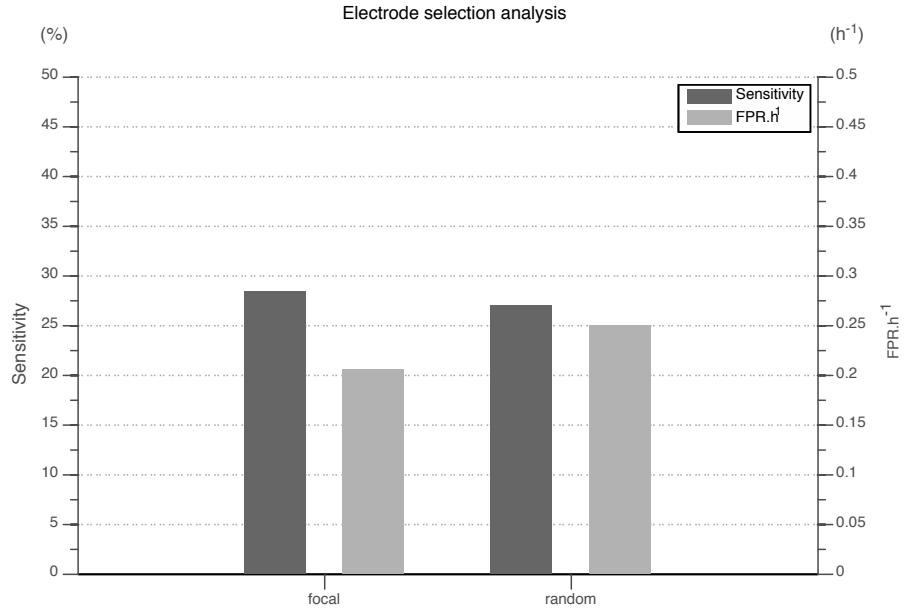


Figure 3.19 – Results in terms of sensitivity and FPR.h^{-1}, grouped by the electrode array selected.

3.4 Feature Selection

Despite of the recognition of the SVM algorithm as one of the most effective classification methods (Byvatov et al., 2003; Singla, 2011), high dimensional feature spaces may disturb their performance. In certain cases, the *curse of dimensionality* can be minimized using methods of feature selection (Chen and Lin, 2006).

Feature selection presents several benefits: facilitates data visualization and data understanding (ranking the feature importance according to a determined measure), reduce the computational requirements of the classification method, and reduce noise in order to improve classification accuracy (Ding and Peng, 2005). Feature selection methods are important to select the appropriate subset of features useful to build a good predictor (Guyon and Elisseeff, 2003).

Feature selection methods can be divided into two categories: the *filters* and the *wrappers*. A filter is a preprocessing step (and independent of the choice of the predictors) that ranks the features according to a defined measure. Wrappers utilize machine learning as an alternative to characterize subsets of features.

We applied two different feature selection techniques in order to obtain an optimal subset of features selected from a candidate set of electrodes and quantitative features. A filter method and a wrapper method were used to improve the classification performance and understanding about the feature sets.

The low-dimensional feature sets were used as the input of SVMs to predict seizures in out-of-sample data, i.e., the selection methods were used only in training data.

3.4.1 Comparison of feature selection methods in a dataset with 10 patients

The feature selection study was based on a subset of 10 patients with continuous long-term scalp EEG recordings from the EPILEPSIAE database. These patients were selected among the 185 patients considered in Section 3.2. This is a computational extensive study; therefore we only selected a subset of patients to investigate the feasibility and usefulness of feature selection in seizure prediction and to analyze the possible improvements of performance.

The patients were selected according to the p-value estimated (see equation (3.26)), i.e. we selected 10 patients whose results are close to the upper critical sensitivity value of a random predictor. The idea is to determine if feature reduction methods are able to contribute to the improvement of the results.

The dataset included a total of 106 seizures.

Using two feature selection methods (mRMR and SVM-RFE), we selected a subset of 66 features and classified according to the guidelines presented in Section 3.2.

The high dimensional datasets used consist of 22 univariate features per channel. Consider, for instance, an array of 27 electrodes; the total number of features considered in the original problem would be 594. The feature extraction scheme is similar to the one described in Section 3.2.2.

The feature selection methods aim the identification of a feature subset useful to optimize the seizure predictor and characterize all potentially important feature-electrode combination.

3.4.1.1 *minimum Redundancy - Maximum Relevance (mRMR)*

According to Peng et al. (2005) the condition for feature selection is to "identify the most characterizing features which is critical to minimize the classification error". However, the set of best features individually considered, may not translate into the best set of features when considered as a group. The method proposed by Ding and Peng (2005), intends to maximize the information relevant

to characterize the target while minimizing redundancy within the feature set and appears to well-suit the seizure prediction research.

One of the main shortcomings of common ranking approaches is that the subset of selected features can present high correlation among them. According to Peng et al. (2005), the selection of two "very effective" features does not form a better feature pair, mainly due to possible correlation between them. The redundancy among the subset of features presents two inconvenient: low efficiency (highly correlated features) and low broadness (the features would not represent the entire feature space covered by the original set, important for the ability of classifiers to generalize the results). The minimum redundancy criteria combined with the relevance factor tackles these problems.

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

Let us denote $\boldsymbol{\Omega}$ as the subset of selected features, F the p-value returned by the F-test, and N_{feat} the total number of features in the dataset. The first step of mRMR algorithm is the selection of the most relevant feature based on a statistical F-test (the hypothesis-test for the equality of two variances is used here as a relevance measure) (Montgomery and Runger, 2003). The features that remained unselected are ranked according to a score (*F-test correlation difference*) that takes into account relevance and redundancy. *F-test correlation difference* score is defined by (3.27),

$$mRMR_{score} = \max_{h \in H} \left\{ F(\mathbf{h},\mathbf{y}) - \left(\frac{1}{N_{\Omega}}\right) \sum_{j \in \Omega} \left| c_{hj} \right| \right\} \tag{3.27}$$

where N_{Ω} is the number of features in the subset $\boldsymbol{\Omega}$, H is the set of unselected features, and c_{hj} represents the cross-correlation between feature $\mathbf{h}$ and $\mathbf{j}$. $F(\mathbf{h},\mathbf{y})$ is the F-statistic computed between feature $\mathbf{h}$ and the target $\mathbf{y}$, and represents the relevance coefficient, while the term $\left(\frac{1}{N_{\Omega}}\right) \sum_{j \in \Omega} \left| c_{hj} \right|$ represents the average redundancy between feature $\mathbf{h}$ and the group of previously selected features $\boldsymbol{\Omega}$.

3.4.1.2 *Recursive Feature Elimination - Support Vector Machine (RFE-SVM)*

A different approach is used in the RFE-SVM algorithm described by Guyon et al. (2002). According to the authors "a good feature ranking criterion is not

necessarily a good feature subset ranking criterion". The authors proposed to test the weight vector $\mathbf{w}$ of the SVM solution to produce a feature selection method. The weights w_i (where w_i is *ith* position of the weight vector $\mathbf{w}$) of the decision function $D(\mathbf{x})$ (see equation (3.2)) are a function of the support vectors and can be used as feature ranking criterion.

RFE-SVM follows an iterative procedure based on the following three steps:

i) Train the classifier (optimization of the SVM parameters and compute the weight vector);

ii) Compute the ranking criterion based on the weight vector - $(w_i)^2$

iii) Remove the features with lowest ranking criterion.

In theory it is possible to use any kind of kernel of SVM in the RFE-SVM method. However, the computational effort involved in model optimization is higher in nonlinear kernels. Therefore we restrict the approach to the linear kernel.

Similarly to the mRMR approach, RFE-SVM can also be used to conclude about the most important feature-channel combinations.

3.4.2 Results

The dimensional reduction performed by the mRMR method exhibited a small improvement in the performance of the predictors. Considering the average *F-measure* of the best models (epoch-based metric), the mRMR subset with N_Ω =66 features achieves 0.14, while the complete feature set N_{feat} =132 presents 0.12. On the opposite, the subset obtained by the RFE-SVM method did not change the performance of the classifier (also registered 0.12) (Figure 3.20).

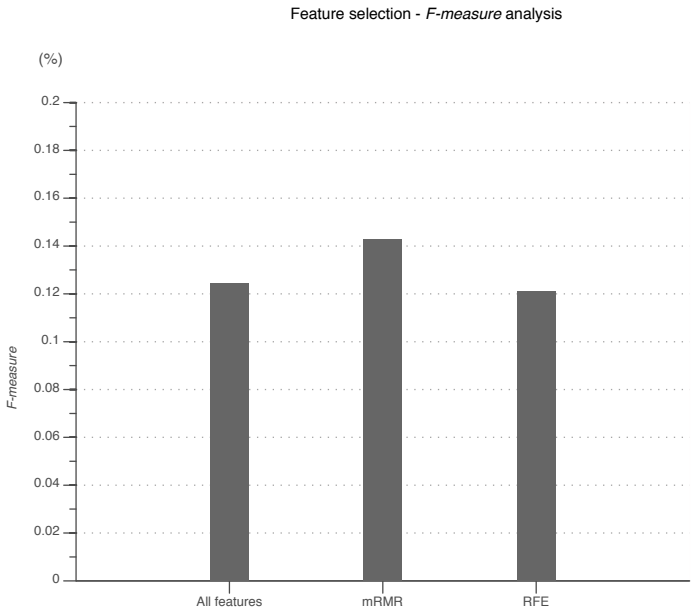


Figure 3.20 – Average *F-measure* presented by the different methods. The results obtained using mRMR present an higher *F-measure* compared to the results obtained using RFE.

In terms of event-based measures (alarm sensitivity and FPR.h^{-1}), the results show that the false positive rate decreases with the use of both feature selection methods, specially using the mRMR approach. However, the alarm sensitivity tends to decrease using feature selection methods (Figure 3.21).

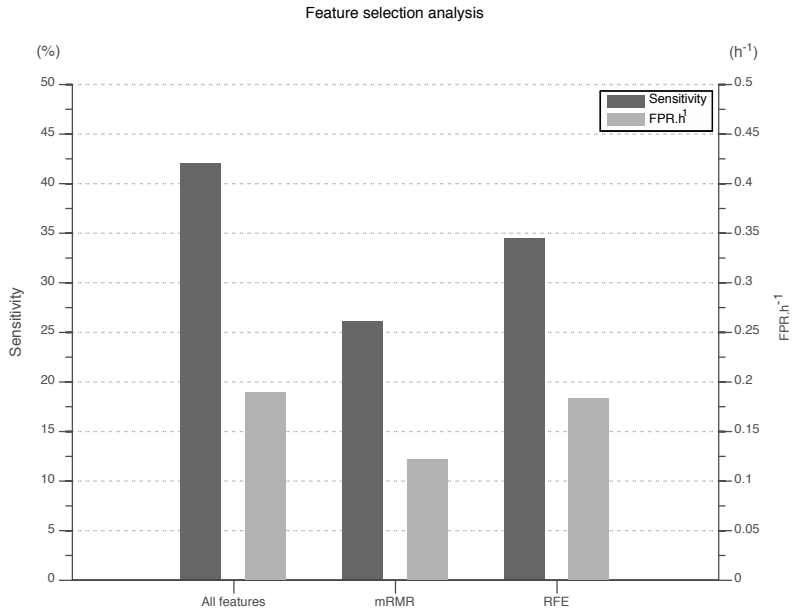


Figure 3.21 – Results obtained using the complete set of features, the mRMR subset and the RFE-SVM subset of features. The mRMR results present the lowest FPR.h^{-1}.

3.4.2.1 *Significance*

We have compared the performances obtained by the two different subsets with the upper critical sensitivity of a random predictor. Since all the optimization methods are applied in the training stage (using only the training set), the statistical analysis performed is similar to the one presented in Section 3.2.5, i.e. we considered the unspecific random predictor defined by equation (2.25) and 12 independent features (one for each dataset).

We found that five out of the 10 patients analyzed with mRMR improved their results, surpassing the threshold settled by the upper critical sensitivity of a random predictor. It is possible to conclude that the improvement is due to the decrease in the false positive rate. RFE-SVM also improved the results achieving statistical significance in two patients.

We performed a statistical analysis to study the influence of the feature selection methods. We evaluated the *F-measure*, alarm sensitivity and $FPR.h^{-1}$. Since the performance measures did not present a normal distribution (according to a Kolmogorov-Smirnov goodness of fit test), we performed the KW test (at the 5% significance level). The results lead us to conclude that no significant differences exist among the different strategies.

3.4.2.2 *Electrode-features combinations*

One of the objectives of this study is to find optimal *electrode-feature* combination using feature selection methods. The results for two patients are presented in the following.

Patient 3500

According to the epileptologist, patient 3500 presented right frontal epilepsy. The test set includes two seizures and the best model in the initial evaluation (based on the dataset formed by a *random* selection of electrodes and SOP of 40 minutes) achieved a alarm sensitivity of 100% (successfully predicted the two seizures) and $FPR.h^{-1}$ of 0.14.

The performance of the predictor improved above the threshold imposed by the upper critical sensitivity of a random predictor using the subset of features determined by mRMR. The alarm sensitivity decreased to 50% (missed one of the seizures) but no false positive was raised in the testing set. Figure 3.22 summarizes the subset of features selected according to the mRMR algorithm. The subset of features emphasizes the importance of the focal electrodes (electrodes F8 and FZ).

On the opposite, no particular pattern was found in the type of features selected (spectral features from different electrodes appear more often in the subset).

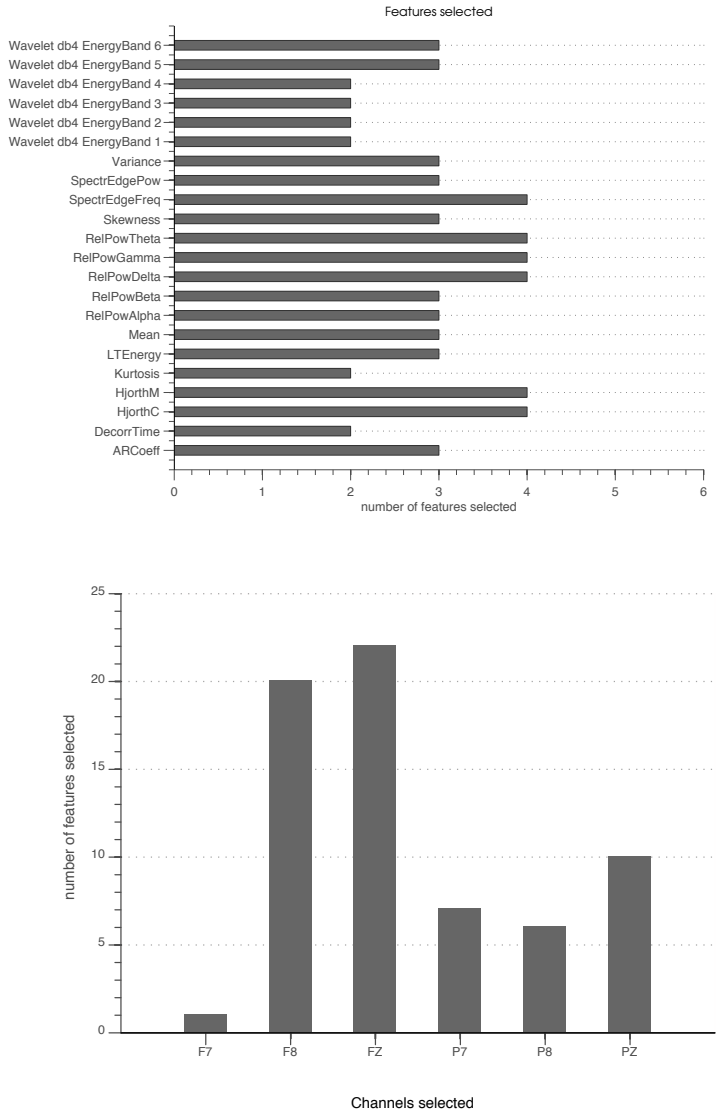


Figure 3.22 – Feature and channel selection performed using mRMR. On the top, the feature selection suggests that all the features play an important role in the predictor ability to classify patterns. On the bottom, the channel selection suggests that channels F8 and FZ are the most important in terms of information content.

Patient 1200

The patient, diagnosed with temporal right epilepsy, presented 20 seizures in the testing set. Both classification strategies (with and without feature selection) preferred the same electrode set and pre-ictal periods.

The model that presented the best results after the application of feature selection (RFE-SVM) achieved an *F-measure* of 0.18, an improvement when compared to the original set (0.11). Concerning event-based performance measures, the results present an increase in alarm sensitivity and FPR.h^{-1}. The alarm sensitivity, according to the statistical evaluation, is higher than the one presented by an unspecific random predictor.

Figure 3.23 summarizes the RFE-SVM-based subset of features. The subset highlights frontal lobe electrodes. Kurtosis and features characterizing lower frequencies from the different electrodes appear more often in the subset; the results seem to suggest their importance to increase the separability between classes.

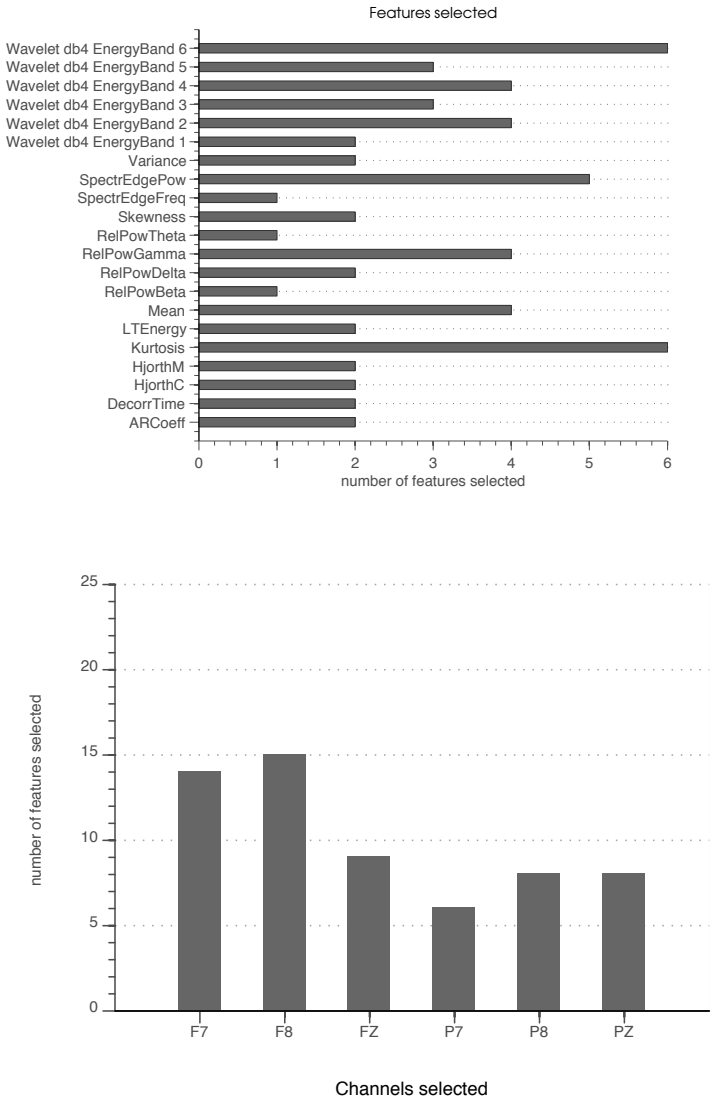


Figure 3.23 – Feature and channel selection performed using RFE-SVM. On the top, the feature selection suggests that kurtosis and wavelet coefficients (low frequency band) are important.

On the bottom, the channel selection suggests that frontal channels F7 and F8 are the most important in terms of class separability.

3.5 Feature Reduction

We applied classical MDS to a high dimensional feature set and the resulting reduced feature vectors were used as input to an SVM to determine the validity of the method.

3.5.1 Using classical MDS in seizure prediction studies

Our study followed the usual two-stage approach, training the SVM using the output of classical MDS on the training set, and testing of the model in the testing set.

Let us assume that the training set has N_P samples and the testing set has N_T samples. The most intuitive solution would be to embed all data, i.e. the complete training set and the known testing samples ($N_P + N_T$ examples), estimate a new transformation matrix and train a new SVM model. However this approach is unfeasible in real-time applications due to memory and computational requirements, i.e. every new example would require the computation of a new dissimilarity matrix and train, optimize and validate a new model, which, after a certain number of examples, would become prohibitive.

Hence, the interest lies in the performance of the classifier trained using only the original, smaller, training set. The general idea is to choose $\hat{N}_P$ samples ($\hat{N}_P < N_P$), also referred to as *landmarks* (Kruskal and Hart, 1966), and to apply MDS on this subset, mapping them into the subspace $\Re^{N_{mds}}$. These transformed samples are used as the training set of the classifiers. The remaining samples are then iteratively mapped to $\Re^{N_{mds}}$ using the transformation matrix estimated using the original landmarks and compose the testing set for the trained SVM (Trosset and Priebe, 2008).

The number of dimensions of the transformed hyperspace $\Re^{N_{mds}}$, input of the classifier, is determined according to the number of dimensions required to explain 80% of the variance of the original training set. The number of dimensions N_{mds} can be easily estimated adding the sorted eigenvalues of the transformation matrix.

3.5.2 Experiments in a dataset with 10 patients

The study was made on the same patient set considered in Section 3.4.1.

The dimensional reduction obtained by classical MDS presented an improvement in the average *F-measure*. The reduced set of features obtained using classical MDS achieved 0.15 compared to the 0.12 obtained with the complete set of features. However, this improvement was not noticeable considering event-based metrics (alarm sensitivity and FPR.h^{-1}).

The average sensitivity decreased from 41.99% (complete set of features) to 38.40% (features obtained using classical MDS), and the FPR.h^{-1} increased from 0.19 to 0.28 per hour (Figure 3.24).

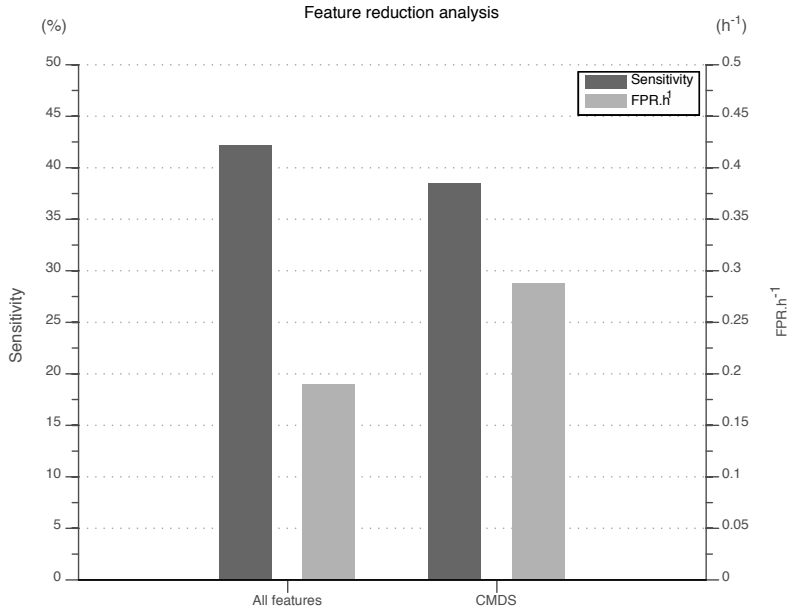


Figure 3.24 – Performance of classical MDS. The results tend to worsen compared to the complete set of features.

In terms of statistical significance, one of the 10 patients analyzed surpassed the upper critical sensitivity of a random predictor.

Figure 3.25 displays the reduced feature vectors (corresponding to the training set of patient 202, *focal* electrode array and two different SOP) embedded in a three-dimensional space. The plot shows the complexity and dispersion of the data in the embedded space.

It is especially interesting to notice that, in certain cases, different pre-ictal periods do not influence the dispersion of the data, i.e., the data agglomerations remain similar considering different pre-ictal periods.

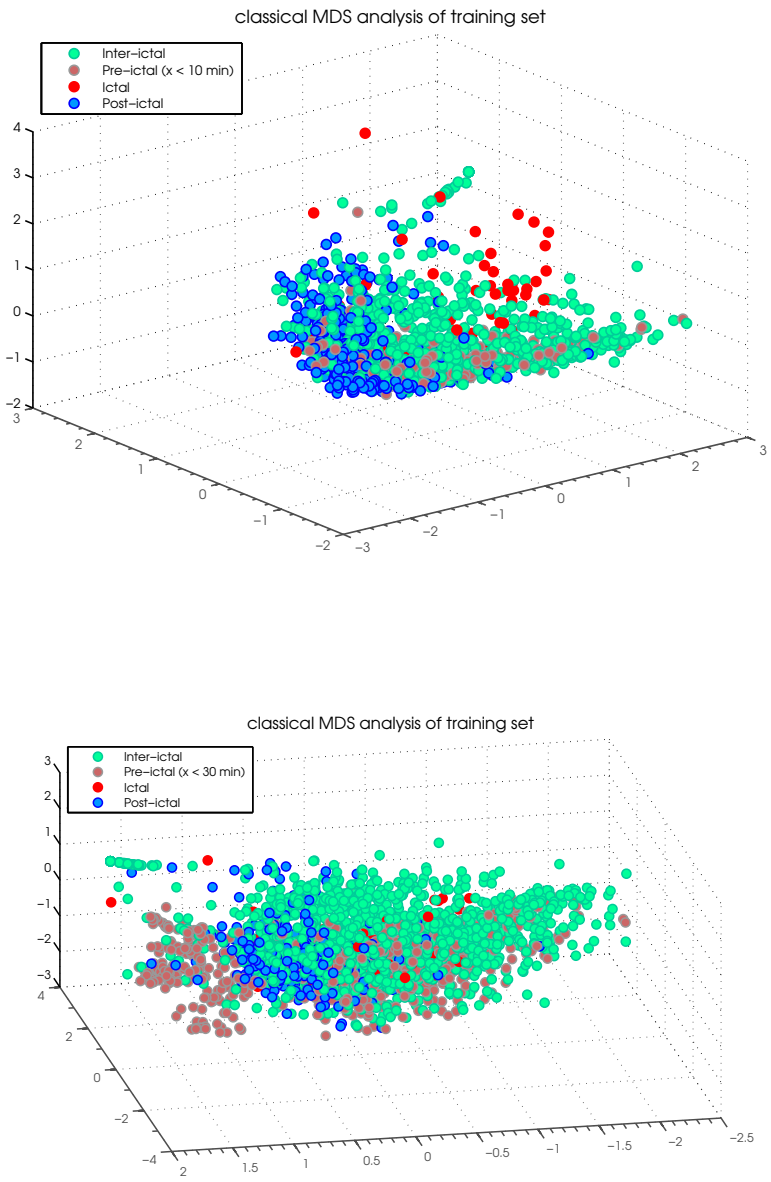


Figure 3.25 – Projections of three-dimensional data obtained by classical MDS. The plot on the top represents a 10 minutes SOP and the plot on the bottom a 30 minutes SOP. The figures present the complete training sets of patient 202, *focal* electrode array.

Another interesting property is related to the different combination of channels; in some cases different channels do not cause significant variations.

3.6 Regularization of the classification output

Epoch-based metrics such as *F-measure*, sensitivity or specificity may be insufficient (Temko et al., 2011) to measure the performance of a classifier. Since the aim is to predict seizures, the output of the classification scheme at each instant must be an alarm or its absence. So an event-based classification is needed. An additional step is necessary to transform a epoch-based metric (i.e., the output of the classifier every five seconds) into an event-based metric (series of alarms), so that prediction methods may be compared using frameworks such as the seizure prediction characteristic, proposed by Winterhalder et al. (2003).

The epoch-based approach (every sample is considered as a separate testing example) may be misleading. A binary problem has four possible outputs: true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN). The metrics usually used to characterize the output are the sensitivity $TP/(TP+FN)$, the specificity $TN/(TN+FP)$, and accuracy $(TN+TP)/(TN+TP+FN+FP)$. The unbalance between classes in seizure prediction, seriously compromises these metrics. For instance, a high accuracy is obtained if the classifier output is limited to 'non pre-ictal' labels. The *F-measure* tackles this problem in part. It considers both sensitivity and specificity, and is defined by equation (3.22).

In seizure prediction, the output of the classifiers can be divided into two states: pre-ictal and non pre-ictal. Each pre-ictal sample can be used to generate an alarm (e.g. Mirowski et al., 2009). However in order to reduce the number of false positives, a post processing step is usually necessary. Two approaches were suggested in the literature: the Kalman filter (Chisci et al., 2010; Park et al., 2011) and the firing power (Teixeira et al., 2011).

3.6.1.1 *Kalman filter*

The post-processing of the classification output by means of a Kalman filter (Kalman, 1960; Welch and Bishop, 2006) intends to decrease the number of false positives, smoothing the output of the two-class SVM classification. A second order Kalman filter has been considered, as was done by the reference Chisci et al. (2010).

Let us consider the smoothed SVM output moving in the state space with constant velocity. Since there is no acceleration, we can model its dynamics by a second order continuous state model where the states are the position and the

velocity. The autonomous continuous state equation and the position equation become (3.28), neglecting the random noise (supposed white noise).

$$\begin{bmatrix} g' \\ g'' \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} g \\ g' \end{bmatrix}$$
$$z = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} g \\ g' \end{bmatrix} \tag{3.28}$$

Now with a discretization interval of Δt, the discretized state equation becomes (3.29),

$$\begin{bmatrix} g_{t+1} \\ g'_{t+1} \end{bmatrix} = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 0 & \Delta t \\ 0 & 0 \end{bmatrix} \right\} \begin{bmatrix} g_t \\ g'_t \end{bmatrix} = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix} \begin{bmatrix} g_t \\ g'_t \end{bmatrix}$$
$$z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} g_t \\ g'_t \end{bmatrix} \tag{3.29}$$

resulting in the equation (12) of Chisci et al. (2010).

The aim is to compute the first component of g, the position. When the position reaches the SVM decision threshold, in an ascending state, an alarm is generated. However the state is not measurable, but can be estimated by a discrete Kalman filter, as in Chisci et al. (2010).

3.6.1.2 *Firing Power*

The firing power approach considers the binary output ('non pre-ictal','pre-ictal'), and generates alarms when the sum of the 'pre-ictal' elements rises above a threshold (Figure 3.10). Let us consider $D(x_t)$ the output of the SVM at instant t. If we consider the posterior probability of the classifier, the procedure is as follows:

i) Binarization of the SVM output (3.30);

$$output_t = \begin{cases} 1, \; if \; D(x_t) \geq 0 \\ 0, \; if \; D(x_t) < 0 \end{cases} \tag{3.30}$$

ii) Sum of the samples considered 'pre-ictal' in the past pre-ictal period (3.31),

$$F_{power_t} = \frac{\sum_{n=t-SOP}^{t} output_n}{SOP} \tag{3.31}$$

where F_{power_t} is the firing power at the instant t and SOP is the size of the window considered for correct predictions;

iii) Generate alarms through the use of a threshold. To determine the best threshold we used different values, {0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85}.

3.6.2 Comparison of both techniques with 10 patients

A comparative study between the Kalman filter and firing power technique was performed, considering exactly the same conditions.

The post-processing study was based on the long-term recordings of 10 patients with scalp EEG, selected accordingly to some quality conditions important to validate the approach.

If discontinuities exist in the datasets, a "re-set" of the post processing strategies would be necessary and the comparison between them would be biased. Therefore, the patients' dataset contain at least five days of data and more than 99.5% of effective recording time (i.e., minimum intervals or discontinuities in the dataset). Fifteen patients fulfill these conditions, and 10 were randomly selected. All the selected patients suffer from temporal lobe epilepsy.

The dataset included 58 seizures over a total of 751 hours of testing data. We used 22 univariate features and two different arrays of electrodes (*focal* and *10-20*).

Considering the different parameters of the framework (different pre-ictal ranges, electrodes arrays, and post-processing parameters), we analyzed 6400 outputs for each patient. The models were evaluated according to alarm sensitivity and FPR.h^{-1}. The best predictor was selected based on the closest distance to the optimal point, i.e., alarm sensitivity of 100% and FPR.h^{-1} of 0.0.

In general, the results reflect a trade-off between sensitivity and specificity (or FPR.h^{-1}). In other words, an increase in sensitivity is usually associated to an increase of FPR.h^{-1}.

The results using the Kalman filter lead to higher sensitivities. In average the sensitivity was approximately 84%. However, this high sensitivity is accompanied

by a high $FPR.h^{-1}$ (1.51 per hour). The usefulness of such results is negligible (Schulze-Bonhage et al., 2010). The firing power approach yields a significant decrease in the $FPR.h^{-1}$ (0.20 per hour). The sensitivity also decreases (slightly) to an average of 77%.

The best model favored longer pre-ictal periods, while no particular preference was found in electrode selection method.

3.7 Circadian Rhythmicity

Several studies reported evidences of interactions between epilepsy and circadian rhythms, i.e. any physiological activity that regularly oscillates with a period of 24 h (Durazzo and Zaveri, 2011). The occurrence of seizures is not entirely random in terms of their distribution over the day. The reason for these specific patterns is not clear but it is believed that the circadian clock plays an important role. Moreover, this modulation may be influenced by the location of the epileptic foci (Quigg, 2000; Durazzo and Zaveri, 2011).

Chronic disorders of the central nervous system have shown to alter the circadian rhythmicity. Some epileptic patients were reported to present disorders in the neuronal excitability associated with the sleep-wake cycle and daily rhythms of hormone release and core body temperature (Bazil et al., 2000). Moreover, sleep deprivation is known to enhance seizures (Malow, 2004).

Different studies analyzed the temporal evolution of features and the different timescales involved while studying the pre-ictal patterns of the features (Schelter et al., 2006b; Kuhnert et al., 2010). According to the study by Schelter et al. (2006b) changes in the EEG dynamics are related to the sleep-wake cycle and interfere with the seizure-prediction algorithms.

These evidences lead us to analyze the output of the classifiers, and look for possible circadian dependencies in classification.

3.7.1 Analysis with 10 patients

We analyzed the classifiers' output of 10 patients suffering from temporal lobe epilepsy available at the European Database on Epilepsy (see Section 3.6.2). An important requirement for selection of patients was the absence of significant gaps without data in the continuous EEG recordings.

The best model was determined based on the *F-measure*. The idea was to find the best model in terms of the description of the dataset and not in terms of the alarm generation series.

The temporal evolution of the classifier output of some patients seems to exhibit large fluctuations over time. To investigate possible circadian influence we divided the data from the period between 9 p.m. and 9 a.m. (considered night time) and the rest of the data (day time) (Figure 3.26).

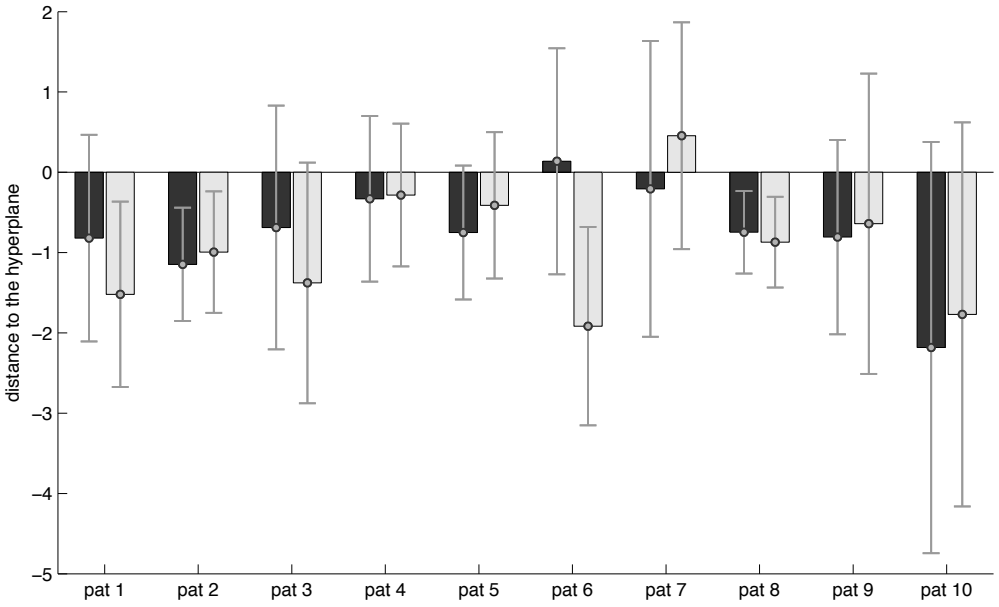


Figure 3.26 – Differences of the classification output (distance of each sample to the hyperplane separating) during the day and the night. The dark gray represents the day data and the light gray represents the night data.

To confirm differences between the two periods of the day, we performed a non-parametric statistical test. The Wilcoxon rank-sum test (Wilcoxon, 1945) is used to test the hypothesis of equality between the mean of two independent variables (Montgomery and Runger, 2003). We considered the data from the night time and the data from the day time of the 10 patients analyzed.

We observed statistically significant differences between the classification output of the two periods of the day for all the 10 patients ($p<0.05$).

We also estimated the power spectral density of the unsmoothed output of the best classifier for each patient. For most of the patients we found a strong component at about 24 hours (and sub-harmonics at about 12 and eight hours).

In Figure 3.27 we present the power spectral density of the classification output of patient 6. The trace presented shows strong contributions form the eight hours and 24 hours timescales.

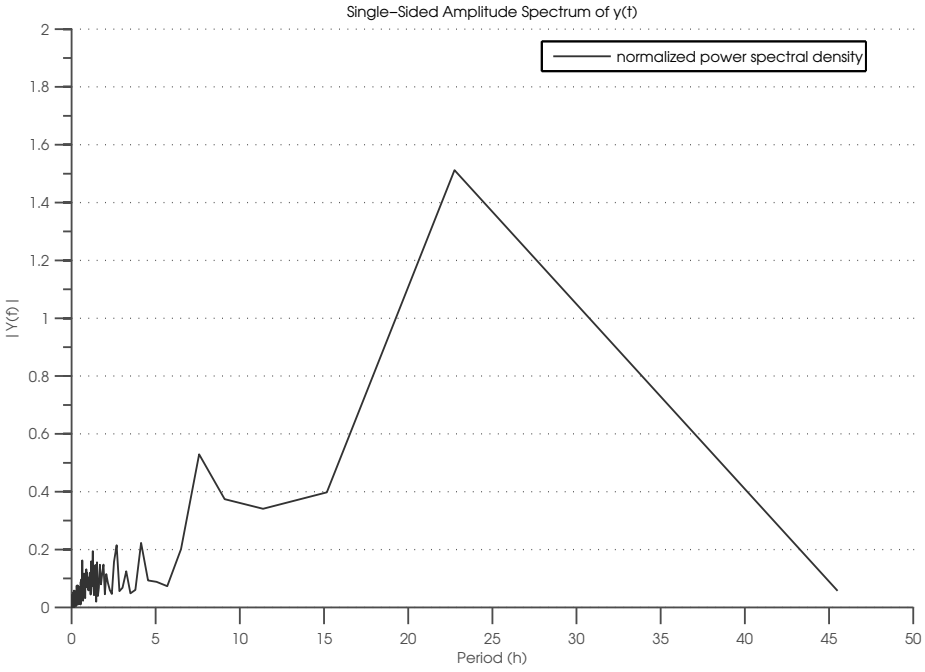


Figure 3.27 – Power spectral density of the temporal evolution of the classifier output. The results show a peak around 24 h.

These results suggest that the pre-ictal patterns are related to processes acting on timescales of days (daily rhythms). The training patterns of the class 'pre-ictal' (class '1') seem to adapt the model to the sleep-wake cycle, suggesting the idea of a 'pro-seizure state' when the patient is in a particular period of the day.

3.8 Summary and Discussion

Although there is growing evidence that a pre-ictal stage exists, a rigorous study, based on a sound statistical framework is still missing. The comprehensive study presented in this chapter aimed to investigate the feasibility of epileptic seizure prediction, researching predictors with a performance above chance level (meaning by this that the results obtained are better than the upper critical sensitivity of an analytical random predictor defined by (2.25)), considering the

recommendations suggested by several authors (Mormann et al., 2007; Schelter et al., 2006a; Winterhalder et al., 2003).

Considering the set of 216 patients used in this chapter (185 monitored with scalp electrodes and 31 monitored with intracranial electrodes), we determined that 24 patients present sensitivities higher than the ones presented by the upper critical sensitivity of a random predictor (at the significance level of 5%). Moreover, the binomial test applied at the group level allows us to conclude that the results are statistically significant.

Considering intracranial and scalp recordings individually, we noticed that the statistical significance at the group level is also achieved demonstrating the usefulness of the predictor proposed in both scenarios. These results are rather surprising due to the different nature of scalp and intracranial recordings (Nunez and Srinivasan, 2006). On the one hand, the features considered in this study may be insufficient to extract (possible additional) information contained in iEEG, specially considering the local networks involved in seizure generation. On the other hand, information about the global behaviors of the brain (also available through scalp EEG) may contain the 'predictive information'.

The comparison of these results with other recent machine learning studies is to a large extent impossible (Mirowski et al., 2009; Netoff et al., 2009; Chisci et al., 2010; Park et al., 2011). These studies present promising results using customized algorithms based on machine learning algorithms and multiple features (for example, some authors report sensitivities as high as 97.5% and FPR.h^{-1} as low as 0.27). However the lack of an appropriate statistical framework to analyze the results do not allow concluding about usefulness of such studies. Moreover, these results were based on short segments of iEEG data with discontinuities and limited number of patients.

Different variables were studied in the performance measures considered. The data seem to suggest that temporal epilepsies present the best sensitivities and central epilepsies the lower FPR.h^{-1}. However, no statistical significant changes were found based on the localization of the epileptic focus. Based on the KW test performed, lateralization (right, left or both) induces significant differences in the *F-measure*. According to our results, longer pre-ictal periods seem to be more appropriated for seizure prediction studies (average of 27 minutes). A statistically significant difference was determined considering the *F-measure* among the SOP periods. These results lead us to conclude that a 10 minutes SOP provides

insufficient information to the model to generalize in the testing set. Considering the three different approaches that were used to select electrode subsets, and despite the higher sensitivity presented by the *random* selection, the variations were not significant.

The fact that the results obtained for some patients surpassed the upper critical sensitivity presented by a random predictor is not enough to reach clinical relevance. Further research is needed and we presented in this chapter several studies in this direction, for a smaller set of patients, just to measure the improvements that these techniques could eventually permit.

In this sense, we investigated the ability of pre-processing methods (feature selection and reduction methods) to improve the results. These computationally expensive approaches were tested in a subset of 10 patients and intended to test the influence of these strategies in the overall performance of the algorithm. The results proved to marginally increase (not statistically significant differences) the performance of the prediction algorithm. In particular, the mRMR method presented the best results. The use of these strategies also allowed us to analyze feature subsets (selected according to different parameters), and conclude that there are specific patterns for each patient, confirming the need for individually tailored algorithms, and the selection of an appropriate combination between electrodes and features.

Analyzing the long-term variability of the classifiers' output, we identified circadian dependencies, which is in close agreement with previously published studies (Schelter et al., 2006b; Kuhnert et al., 2010). The results support the hypothesis that pre-ictal changes reflect different vigilance states.

Nevertheless, we consider that prediction may be further enhanced. Good performance of the predictors is only achieved for a small subset of patients.

Different approaches may be considered towards the improvement of the seizure prediction algorithm such as: including other univariate or multivariate features (adding information about the local and global behavior of the brain), considering adaptive classification methods such as incremental and decremental SVMs (Cauwenberghs and Poggio, 2001) to tackle circadian rhythms involved in seizure generation, etc.

In the following chapters we will present research made with this aim.

4. Identification of the epileptic brain states using EEG spectral analysis and topographic mapping

In the previous chapter we presented the main results of a seizure prediction algorithm based on several machine learning techniques applied to a large database of 216 epileptic patients. The quantitative evaluation of the algorithm presented sustainable results for a small group of patients. The data-driven approach used creates a black-box model in which very little may be inferred about the model constructed by the classification algorithm.

Despite the encouraging results and acknowledgment that predictive information was found for a group of patient, the performance of the algorithm is still largely insufficient for clinical usability. Different approaches must be searched for the improvement of the results.

In this chapter we hypothesize that important pre-ictal information may be contained in the spatio-temporal structure of the EEG time series. We shift from a data-driven approach (in which very little assumptions are made about the data) to a more qualitative approach, in which we pretend to enhance our comprehension about the data structure. The main idea is to unveil the structure of the multichannel EEG recordings through innovative approaches, highlighting regions and features that present specific patterns in the pre-ictal period. This may lead to the identification of the different states of the epileptic brain. The method is based on the topographic mapping of the time varying relative power of delta, theta, alpha, beta and gamma frequency bands, estimated from EEG recordings; it characterizes spatio-temporal patterns and models the data using the electrode configuration and points of interest estimated from the data. The method is tested in 10 patients of the EPILEPISAE database. We believe that this patient subset (containing a total of 497.3 hours and 30 seizures) is appropriate to demonstrate the feasibility and usefulness of the method.

The interpretation of the results focuses on the ability of the method to reproduce and characterize the pre-ictal state. This work has been previously published in J.1 and P.1.

4.1 Background and Related Work

Topographic mapping represents a complete picture of the area of the brain covered by electrodes and is one of the main tools to analyze the spatial content of the EEG data. The topographic EEG map is a neuroimaging technique for representing spatial content contained in the EEG using colored contours of the brain activity expressing the amplitude of a signal derived from the EEG (either the original EEG signal or some characterizing measure).

The first comprehensive studies using topographic mapping date back to 1980's (Sandini et al., 1983). The approach is generally used to determine the presence of tumors and focal diseases of the brain (for example, malformation, stroke or epileptic focal regions) assisting the clinical diagnosis by complementing the temporal dimension of electrophysiological changes with a spatial dimension (Ten Caat et al., 2008; Alba et al., 2010).

The study of the topographic information of the EEG is especially important in epilepsy. Some authors have hypothesized that epileptic seizures may be related to synchronization patterns originating from critical groups of neurons that may spread over other regions of the brain (Mammone et al., 2010). Other authors consider the possibility that the transition between inter-ictal, pre-ictal and ictal epileptiform events is a consequence of dynamical changes in the spatio-temporal patterns across different brain structures (Hegde et al., 2007; Mammone et al., 2010).

However, to understand the pre-ictal dynamics (and model the behavior of the epileptic brain), long-term continuous analysis might be necessary from either the spatial or temporal point of views (Mammone et al., 2006; Schulze-Bonhage et al., 2011). Studies proposing quantitative methods to model long-term spatio-temporal dynamics are still relatively rare (Mirowski et al., 2009; Kuhnert et al., 2010). In particular, the analysis of the long-term continuous space-time-frequency hyperspace is to a large extent unexplored.

The method proposed aims to model the long-term spatio-temporal activity of the epileptic brain. Four different states are considered: pre-ictal, ictal, post-ictal and inter-ictal states.

The first step is the computation of the relative power in different EEG frequency bands (δ_{rp}, ϑ_{rp}, α_{rp}, β_{rp} and γ_{rp}). According to Mormann et al. (2005) there are statistical significant differences between the power in the pre-ictal and the inter-ictal period in different frequency bands. The authors mentioned, in particular, a decrease in the delta band and an increase in the remaining frequency bands as seizures approach. We hypothesize that the changes in the frequency content over time may be associated to signature variations in the spatial domain and that the behavior of particular regions varies before the epileptic seizures.

The second step is the computation of topographic maps of each sub-band to evaluate spatial asymmetries and abnormalities in the spectral content of the EEG.

The next step consists of the segmentation of each topographic map based on the similarities of the values of the relative power in each point of the map where an electrode is positioned; the values of the relative power in the points other than the electrode location are computed by interpolation. This segmentation finds the regions with similar relative power. The centers of the regions with the maximum relative power and minimum relative power are our points-of-interests (POI). These POI move over time, i.e. they are defined in time-space-frequency.

Another important variable is the duration of the pre-ictal period, the SOP. Mormann et al. (2007) suggested that electrophysiological changes develop minutes to hours before the actual seizure onset. We determined the pre-ictal period for each patient using statistical test of hypothesis (the Wilcoxon rank-sum test), maximizing the differences between pre-ictal and non pre-ictal samples. We also used the same method to determine the best frequency band. Finally, based on the pre-ictal period and frequency band selected, the locations of the POI over time are used to perform two tasks:

i) to generate a two-dimensional histogram for each of the four brain states;

ii) to train a Hidden Markov Model (HMM) of the different epileptic brain states and the transitions among them.

4.2 Dataset description

Ten patients from the EPILEPSIAE database (see Section 3.2.1) were selected: patients B, F, G, H, and I suffering from frontal lobe epilepsy, patients A, C, D, E, and J from temporal lobe epilepsy (Table 4.1).

Table 4.1 – Sampling frequency, number of seizures analyzed, focal region and total recording time for each patient. "Localization" – "T" = Temporal region, "F" = Frontal region.

Patient	**Sampling frequency (Hz)**	**Number of seizures**	**Localization**	**Duration (h)**
A	512	2	T	41.1
B	512	3	F	43.1
C	1024	2	T	23.9
D	1024	2	T	20.1
E	256	3	T	60.2
F	256	3	F	68.8
G	400	6	F	63.3
H	400	3	F	111.8
I	400	3	F	26.1
J	256	3	T	38.9

The patients, from different epilepsy centers (Ihle et al., 2012; Klatt et al., 2012), were monitored through multi-channel scalp EEG, placed according to the 10-20 system. The patients were randomly selected among the patients of the database fulfilling the condition of minimal discontinuities in the EEG recordings to ensure the quality of the study. In this sense, for each patient, continuous segments with at least two seizures were selected for analysis.

A total of 30 seizures over 497.3 hours of multichannel EEG data were analyzed.

4.3 Methods

The block diagram presented in Figure 4.1 summarizes the main stages of the method: preprocessing, feature extraction, and reconstruction of a sequence of

states using POI location over time. Each step is discussed in the following subsections.

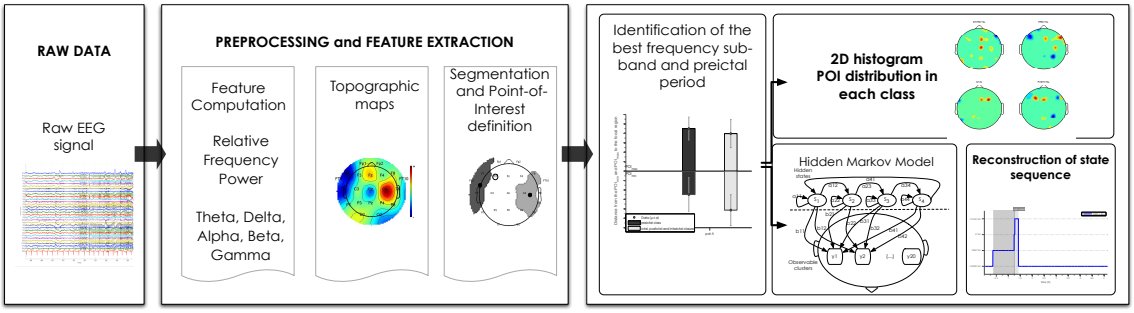


Figure 4.1 –The algorithm consists in: (i) preprocessing, (ii) feature extraction, that is, iterative computation of the maximum and minimum POI, and (iii) feature analysis. The analysis of the feature set results in two different output: (a) two-dimensional histograms describing the areas most relevant to each class (inter-ictal, pre-ictal, ictal and post-ictal), and (b) training of the HMM and reconstruction of the state sequence using the Viterbi algorithm.

4.3.1 Relative spectral power and topographic mapping

In order to estimate the spectral features from the EEG data we used a five seconds non-overlapping sliding window approach as in the previous chapter. For the purpose of seizure prediction, we considered that a five seconds interval is suitable to represent the variations of the EEG raw data. Moreover, it represents a good compromise between long windows (losing temporal resolution and the stationarity assumption) and short windows (losing frequency resolution) (Blanco et al., 1995).

The data preprocessing was based on a 50 Hz notch filter to remove possible artifacts related to power line noise (see Section 3.2.2 for a complete description of the notch filter).

In each window we compute the relative power in the different sub-bands: δ_{rp} (0.5 - 4 Hz), ϑ_{rp} (4 - 8 Hz), α_{rp} (8 - 15 Hz), β_{rp} (15 - 30 Hz) and γ_{rp} (30 Hz to *Nyquist frequency*), as described in Section 3.2.2.5.

An example of the temporal pattern of the relative power in delta sub-band for electrode C3 on an interval of two hours is presented in Figure 4.2a.

To obtain the topographic maps, the relative power is computed for all the electrodes and all the sub-bands.

Let us consider for example the frequency sub-band beta, all the electrodes and a certain instant. We compute the relative power according to the frequency interval of the beta sub-band β_{rp} (15 – 30 Hz) in each electrode. A color is associated to each value; red corresponds to the highest and dark blue to the lowest. Then by cubic spline interpolation for the points between the electrodes, a colored map (as depicted in Figure 4.2b) is obtained, corresponding to the topographic map for the considered instant. Contour lines are traced passing by equal-color points (Delorme and Makeig, 2004). As time elapses, the map changes dynamically, capturing the variations in the brain activity.

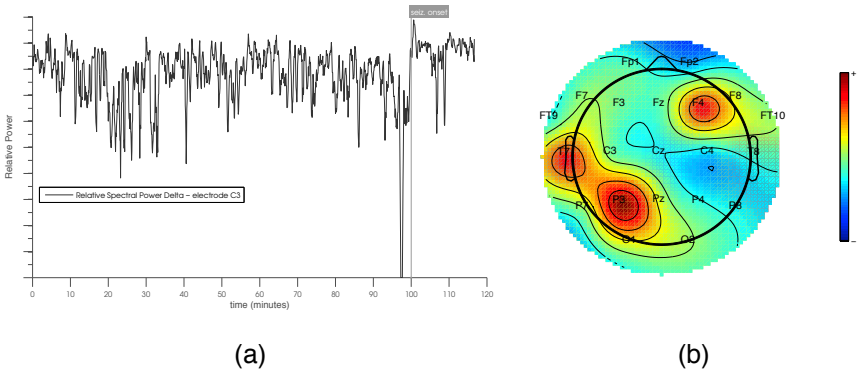


Figure 4.2 – (a) Relative Spectral Power (delta sub-band and electrode C3) in a two hours period of inter-ictal data; (b) The topographic representation of the first sample of the feature Relative Spectral Power (beta sub-band considering all the electrodes).

4.3.2 Segmentation and point of interest

Image segmentation is the process of identifying distinct regions in an image, such that each region is homogeneous with respect to some property (Roerdink and Meijster, 2001). We aim to segment the regions with lowest and highest relative power over a temporal sequence of topographic maps; this sequence defines a trajectory of these regions, which are used to characterize the pre-ictal period.

Image segmentation is performed, based on the graph theory, by the normalized cuts algorithm (Shi and Malik, 2000; Cour et al., 2005).

4.3.2.1 *Normalized cuts*

The normalized cuts algorithm treats a pixel in the topographic map as a node of a graph, and considers the segmentation as a graph-partitioning problem.

A graph $G=(V,E)$, where V represents the vertices (or nodes) and E represents the edges, can be partitioned into two disjoint subsets, A and B. The degree of similarity between the subsets can be computed as (4.1):

$$cut(A,B)=\sum_{u\in A,v\in B} w(u,v) \tag{4.1}$$

where $w(u,v)$ is the similarity between nodes u in A and nodes v in B. The similarity between nodes is defined according to the intensity (amplitude of the variable), which is here interpreted as a distance (Cour et al., 2005). Therefore, finding the optimal bipartitioning corresponds to the minimization of the *cut* function (4.1).

Finding the minimum *cut* is a well-studied problem, for which efficient algorithms have already been proposed. However, most of the methods favor the creation of small sets of isolated nodes in the graph. To lift these drawbacks, Shi and Malik (2000) proposed a normalized version of the algorithm, *Ncut*, where the similarity between two subsets is normalized by the total similarity of each subset (4.2):

$$Ncut(A,B)=\frac{cut(A,B)}{assoc(A,V)}+\frac{cut(A,B)}{assoc(B,V)} \tag{4.2}$$

where $assoc(A,V)=\sum_{u\in A,t\in V} w(u,t)$ is the total weight connection from the nodes of A to all nodes in the graph; similarly for the subset B, $assoc(B,V)=\sum_{v\in B,t\in V} w(v,t)$. By minimizing (4.2) we obtain an optimal solution for the graph partitioning.

Figure 4.3 illustrates the result of the segmentation algorithm applied to the topographic map in Figure 4.2b. The dots represent the centers of each segment and are identified as point of interest *max* (segment with the maximum value) and point of interest *min* (segment with the minimum value of the topographic map). Mormann et al. (2005) reported that the different frequency sub-bands present opposite behavior before seizures (some present an increase in the pre-ictal period while others show a decrease). Therefore both points of interest *max* and *min* may be important to characterize EEG data. If the relative power of a particular sub-band increases in the pre-ictal period it may be interesting to follow the point of

interest related to the *max*. On the opposite, if the feature decreases the point of interest related to the *min* is possibly more meaningful.

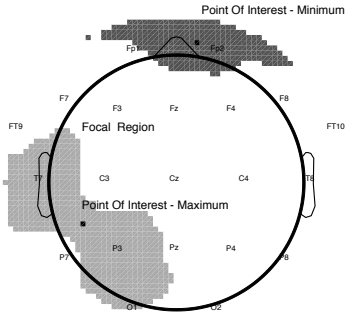


Figure 4.3 – Segmentation of the topographic map presented in Figure 4.2. The light gray and dark gray regions contain the maximum and the minimum values of the relative spectral power. The black dots represent the centers, i.e., points of interest (POI), of each region.

4.3.3 Selection of the frequency sub-band and optimization of the pre-ictal period using statistical analysis

An important part of this study is to determine which of the frequency bands (and corresponding features) better characterizes the pre-ictal period, considering the changes in the time and space domains. Previous studies (Mormann et al., 2005), despite mentioning these changes do not refer which combination *feature–pattern* better explains the transition between states. To determine the best feature we performed a statistical evaluation to analyze the difference between the means of pre-ictal class and the combination of the rest of the classes.

Let us consider the focal electrodes identified by the epileptologists and their placement in the topographic mapping as spatial references. The distance between the references and each POI (sequence of *max* and *min*) over time is computed, and two time series of distances are obtained. Each time series is divided into two variables, pre-ictal and non pre-ictal. To estimate the optimal pre-ictal period we considered periods ranging from 2 minutes to 60 minutes (with 2 minutes interval) before the seizure onset.

Firstly, using the Jarque-Bera test, we tested if the time series are non-Gaussian. This test measures the skewness and kurtosis of a time series matching with the skewness and kurtosis of a Gaussian distribution (Jarque and Bera, 1987). Since the time series did not exhibited a normal distribution, a non-parametric statistical hypothesis test, the Wilcoxon rank-sum test (Wilcoxon, 1945) was used

to test the null hypothesis, i.e., if the means of the distributions of both time series were equal.

The test gives a p-value that can be used as the criteria for rejection of the null hypothesis, or to establish a confidence interval. In our case, we interpret the p-values as a measure of separability between classes, in other words, the smallest the p-value the more confident we are that the statistical distributions are different.

The results obtained from Patient A using the POI related to the minima are illustrated, as example, in Figure 4.4. The p-values are very low for pre-ictal periods longer than six minutes, while the minimum is obtained for 52 minutes. These values demonstrate that the distribution during the pre-ictal period is different from the distributions of the other periods and they suggest that the most confidence in this statement is obtained for 52 minutes.

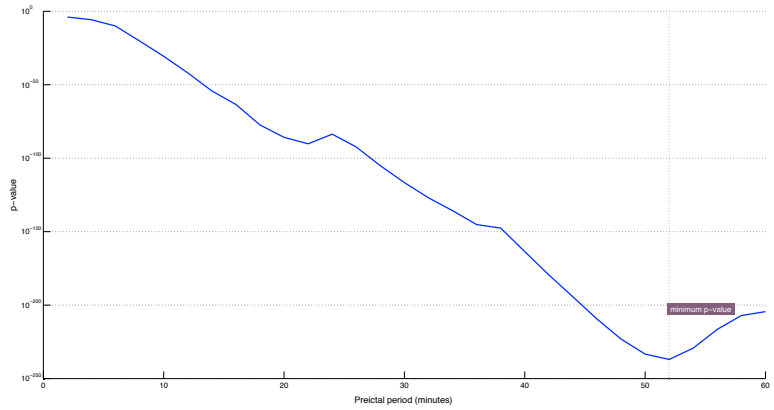


Figure 4.4 – Results of the Wilcoxon rank-sum test applied to the non pre-ictal and pre-ictal data for minimum-based POI, patient A, sub-band beta. The minimum of the p-value function determines the 'optimal' separation between sets, that is, the best pre-ictal period.

4.3.4 Two dimensional histograms of POI

The POI are defined in a grid with 67 rows and 67 columns that discretizes the topographic map. To better understand the spatial distribution of each class, two-dimensional representations (one for each class) of the feature selected were designed. Each position available corresponds to a bin of a two-dimensional histogram. An example is presented in Figure 4.5.

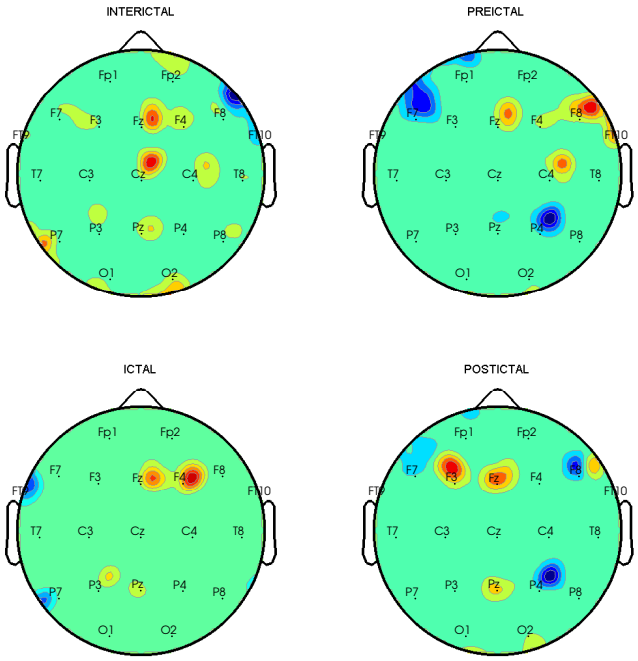


Figure 4.5 – After identifying the optimal pre-ictal period, the data is grouped into four different classes. Each two-dimensional histogram represents the spatial distribution of the minima and maxima POI over the scalp for each class. The red regions correspond to high concentration of maxima and blue areas correspond to high concentration of minima.

Patient A, depicted in the figure, has the seizure origin in the left frontotemporal area. One of the seizures appeared bilaterally as AF7 and AF8 electrodes were both identified in the seizure origin. The figure shows the results of the beta frequency range with a pre-ictal period of 52 minutes. The inter-ictal segments show the frontal region as the most noticeable. The presence of maxima and minima is especially dense in the frontotemporal areas (center and right regions). The maxima tend to appear around FZ (also around CZ) and the minima are very common around F8. The pre-ictal segments present a different pattern, the minima tend to appear in the left frontal areas (also in the right parietal region) while the maxima are common around FT8 (right frontotemporal regions) but also appear around the FZ region. The ictal segment present very well defined areas: the minima appear in the left frontal (region related to the seizure origin) and the maxima usually appear around F4 and FZ. The post-ictal state presents a different configuration, maxima in centrotemporal areas and the minima around the P6 region.

4.3.5 Hidden Markov Model

A Hidden Markov Model (HMM) is an interesting method for the estimation and analysis of state transitions especially in multi-stage processes (Chiu et al., 2011). The HMM is especially well suited for our time varying feature (POI positions), which is governed by different underlying dynamical states (Wong et al., 2007).

HMM model the sequence of observations - the POI positions - based on the four underlying 'hidden' states, pre-ictal, ictal, post-ictal and inter-ictal, having probabilistic inter-state transitions (Durbin et al., 1998).

The POI are defined in a grid with 67 rows and 67 columns. Considering all possible positions would increase the computational requirements and could result in spatial overfitting of the data. As a compromise between complexity (high number of clusters) and ability of the method to describe the data (*F-measure* of the HMM output), we grouped the POI into 20 clusters.

Figure 4.6 shows the averaged performance among the patients according to the number of clusters considered and presents a significant improvement in performance from 4 to 20 clusters. Increasing the complexity beyond 20 clusters does not present substantial variations.

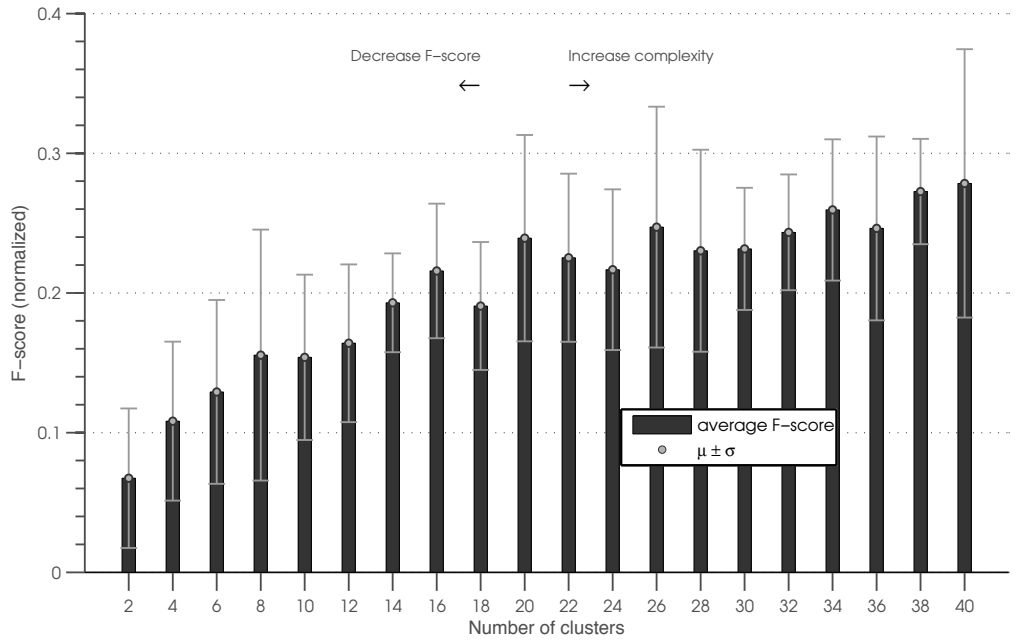


Figure 4.6 – Averaged *F-measure* obtained among the patients considered according to the number of clusters selected. The number of centers used to cluster the POI over the discretized map represents a compromise between the complexity of the model (if we increase the number of clusters) and the ability of the method to describe the data (*F-measure* of the HMM output). Moreover, the increase of the performance of the method tends to stabilize around the 20 clusters.

The use of these custom reference points (clusters based on POI) instead of the location of the electrodes presents some advantages. The sequence of clusters cl_{i_t} was used as the observable variable of the HMM. The k-means algorithm was used to determine the position of the clusters. Then, as illustrated in Figure 4.7, e_{ij} represents the probability (also known as the emission probability) of a certain brain state bs_i be connected to a cluster cl_j. q_{ij} is the probability of transition between state bs_i and state bs_j (Jeung et al., 2007).

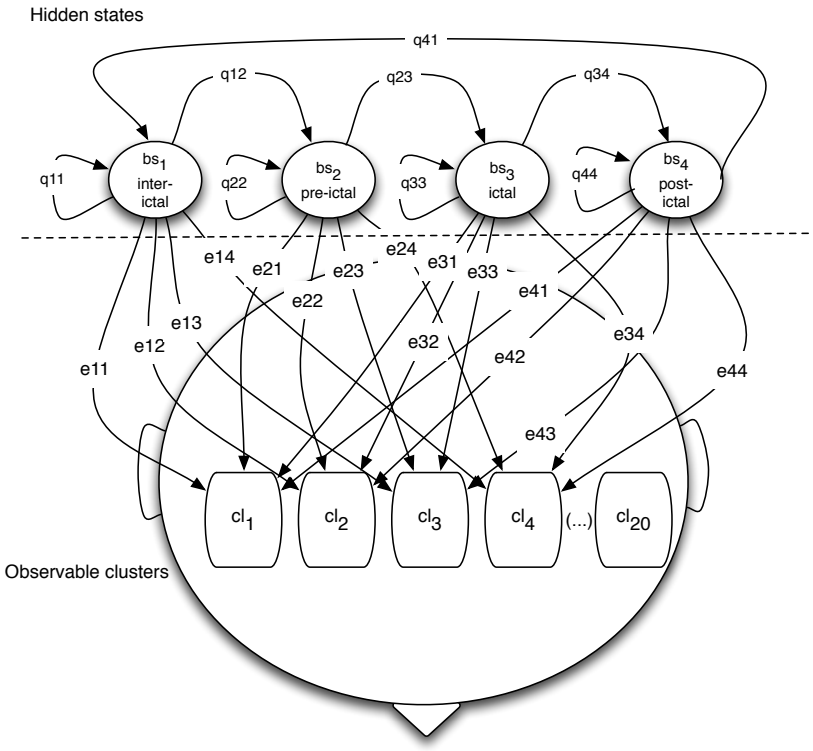


Figure 4.7 – HMM describing state transitions based on a sequence of observations extracted from a patient data. The model considers four states: i. inter-ictal, ii. pre-ictal, iii. ictal, and iv. post-ictal. q_{ij} and e_{ij} represent the state transition probabilities and the observation probability respectively. cl_i represent the set of clusters obtained using k-means to define a cluster-coded trajectory. We used the Baum-Welch algorithm to train the model and the Viterbi algorithm to reconstruct the state-sequence.

In order to construct a trajectory-based HMM, we must define (i) the number of states N_{BS}, (ii) the number of observation clusters N_{CL}, (iii) a set of state transition probabilities $\mathbf{Q}$ and (iv) a set of observation probabilities $\mathbf{E}$.

The set BS of possible hidden states is expressed as (4.3),

$$BS = BS\left\{bs_1(inter\text{-}ictal), bs_2(pre\text{-}ictal), bs_3(ictal), bs_4(post\text{-}ictal)\right\} \tag{4.3}$$

and additionally one can define a possible observation set as (4.4).

$$CL = \left\{cl_1, cl_2, ..., cl_{N_{CL}}\right\} \tag{4.4}$$

The state transition probabilities are defined as a four by four matrix $\mathbf{Q}$ where q_{ij} is defined as the probability of state bs_j in the next instant knowing that the state in the present instant is bs_i, or formally as in (4.5).

$$q_{ij} = Prob\left[bs_{j_t} \middle| bs_{i_{(t-1)}} \right]; 1 \le i, j \le N_{BS} \tag{4.5}$$

The 4 by 20 matrix $\mathbf{E}$ represents the probabilities among observations and is defined by (4.6).

$$e_{ij} = Prob\left[cl_{j_t} \middle| bs_{i_t} \right]; 1 \le j \le N_{CL}; 1 \le i \le N_{BS} \tag{4.6}$$

where e_{ij} denotes the probability of observing the cluster $cl_j \in [1, N_{CL}]$, while the system is in the state $bs_i \in [1, N_{BS}]$.

Due to the nature of our problem the following constraints were imposed during the training (4.7), (4.8), and (4.9)

$$q_{13} = q_{14} = q_{21} = q_{24} = q_{31} = q_{32} = 0 \tag{4.7}$$

$$q_{11} = 1 - const \tag{4.8}$$

$$q_{12} = const \tag{4.9}$$

The value *const* represents the probability of transition between inter-ictal and pre-ictal states and is estimated in order to optimize the sensitivity (and specificity) of the observed output.

The probabilities q_{ij} and e_{ij} are computed using the Baum-Welch algorithm, which is based on the expectation maximization (Durbin et al., 1998). We used the Viterbi algorithm (Durbin et al., 1998) to reconstruct the most likely state sequence using the trained HMM and the sequence of real observations. To assess the quality of the model we compare the reconstructed sequence of states with the target proposed.

To evaluate the output of reconstructed HMM, the sensitivity, specificity, accuracy and *F-measure* (see Section 3.2.4.1) were calculated on the real datasets.

The *F-measure* parameter β was set as 0.5 emphasizing the importance of precision due to the nature of the problem, that is, the number of pre-ictal samples is reduced compared to the remainder of the data.

4.4 Results and Discussion

In the following subsections, we present the results and detail the discussion concerning the two-dimensional histograms. We also evaluate and discuss the reconstruction of the state-sequence using the trained HMM.

4.4.1 Feature selection, POI and two-dimensional histograms

First, the optimal pre-ictal period and frequency sub-band were estimated taking into account the results obtained using the Wilcoxon rank-sum test. The results are presented in Figure 4.8 and summarized in Table 4.2. Based on these criteria, we classified the samples of the frequency sub-band selected (the POI related to the maxima or minima) into four classes according to the optimal pre-ictal period.

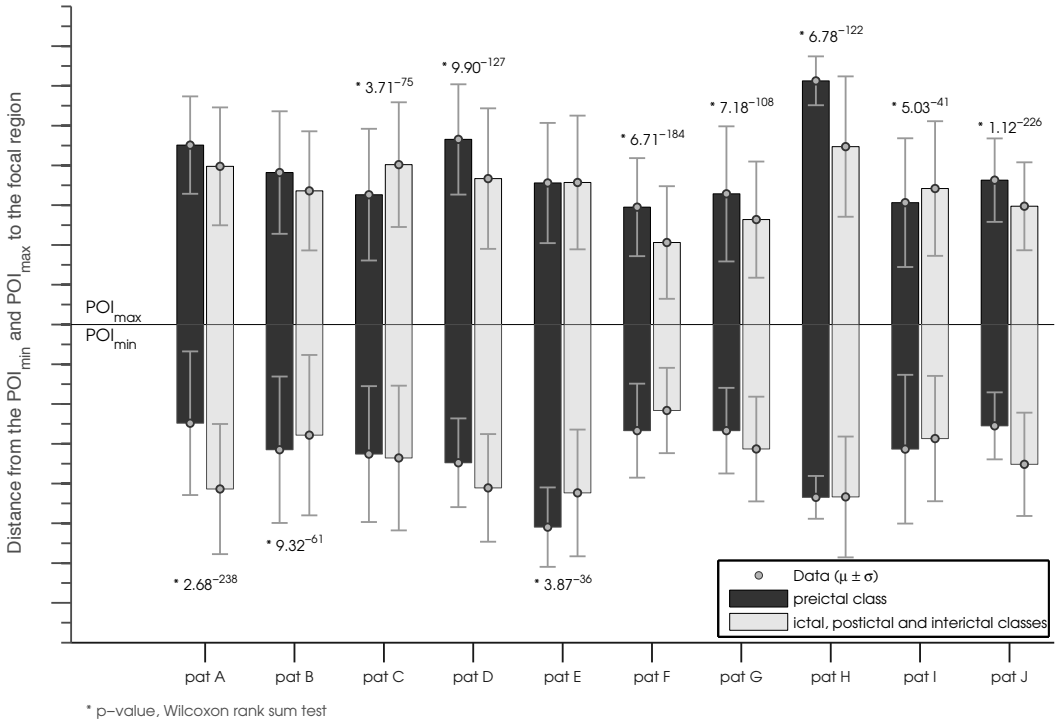


Figure 4.8 – The p-value of the Wilcoxon rank-sum test. The bars represent the mean distance between the POI and the focal region according to the classes pre-ictal and non pre-ictal. According to the analysis of the differences between classes, minima sequence was selected for patients A, B and E, while the maxima sequence was selected for the rest off the patients. The frequency band selected for each patient is described in Table 4.2.

Looking at Table 4.2, we can point out that no particular pattern was identified in frequency sub-bands selection, apart from the absence of the delta

frequency sub-band. Furthermore, the method selected the POI sequence related to the minimum areas for three patients (patients A, B and D) and the POI sequence related to the maximum for the rest of the patients.

4.4.1.1 *Two-dimensional histograms*

After obtaining and selecting the sequence of maps representative of the temporal evolution of POI, these were grouped according to their class to design two-dimensional histograms, one for each class (inter-ictal, pre-ictal, ictal and post-ictal).

We did not observe any correlation between the states and the distance between the POI and the focal region. The distribution of the POI in the pre-ictal state presents a different behavior for each patient (Figure 4.8 allows to conclude that the POI are closer to the focal region in six patients and farther from the focal region in the rest of the patients), interestingly the POI labeled as pre-ictal were mostly identified in areas outside the focal region, either contralateral or extra-focal. This suggests that different areas may be involved in seizure generation or that these are part of epileptic networks (Kuhnert et al., 2010; Feldwitch-Drentrup et al., 2011).

The POI distribution for each class presents different patterns. The inter-ictal and post-ictal POI distributions present several regions related to extremes (maxima or minima). These are usually spread through the map. The pre-ictal class has fewer relevant areas. In the ictal class the regions are smaller. Patient A is given as example in Figure 4.5.

4.4.2 Modeling of POI using trajectory pattern HMM

In order to evaluate if the trajectory defined by the POI can identify the sequence of states, i.e., inter-ictal to pre-ictal, pre-ictal to ictal, ictal to post-ictal and post-ictal to inter-ictal (i.e., the target), we trained an HMM (as described in Figure 4.7) based on the sequence of POI.

The estimation of clusters based on POI over time allows the creation of asymmetries in the map. In fact, the distribution of new reference points (compared to a normal 10-20 electrode positioning) increases the spatial resolution in regions with a higher number of POI and decreases the spatial resolution in regions with lower number of POI.

Table 4.2 – The first column represents the patients (A to J). The following three columns in the leftmost part present the best pre-ictal period and best frequency band selected (POI sequence) according to the lowest *p*-value using the Wilcoxon rank-sum test. The measure maximizes the differences between the pre-ictal class and the remainder of the data. The last four columns in the rightmost part present a summary of the results obtained using the POI sequence selected and the HMM state sequence reconstruction for each patient.

Patient ID	**Pre-ictal period (minutes)**	**Frequency sub-band**	**Max or Min**	**Sensitivity** (%)	**Specificity** (%)	**Accuracy** (%)	***F-Measure*** (%)
A	52	Beta	Min	99.68	98.84	98.32	82.58
B	26	Beta	Min	97.86	97.03	96.51	56.05
C	60	Alpha	Max	96.25	98.24	94.05	89.68
D	60	Theta	Max	100.00	89.23	89.62	56.13
E	12	Theta	Min	91.20	88.76	85.29	9.25
F	60	Gamma	Max	97.64	81.78	81.21	23.39
G	34	Alpha	Max	71.94	82.73	77.26	22.42
H	28	Beta	Max	97.32	95.14	92.03	24.06
I	30	Gamma	Max	94.86	97.55	88.75	76.22
J	60	Alpha	Max	99.17	92.87	90.07	59.17
Average	42.2			94.59	92.22	89.31	49.89

The comparison of the results obtained using the position of the electrodes as references (without performing any segmentation) with the ones obtained using references based on POI obtained after segmentation for patient A, allowed us to conclude that the results presented by the latter are better than the ones presented by the references based on the position of the electrodes (Figure 4.9). Considering patient A, the *F-measure* of the HMM reconstruction based on the position of the electrodes is 64.29% while the performance of the method based on references using segmentation and clustering achieves 82.58%.

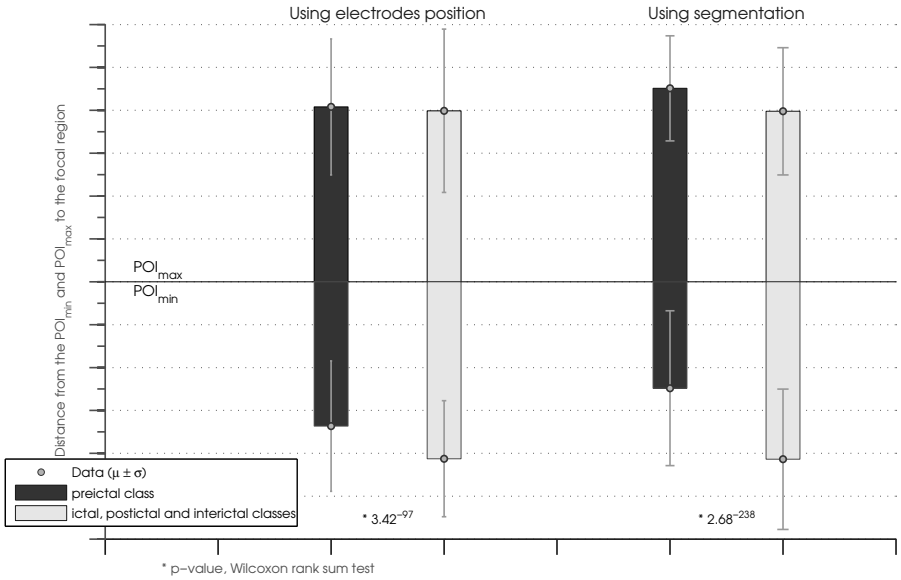


Figure 4.9 – Difference between the performance of the method using electrodes as POI and the performance using POI based on segmentation and clustering for patient A.

Using the k-means algorithm, we determined 20 clusters from the POI data and derived a trajectory based on the sequence of clusters (observations) over time.

Figure 4.10 presents the results obtained using the Viterbi algorithm to reconstruct the state sequence. The output of the models yielded, in general, very good results, summarized in Table 4.2.

Patient C presented the best results in terms of *F-measure*, although with one false positive in almost 24 hours of data. The two seizures of the dataset were correctly modeled. The *F-measure* was 89.68%, showing a good trade-off between sensitivity and specificity.

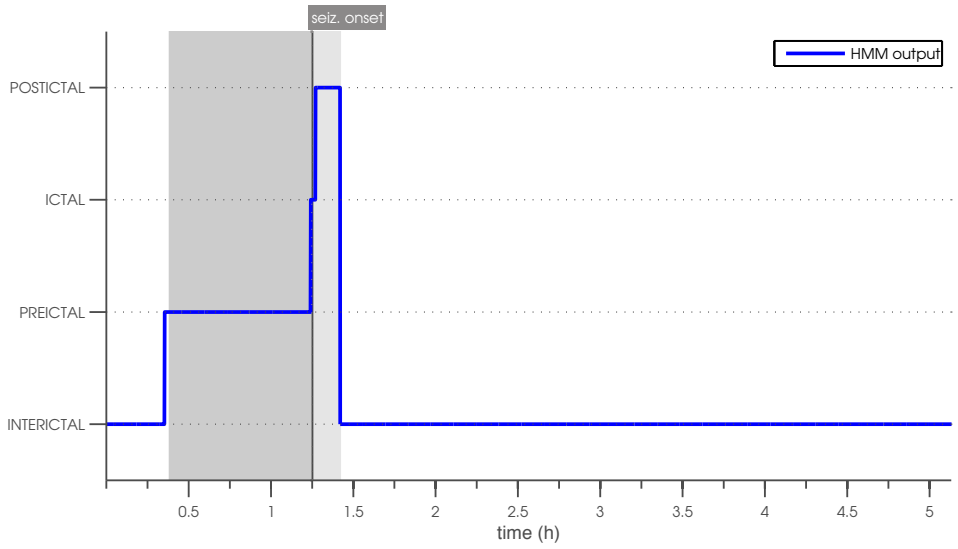


Figure 4.10 – Reconstruction of the state sequence (using the Viterbi algorithm) obtained with the HMM trained with the Baum-Welch algorithm; a 5 hours segment of data of patient A is shown in the blue line while the dark gray, and light gray backgrounds represent the pre-ictal (52 minutes), and post-ictal periods considered in the target. The target presented is defined according to the optimal parameters determined.

In terms of sensitivity, patient D achieved 100%, all pre-ictal points were correctly classified. The estimated model, although presenting one false positive, correctly captured two seizures.

Patient G presented 6 seizures along with 63 hours of data. The results were slightly worse in terms of sensitivity (the output of the HMM did not match exactly the target proposed). However all the seizures were correctly identified.

The duration of patient H recordings was around 112 hours (longer inter-ictal segments among the analyzed patients) and the results show an accuracy of 92.03%.

The HMM-Viterbi output was able to identify the different brain states in all 30 seizures analyzed, however the output yielded several false positives as well. The mean sensitivity of the point-to-point analysis was 94.59%, the mean specificity was 92.22%, the mean accuracy was 89.31% and the mean *F-measure* was 49.89%.

4.5 Conclusion and Discussion

We developed a method to model the relative power in frequency bands based on topographic maps using HMM and presented the results for 10 patients affected by refractory focal epilepsy. Previous studies based on long-term EEG data have focused in 'non spatial' data. The proposed method is intended to improve seizure prediction by adding spatial information to previous approaches of Chapter 3. Considering the seizure onset areas identified by the neurologists, one can state that the method is able to localize EEG changes allowing the identification of different regions that may be related with the generation of focal epileptic seizures. All patients were monitored through scalp EEG and the datasets used in this study totalized 497.3 hours. Each channel was processed by FFT and the topographic maps based on the relative spectral power of five frequency bands were generated iteratively. Points of interest were estimated over time taking into account the highest and lowest values of each map. Two approaches were followed to analyze data: the analysis of two-dimensional histograms for the different brain states assumed (inter-ictal, pre-ictal, ictal and post-ictal) and the training of an HMM to model the trajectory of the POI.

Statistical analysis was used to determine the frequency sub-band that better described the pre-ictal period, as well as to find the optimal pre-ictal period. The mean pre-ictal period, 42.2 minutes, suggests that longer pre-ictal periods may provide a better separability between classes.

The two-dimensional histograms based on the POI location allowed us to conclude that different spatio-temporal patterns occur in each epileptic brain state, also suggesting a relationship between the onset zone and the position of the POI throughout the pre-ictal, ictal and post-ictal periods. According to the pre-ictal periods considered for each patient, the relation between the onset area and the computed POI may develop several minutes before the seizure onset. Furthermore, the variations between the inter-ictal and pre-ictal states are not restricted to the onset zone, which suggests that areas outside this region may also be involved in the seizure generation process.

The results obtained by the Viterbi reconstruction of the data using the trained HMM based on the sequence of observations also suggest the existence of specific state-related patterns. The algorithm for the modeling of the trajectory patterns allowed us to hypothesize that the trajectory of the POI might be used to

model the sequence of brain states in epileptic patients. The average accuracy of 89.31% indicates that the proposed method can accurately model the spatio-temporal dynamics underlying the four states.

In summary, the method achieved promising results in the modeling of the epileptic brain states: the two-dimensional distribution of data presented variations between states and the trained HMM was able to learn the transition between the proposed sequence of states.

A real-time approach would rely on a preliminary patient-specific train stage in which the probabilities e_{ij} and q_{ij} would be estimated (as well as optimization of parameters). Then based on the trained model, it would be possible to generate a probabilistic tool for seizure prediction by identifying patterns related to the transition to pre-ictal period.

In Chapter 5 we further explore the spatio-temporal patterns related to the pre-ictal state using multi-way analysis. This technique represents an alternative mathematical framework for the analysis of the space-time-frequency space. The goal is to characterize the pre-ictal state by means of the temporal, spectral and spatial signatures, while evaluating their variations in the pre-ictal period.

5. LOOKING FOR THE ICTAL EVENTS BY MULTI-WAY ANALYSIS

Multi-way analysis is a technique developed in the 1920's (Hitchcock, 1927). Its use in multichannel EEG recordings (namely in the analysis of the space, time, and frequency decomposition of the EEG) dates back to 2004 (Miwakeichi et al., 2004). Later, in 2006, Acar and colleagues applied the technique to the "understanding of the structure of epileptic seizures" (Acar et al., 2006). We introduce it here for tentative identification of the pre-ictal state.

We hypothesize that the decomposition of multi-way datasets (space-time-frequency representation of EEG recordings) can be used to identify the properties of the pre-ictal period.

In the first step we construct a three-way tensor, in which the first way represents time, the second way represents frequency coefficients and the third way represents the channels. We then use the PARAFAC model, to identify the temporal signature related to the pre-ictal period. We analyzed a set of four patients (including 33 seizures) from the EPILEPSIAE database.

The study presented in Section 5.2 has been published in C.1.

5.1 Introductory Elements of Multi-way Analysis

Recent advances in Brain-Computer Interfaces (BCI) based on EEG recording are closely related with two main aspects: (i) improvements in the acquisition of the EEG with high spatial density and high frequency ranges, (ii) development of multidimensional signal-processing techniques able to extract different information compared to the traditional one- and two-dimensional approaches (Cichocki et al., 2008a; Cichocki et al., 2008b). According to Cichocki et al. (2008b) standard two-dimensional matrix factorizations (PCA, Independent Component Analysis (ICA), Nonnegative Matrix Factorization (NMF), etc.) or

flat-world view, might be insufficient to analyze large-volume stream of data encompassing many dimensions.

Consider a multichannel EEG data. The data contain structures and higher-order dependencies related to different intrinsic variables such as the channels (spatial dimension), samples over time (temporal dimension), frequency, number of trials, condition of the patients, etc. The ability of decomposing the data using a two-dimensional approach might be insufficient to explain the complete structure of the data.

One promising approach to enhance the interpretability of multichannel EEG recordings and extract significant features, is the use of a more natural representation of the multi-dimensional data. Instead of the two-dimensional matrices, "the data can be arranged in a cube" or other higher-order representation (Bro, 1997). Multi-way data can be decomposed by several methods such as parallel factor analysis (PARAFAC) (Harshman, 1970), Tucker3, and two-way PCA after the unfolding of the multi-way data into two-dimensional matrices (Bro, 1997). These three alternatives represent multi-linear decomposition methods in the sense that the multi-way data is decomposed into sets of coefficients and components. The decomposition outputs can be used to estimate some hidden structures as well as correlation between the original variables. Some of the advantages of multi-way analysis are simpler interpretations frameworks, solution uniqueness and robustness to noise (Acar and Yener, 2009).

5.1.1 Multi-way analysis and EEG signal processing

Several authors have proposed methods based on the decomposition of multi-way arrays for the purpose of summarizing and interpreting EEG data.

Miwakeichi et al. (2004) proposed a unique decomposition, based on a three-way PARAFAC model (space, time, and frequency) (Figure 5.1). These authors compared the EEG recordings of five patients, during two different states: (i) resting and (ii) mental arithmetic. The authors reported two spectral signatures common to all patients. These were modulated by the physiological state (increase of the alpha rhythms during the resting state and increase of theta rhythms during mental arithmetic). The model decomposition allowed the authors to conclude about the localization of the differences between states.

Acar et al. (2007), De Vos et al. (2007a) and De Vos et al. (2007b), proposed a novel approach to study the "epilepsy seizure structure" based on multi-way

analysis. The authors "re-arranged multichannel ictal scalp EEG data as a third-order Epilepsy tensor with time, frequency, and channels in order to automatically localize a seizure origin" (Acar, 2007). De Vos et al. (2007b) using a similar strategy correctly localized the seizure onset zone in 92% of the cases considered, confirming the validity of the approach.

5.1.2 Multi-way array factorization and decomposition

Multi-way arrays (or tensors) are data structures with order greater than two, i.e. higher-order generalization of vectors (one-way) and matrices (two-way). We use $\underline{\mathbf{X}} \in \Re^{I_1 \times I_2 \times \ldots \times I_{N_d}}$ to represent a tensor with N_d ways. The dimension of each way is represented by $I_n, n = 1,\ldots,N_d$.

For simplicity, let us assume a three-way model with dimensions I, J and K; the element x_{ijk} of the tensor $\underline{\mathbf{X}}$ denotes the *ith* row, *jth* column and *kth* tube.

The basic sub-units of tensors are: (i) rows, columns, and tubes (one row in the third dimension) (ii) slices, two-dimensional sections of a tensor, i.e. planes defined by fixing one of the ways (Figure 5.2).

Two of the most common decompositions methods for multi-way tensors are the Tucker3 and the PARAFAC models. Both can be interpreted as generalizations of two-dimensional analysis for higher-dimensional data using additional constraints such as orthogonality, non-negativity or sparsity of hidden factors (Cichocki et al., 2008b).

We summarize these two models for fitting high-dimensional data. For simplicity, let us consider $\underline{\mathbf{X}}$ a three-way tensor.

5.1.3 PARAFAC model

To define the PARAFAC model, first we need to describe the *rank-one* tensor. According to Kolda and Bader (2009) a three-way tensor $\underline{\mathbf{X}} \in \Re^{I \times J \times K}$ is rank-one if it can be written as the outer product of three vectors (5.1),

$$\underline{\mathbf{X}} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c} \tag{5.1}$$

where the symbol $\circ$ represents the outer product operator and $\mathbf{a},\mathbf{b},\mathbf{c}$ represent column vectors with dimension I, J and K. Each element x_{ijk} of the tensor is the product of the corresponding vector elements (5.2).

$$x_{ijk} = a_i \times b_j \times c_k \tag{5.2}$$

A PARAFAC model expresses a tensor as the sum of a finite number of rank-one tensors. A given third-order tensor can be described as (5.3),

$$\underline{\mathbf{X}} \approx \sum_{r=1}^{R} \left(coef_r\right) \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r \tag{5.3}$$

where R is the number of components (Figure 5.1 illustrates this sum of components).

An alternative representation is possible using the *factor matrices*. The factor matrices are the combination of the vectors from the rank-one tensors. We can define the factor matrix as $\mathbf{A} = \left[\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_R\right]$ (and similarly for matrices $\mathbf{B}$ and $\mathbf{C}$). Using this notation, equation (5.3) can be rewritten in a matricized form (one per way) as (5.4),

$$\begin{aligned} \mathbf{X}_{(1)} &\approx \mathbf{A}\left(\mathbf{C} \odot \mathbf{B}\right)^T \\ \mathbf{X}_{(2)} &\approx \mathbf{B}\left(\mathbf{C} \odot \mathbf{A}\right)^T \\ \mathbf{X}_{(3)} &\approx \mathbf{C}\left(\mathbf{B} \odot \mathbf{A}\right)^T \end{aligned} \tag{5.4}$$

where $\mathbf{X}_{(n)}$ represents the matricization of the tensor along the *nth* way, and $\odot$ represents the Khatri-Rao product (see Kolda and Bader (2009) for a detailed explanation).

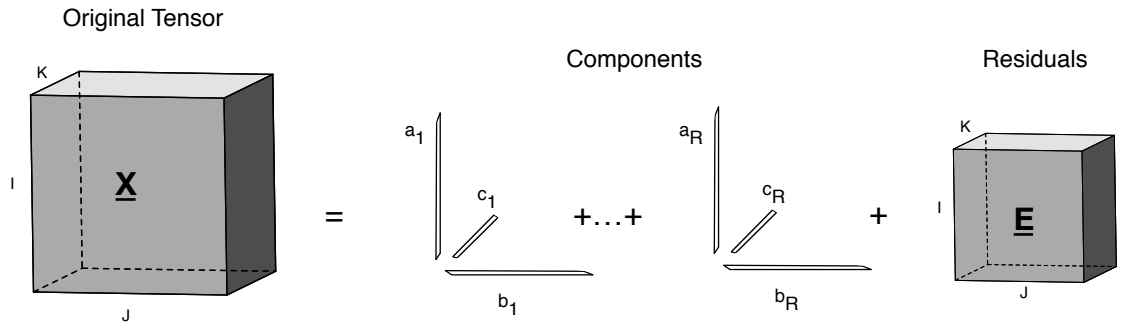


Figure 5.1 – Illustration of a PARAFAC model. The three-way array $\underline{\mathbf{X}}$ is expressed as the sum of R rank-one tensors. $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$ correspond to vectors along the first, second and third ways respectively while $\underline{\mathbf{E}}$ represents the residuals. We assume that the linear combination coefficients are scaled to the unit.

Contrarily to the two-dimensional matrix decompositions, the higher-order tensor decompositions are often unique, "with the exception of the elementary indeterminacies of scaling and permutation" (Kolda and Bader, 2009).

5.1.3.1 *Preprocessing and rank of multi-way arrays*

The appropriate rank of a PARAFAC model, i.e. the number of components, is an "open" subject in the sense that no consensus has been achieved in terms of the best method to determine the rank. In general, the methods used in two-dimensional analysis proved to be useful in multi-way tensor decomposition. Approximation error analysis, interpretability of the solution, and core consistency are some of the methods proposed. However, according to Acar and Yener (2009), no specific method determines an optimal number of components for real data.

Let us consider a three-way tensor. Two preprocessing methods are usually considered in tensor analysis, centering and scaling. According to Acar and Yener (2009), centering across one way of the tensor involves matricization (Figure 5.3) in the desired way and applying two-way centering. Scaling requires dividing each slice of the tensor corresponding to a determined variable by the same scalar. According to Bro (1997), scaling within the first way can be described by (5.5).

$$x_{ijk} = \frac{x_{ijk}}{S_i}$$
$$S_i = \sqrt{\left(\sum_{j=1}^{J} \sum_{k=1}^{K} x_{ijk}^2 \right)} \tag{5.5}$$

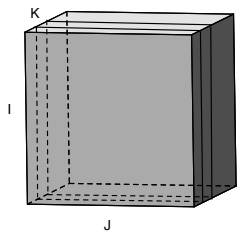


Figure 5.2 – Frontal slices in a three-way model. A slice is one of the faces of the three dimensional object (in this case we fix the third way, i.e. $\mathbf{X}_{::k}$).

5.1.3.2 *Algorithm*

The algorithms used in the decomposition of multi-way models are usually iterative. The Alternating Least Squares (ALS) is one of the first approaches for

the decomposition of the PARAFAC model (Harshman, 1970). Let us consider a three-way tensor and the rank R.

In summary, the goal is to iteratively compute the solution that represents the best approximate of $\underline{\mathbf{X}}$, i.e. (5.6),

$$\min_{\hat{\underline{\mathbf{X}}}} \left\| \underline{\mathbf{X}} - \hat{\underline{\mathbf{X}}} \right\| \tag{5.6}$$

where $\hat{\underline{\mathbf{X}}}$ represents the PARAFAC model $\hat{\underline{X}} = \sum_{r=1}^{R} (coef_r)\, \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r$.

The main idea is to iteratively fix the factor matrices of two ways, let us say $\mathbf{B}$ and $\mathbf{C}$, and solve the optimization problem for the third, in this case $\mathbf{A}$. Then fix $\mathbf{A}$ and $\mathbf{C}$, and solve the problem for $\mathbf{B}$ and finally, fix $\mathbf{A}$ and $\mathbf{B}$ solving the problem for matrix $\mathbf{C}$ (Kolda and Bader, 2009).

The unknown set of parameters is estimated according to the least-squares criterion and the procedure is repeated until one of stopping criteria is achieved.

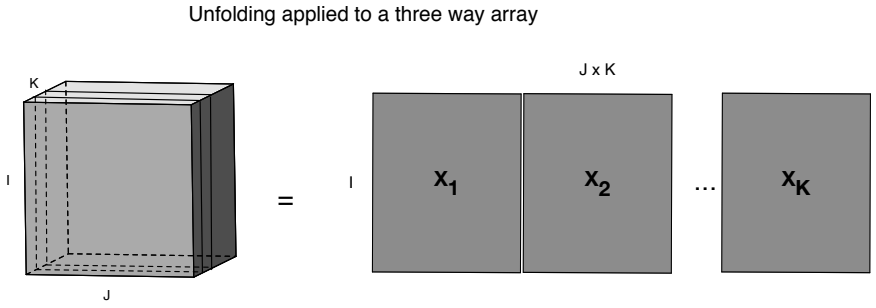


Figure 5.3 - Unfolding or matricization is one of the main steps of the ALS algorithm. This operation transforms a three-way tensor into a two-dimensional matrix allowing performing common two-dimensional operations.

Let us consider the PARAFAC model defined by equations (5.4), and in particular the first equation therein. Considering $\mathbf{B}$ and $\mathbf{C}$ fixed, we can rewrite the minimization problem (5.6) as (5.7),

$$\min_{\hat{\mathbf{A}}} \left\| \mathbf{X}_{(1)} - \hat{\mathbf{A}} \left(\mathbf{C} \odot \mathbf{B} \right)^T \right\|_F \tag{5.7}$$

where $\left\| \underline{\mathbf{X}} \right\|_F$ represents the Frobenius norm of a tensor. An estimate of the solution is given by (5.8).

$$\hat{\mathbf{A}} = \mathbf{X}_{(1)}\left[(\mathbf{C} \odot \mathbf{B})^T\right]^{\dagger} \tag{5.8}$$

The symbol † represents the Moore-Penrose pseudo-inverse.

A common stopping criterion is to determine when the variation between consecutive iterations is below a certain threshold, e.g. 10^{-6} (Algorithm 5.1).

Algorithm 5.1 – Three-component PARAFAC model and the Alternating Least Squares algorithm (adapted from Bro, 1997).

i. Define the number of components R
ii. Initialize factor matrices $\mathbf{B}$ and $\mathbf{C}$
iii. Estimate $\mathbf{A}$ from $\hat{\mathbf{A}} = \mathbf{X}_{(1)}\left[(\mathbf{C} \odot \mathbf{B})^T\right]^{\dagger}$
iv. Estimate $\mathbf{B}$ from $\hat{\mathbf{B}} = \mathbf{X}_{(2)}\left[(\mathbf{C} \odot \mathbf{A})^T\right]^{\dagger}$
v. Estimate $\mathbf{C}$ from $\hat{\mathbf{C}} = \mathbf{X}_{(3)}\left[(\mathbf{B} \odot \mathbf{A})^T\right]^{\dagger}$
vi. Repeat from iii. until convergence criteria is met (little change between consecutive iterations, number of iterations, etc.)

5.1.3.3 *Initialization*

The algorithm does not guarantee convergence to the global minimum of the function. The selection of good candidates as starting points can potentially benefit the model in different aspects, e.g. easy of convergence, interpretability, and reaching the global minimum.

Multiple runs using random initialization may help to conclude about convergence of the method.

5.1.3.4 *Constraints*

Imposing constraints in the PARAFAC model will inevitably decrease the fit of the solution. However, this decrease in the approximation of the solution may be justifiable in terms of stability and interpretability of the solution. The assumptions associated with the constraints in terms of interpretability and realism may justify the decrease in the fit of the solution (Bro, 1997).

The use of non-negativity constraints is usually based on the knowledge of data. In fact, the idea behind EEG data decomposition is usually related to the estimation of the sources. The sources should contribute positively to the final

EEG recordings and features extracted. Non-negative constraints in neurophysiologic data has been the focus of several researchers (Lee et al., 2007; Cichocki et al., 2008a; Lee and Choi, 2009). According to Cichocki et al. (2008b), the components estimated using non-negative tensor factorization (NTF) may be associated to the underlying conditions (with mixed spatial, temporal and spectral information) without prior knowledge on the data.

5.1.4 Tucker Model

Tucker (1966) proposed an alternative to the PARAFAC decomposition model. The Tucker model can be considered as a higher-order generalization of the more restrictive PARAFAC. Let us consider the three-way tensor $\underline{\mathbf{X}} \in \Re^{I \times J \times K}$. The main idea is to estimate a core tensor $\underline{\mathbf{G}}$ that allows more degrees of interaction between the ways (Kolda and Bader, 2009) (5.9).

$$\underline{\mathbf{X}} = \sum_{f=1}^{F} \sum_{u=1}^{U} \sum_{h=1}^{H} (coef_{fuh})\, \mathbf{a}_f \circ \mathbf{b}_u \circ \mathbf{c}_h \tag{5.9}$$

Equation (5.9) can be rewritten using the factor matrices (5.10),

$$\underline{\mathbf{X}} = \underline{\mathbf{G}} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \tag{5.10}$$

where $\mathbf{A} \in \Re^{I \times F}$, $B \in \Re^{J \times U}$ and $\mathbf{C} \in \Re^{K \times H}$ are the factor matrices corresponding to the first, second and third ways, respectively and $\underline{\mathbf{G}} \in \Re^{F \times U \times H}$ is the core array (representing the mixture coefficients). In this model, instead of R components for all ways, the number of components for each way may be different, i.e. F ,U and H, respectively. The core tensor summarizes the interactions between the ways and can be significantly smaller than the original tensor (Kolda and Bader, 2009). An illustration of the Tucker3 model is given in Figure 5.4.

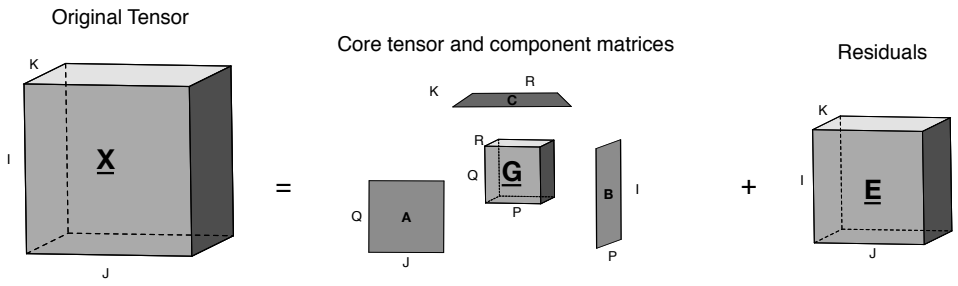


Figure 5.4 – Illustration of the Tucker3 model. The tensor $\underline{\mathbf{E}} \in \Re^{I \times J \times K}$ corresponds to the residuals.

According to Cichocki et al. (2008b), the PARAFAC model represents a simplification of the Tucker model in the sense that the core tensor $\underline{\mathbf{G}}$ is reduced to a super-diagonal tensor (all elements are zero except the elements on the super-diagonal, which act as scaling factors). The full-core tensor structure allows exploring the underlying structure of the multi-way array.

However, the Tucker model also presents some drawbacks. The model is computationally more complex and lacks of uniqueness. The ability to interpret the components and their interactions is also more difficult (Acar and Yener, 2009).

Different algorithms were proposed to decompose Tucker models (see Kolda and Bader (2009) for a review).

5.2 Space Time Frequency (STF) Tensor for the Characterization of the Epileptic Pre-Ictal Stage

Using principles of multi-way data analysis, we present a mathematical framework to characterize the pre-ictal state. The characterization of the pre-ictal period in terms of spatial, temporal and spectral properties is extremely important, for example, to define restricted areas of interest (spatial domain), important frequency sub-bands (frequency domain), etc. This information would also be important to select a subset of features for classification algorithms.

According to Mormann et al. (2005) there is a decrease of power in the delta frequency band in the pre-ictal period in comparison to the inter-ictal period, also accompanied by an increase in the remaining bands in certain EEG channels.

Along these lines, we address the problem of pre-ictal characterization using a STF tensor. Once the STF tensors are constructed, we explore them using PARAFAC models and use their decomposition components to identify temporal, spatial and spectral signatures related to the pre-ictal period. An illustration of the method proposed is presented in Figure 5.5.

For the purpose of multi-way analysis, we used the N-way toolbox for MATLAB (Andersson, 2000). The toolbox contains a comprehensive set of tools to model, decompose and analyze multi-way datasets.

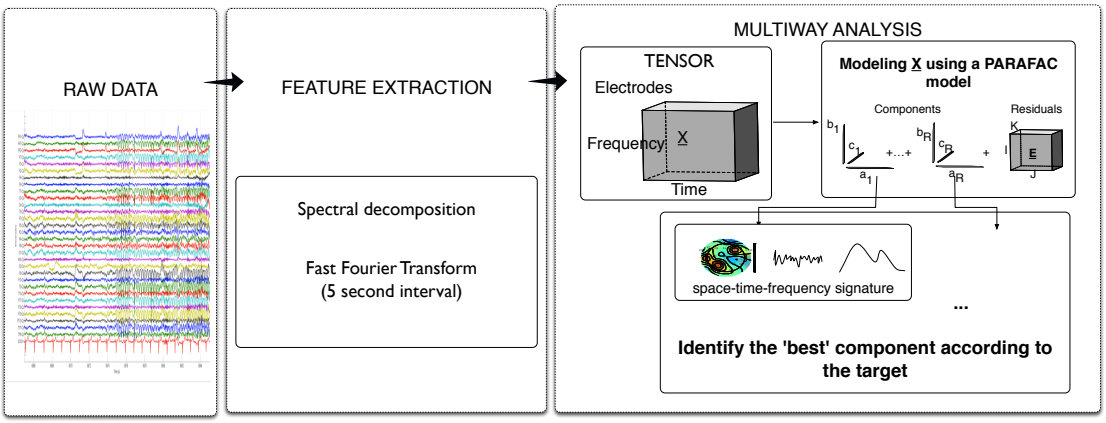


Figure 5.5 – Illustration of the method proposed. The first step is the computation of the spectral features. After the re-arrangement of the data into a three dimensional structure, we used the PARAFAC model and decomposition strategy to determine the 'best' component.

5.2.1 Dataset description

The data are from four patients affected by focal refractory epilepsy. The dataset is part of the EPILEPSIAE database (see Section 3.2.1). The patients were randomly selected among the patients of the database with continuous (without intervals in the data) scalp recordings.

The STF tensor consists (one for each seizure) in a three hours segment (two hours before the seizure onset and one-hour containing ictal and post-ictal data). An overview of the data is presented in Table 5.1.

Table 5.1 – Overview of the dataset

Patient	Sampling frequency (Hz)	Number of seizures	Focal onset region	Number of channels
A	256	9	P4, O2, P8, PZ	27
B	256	9	F3, C3, P3, F8, T8, P8	19
C	256	7	FP1, FP2, F3, F4, C4, P3	19
D	256	8	F3, C3, F7, T7	19

5.2.2 Feature Extraction and STF tensor construction

Multi-channel EEG data originally form a two-dimensional matrix of time and space. This representation of the EEG recordings allows a time domain

analysis. To obtain a time-frequency representation of the data, we apply FFT on the signal, using a five second non-overlapping sliding window in order to quantify the frequency components. In each window we compute the relative power in the different bands: δ_{rp} (0.5 - 4 Hz), ϑ_{rp} (4 - 8 Hz), α_{rp} (8 - 15 Hz), β_{rp} (15 - 30 Hz) and γ_{rp} (30 Hz to *Nyquist frequency*), as described in Section 3.2.2.5.

The STF tensor is a three-way time-varying EEG spectrum array $\underline{\mathbf{X}} \in \Re^{N \times J \times K}$ where N, J and K are the number of samples, number of frequency bands and number of channels, respectively. Once constructed, we model the tensor using a R-component PARAFAC model as in equation (5.3), essentially due to its simplicity and interpretability.

As preprocessing steps, we scaled and centered the tensor within the electrode way. According to Acar et al. (2007), the optimal number of components is three. Taking into account these conclusions, we designed the PARAFAC decomposition model with $R = 3$.

Different approaches were proposed in the literature to determine the initialization values of the PARAFAC model, however none of the methods guarantees convergence to the global minimum. In this study, we used random initialization values and performed three runs to confirm convergence. We also imposed non-negative constraints (NTF) in the temporal, spectral and spatial ways. The convergence criterion defined to interrupt the algorithm was defined as a maximum difference of 10^{-6} between consecutive iterations.

5.2.2.1 *PARAFAC decomposition analysis*

We hypothesize that the spectral features of EEG recordings present pre-ictal variations. Therefore, we can argue that one (or more) components of the PARAFAC model represent these pre-ictal variations. Considering a STF tensor, the pre-ictal period has distinct signatures in space, time, and frequency. Once the temporal signature that reflects the expected pre-ictal variations is identified (see Figure 5.6), we analyze the space and frequency signatures of that component.

The pre-ictal period used to determine the best component was set as 20 minutes (different studies suggest that functional (Federico et al., 2005) and temporo-spatial patterns vary during the 20 minutes prior to seizures (Lange et al., 1983)). Considering this pre-ictal period, we determined the correlation coefficient between the temporal signature of each component and the target proposed. The

best correlation coefficient is used for the selection of the best component, i.e. to determine the component that reflects the pre-ictal variations.

The correlation between a quantitative measure (temporal signature) and a dichotomous (a target) is a particular case of Pearson correlation coefficient and is also known as point-biserial (Tate, 1954). The coefficient ρ represents the strength of association between the two variables and can be either positive or negative. To determine if the correlation between variables is significant we used hypothesis testing at the significance level of 0.05.

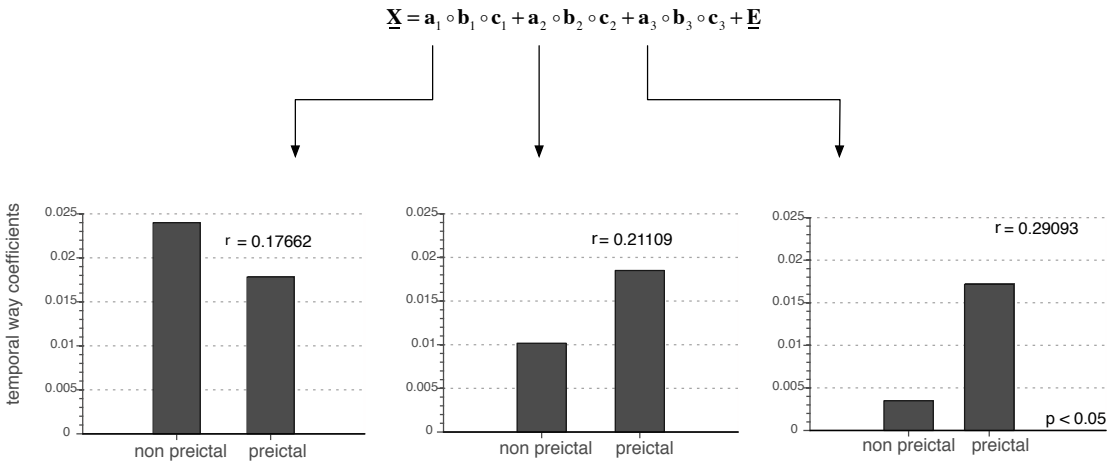


Figure 5.6 – Selection of the best component based on the correlation coefficient between the temporal way of each component and the target. The higher coefficient component represents an higher association strength between the temporal signature and the target.

A three-way PARAFAC model on the tensor $\underline{\mathbf{X}}$ extracts the components $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$ (for $r = 1, 2, 3$). The contributions of time, frequency and electrodes are represented by $\mathbf{a}_r$, $\mathbf{b}_r$ and $\mathbf{c}_r$, respectively (Figure 5.7).

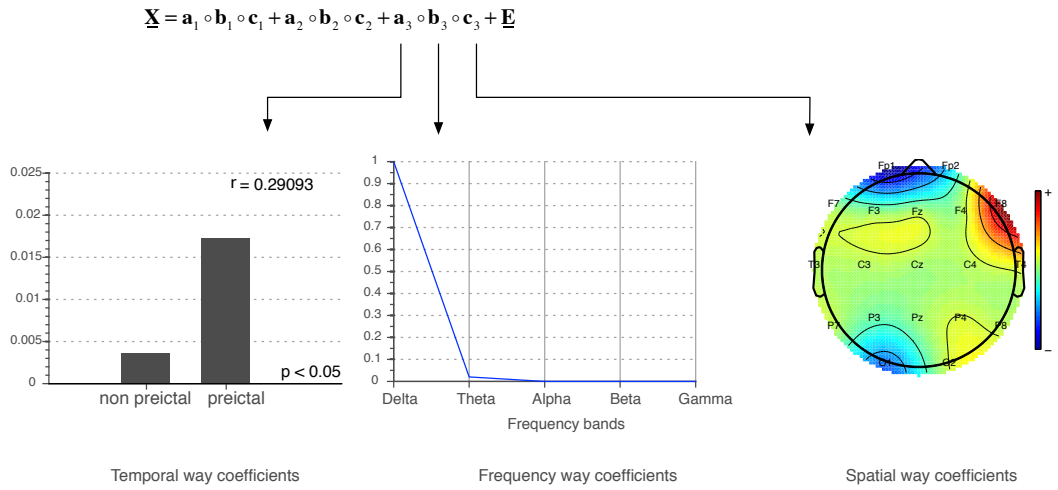


Figure 5.7 – According to the correlation coefficient ρ, the best component in the example presented in Figure 5.6 is r=3. The different signatures of the component are presented here. The results suggest that the pre-ictal variations occur around electrode F8 and in low frequency bands.

5.2.3 Results

We analyzed 33 seizures from four patients. Based on the arguments explained previously, we explored the characteristics of the time, frequency and space signatures.

Considering the analysis of the temporal way, 32 seizures present significant correlation with the target proposed (only one seizure, seizure 6 from patient D, did not present significant correlation). This indicates that there are variations in the STF tensor related to the 20 minutes pre-ictal period suggested.

Based on the hypothesis that these components represent the pre-ictal variations, we analyzed the frequency and space signatures.

Let us consider the spatial signatures exhibited for each seizure of patient C (Figure 5.8). We observe that frontal electrodes exhibit a strong influence (positive or negative) in almost every seizure analyzed. The seizure origin was clinically defined as frontotemporal left for seizures two, three, five, and six; and frontotemporal right for seizures one, four and seven. There is some association between the clinically assessed seizure onset zones and the spatial signatures obtained.

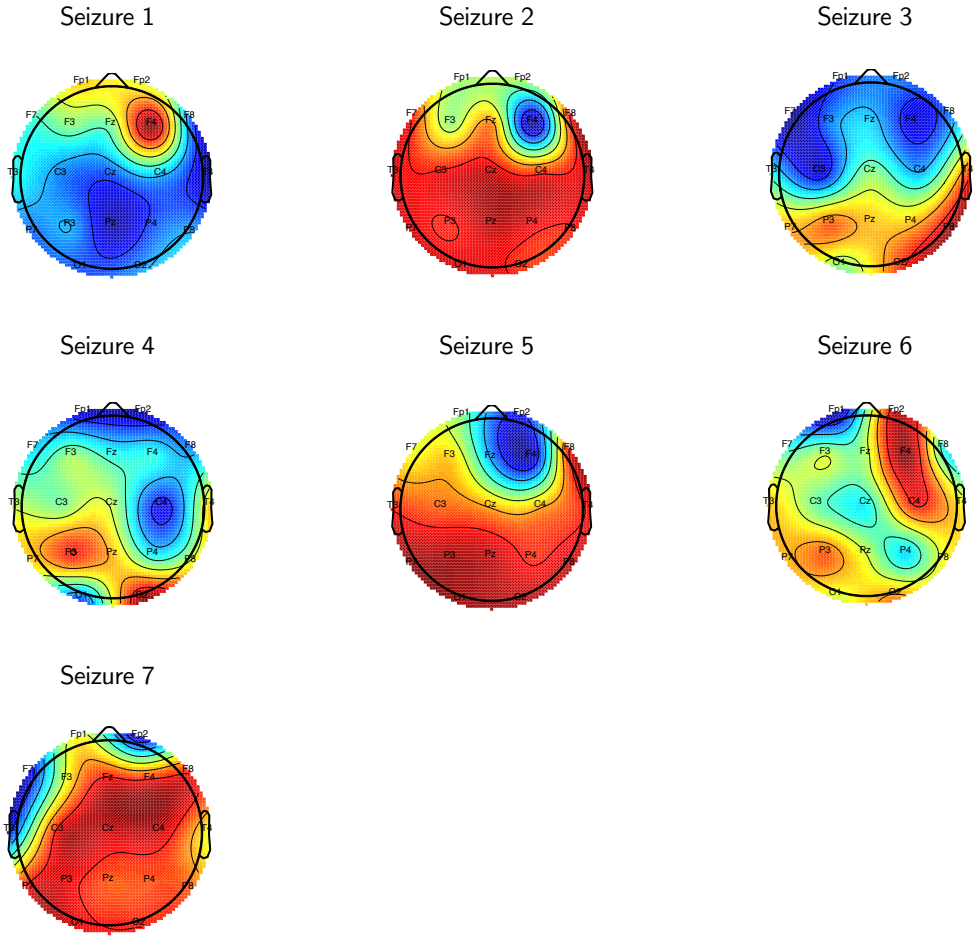


Figure 5.8 – Spatial signatures of the seizures analyzed from patient C. First, we identified the best component. We created the topographical representation of each spatial signature to identify areas of strong influence.

The spatial signatures related to the seizures analyzed from patient A and B present a similar behavior, that is, they present some degree of correspondence between the areas highlighted and the seizures onset zones clinically identified. Patient D presents multiple highlighted regions (some occurring bilaterally) including the areas identified as the seizure onset zones (Figure 5.9).

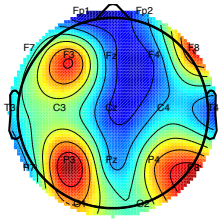


Figure 5.9 – Spatial signature of the third seizure from patient D. The topographic analysis of the spatial signature highlights bilateral frontal and parietal areas.

The third way, that represents the frequency signature, can also be used to characterize the pre-ictal variations. For example, considering the frequency signatures of patient D, low frequency bands are much more predominant than the rest (Figure 5.10). Patient B exhibits different behaviors among the seizures analyzed. Patient C presents a dominance of low frequencies in its frequency signatures. Finally, patient A presents a balance between the different frequency bands considered.

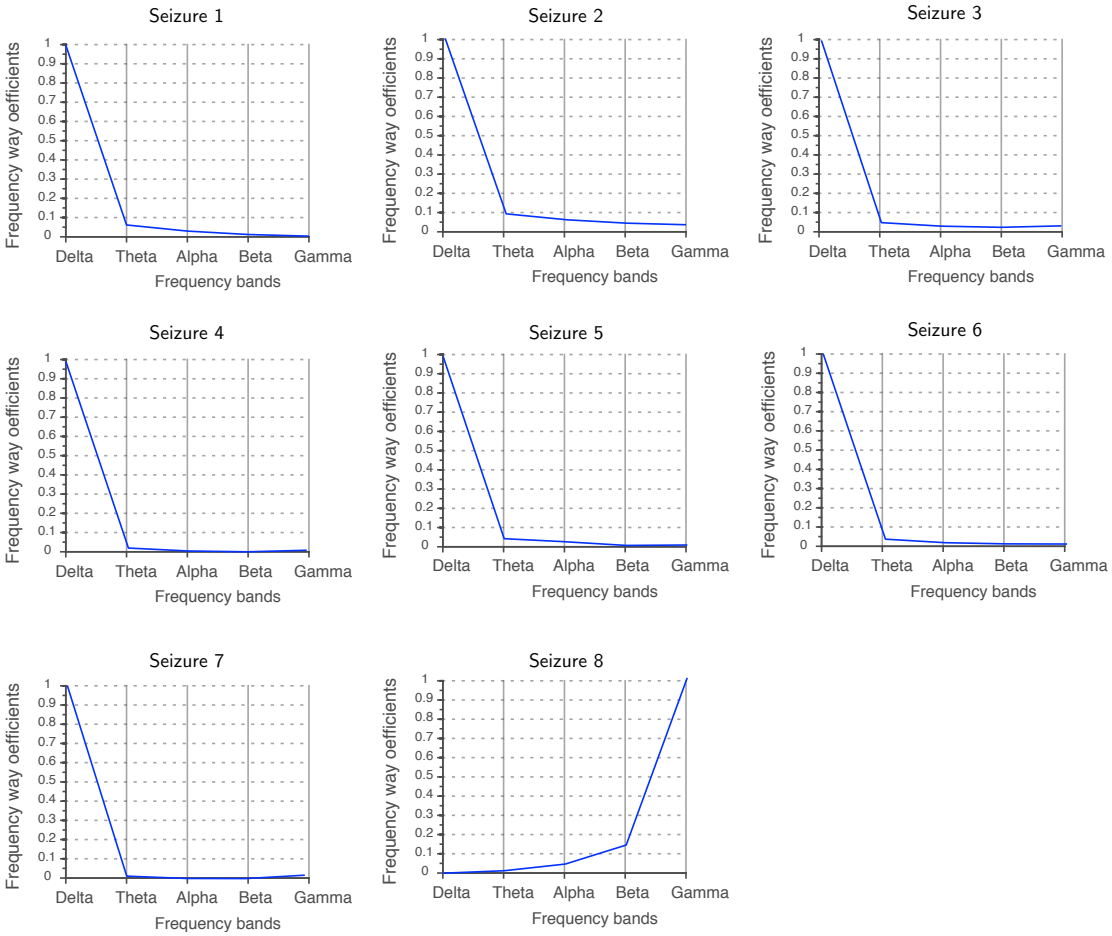


Figure 5.10 - Frequency components of the seizures from patient D. The low frequencies are predominant except in seizure 8.

5.3 Incremental Tensor Analysis and Epileptic Seizure Prediction

Despite of the ability of multi-way models to fit high dimensional data into a single structure, they represent a static vision of the data. Powerful as they may be, the decomposition algorithms are not suitable for the analysis of large, time-evolving data.

Focusing on incremental and on-line applications of multi-way analysis, Sun et al. (2006) proposed the general concepts of tensor stream (Figure 5.11) and incremental tensor analysis (ITA).

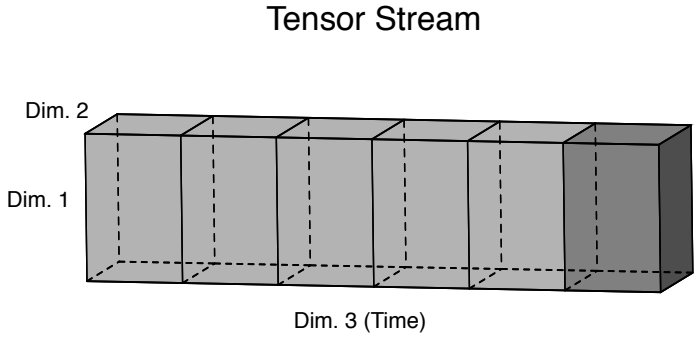


Figure 5.11 – Incremental tensor analysis. The methods allow analyzing the dynamic aspect of tensor stream data.

According to Sun et al. (2008), ITA incrementally summarizes each tensor of the tensor stream as core tensors associated with factor matrices (Figure 5.4). The core tensor can represent a significant summary of the original input tensor and the factor matrices represent the transformation between the original tensor and the core tensor (see equation (5.10)). As new input tensors arrive, factor matrices can be updated. These updates contain important information related to each way.

5.3.1 Algorithm

Sun et al. (2008) introduced Dynamic Tensor Analysis (DTA) – incrementally maintains covariance matrices for all ways and uses the leading eigenvectors of covariance matrices as factor matrices.

The incremental update efficiently avoids redundant computation and storage of data introducing prior knowledge to the decomposition of the new tensor.

5.3.1.1 *Incremental Tensor Analysis*

A tensor sequence is a sequence of R^{th} order tensors $\underline{\mathbf{X}}_t, t = 1,...,T$, where each $\underline{\mathbf{X}}_t \in \Re^{I_1 \times ... \times I_{N_d}}$. The R^{th} order tensor sequence can be viewed as a single $(R+1)^{th}$ order tensor, if one considers that the additional way has a dimensionality T.

Let us consider a tensor stream, and for simplicity, assume that all the new tensors are continuously arriving over time and that they are of the same size. Because of the arriving of new data, it is not possible to determine optimal factor matrices. Moreover, the characteristics of time varying data are likely to change over time.

Unlike the static multi-way model approach presented in Section 5.1 (which assumes that it is possible to determine optimal factor matrices), "it is not feasible to find optimal" factor matrices because of the new data arriving (Sun et al., 2008).

Given a new tensor $\underline{\mathbf{X}}_t \in \Re^{I_1 \times ... \times I_{N_d}}$, and previously estimated factor matrices, the main idea of ITA is to update the factor matrices $\mathbf{U}_i \in \Re^{I_i \times R_i}, i = 1,...,N_d$ and the core tensor $\underline{\mathbf{G}}_t$ such that the reconstruction error E_r (5.11) is minimized.

$$E_{r_t} = \left\| \underline{\mathbf{X}}_t - \underline{\mathbf{G}}_t \prod_{i}^{N_d} \times_i \mathbf{U}_i \right\| \tag{5.11}$$

To update the model, we used the DTA method (Sun et al., 2008) for its simplicity and simple computational requirements.

5.3.1.2 *Dynamic Tensor Analysis (DTA)*

The algorithm is based on the covariance matrices of each way, from which the factor matrices can be easily computed (Figure 5.12) (Algorithm 5.2).

Let (5.12) represent the covariance matrix of the *ith* way of $\underline{\mathbf{X}}$.

$$\mathbf{C}_i = \mathbf{X}_{(i)}^T \mathbf{X}_{(i)} \tag{5.12}$$

From the covariance matrices (5.12), the factor matrices can be obtained through a diagonalization.

The covariance matrix of the *ith* way is updated according to (5.13).

$$\mathbf{C}_i \leftarrow \mathbf{C}_i + \mathbf{X}_{(i)}^T \mathbf{X}_{(i)} \tag{5.13}$$

According to Sun et al. (2008), the updated factor matrices can be computed by diagonalization (5.14),

$$\mathbf{C}_i = \mathbf{U}_i \mathbf{S}_i \mathbf{U}_i^T \tag{5.14}$$

where $\mathbf{S}_i$ is a diagonal matrix.

Time dependent applications usually need to consider a forgetting factor to deal with the balance between new and previous data. A forgetting factor (constrained to the interval [0,1]) may be included in (5.13), resulting in the covariance matrix (5.15).

$$\mathbf{C}_i \leftarrow \lambda \mathbf{C}_i + \mathbf{X}_{(i)}^T \mathbf{X}_{(i)} \tag{5.15}$$

Algorithm 5.2 – main steps of the DTA algorithm (adapted from Sun et al., 2008)

i. / reconstruct the old covariance matrix
   $\mathbf{C}_i = \mathbf{U}_i \mathbf{S}_i \mathbf{U}_i^T$
ii. / update covariance matrix
   $\mathbf{C}_i = \lambda \mathbf{C}_i + \mathbf{X}_{(i)} \mathbf{X}_{(i)}^T$
iii. Set $\mathbf{U}_i$ as the top eigenvectors of $\mathbf{C}_i$
iv. Repeat from i. for all ways $i = 1,...,N_d$
v. / Estimate the core tensor
   $\underline{\mathbf{G}} = \underline{\mathbf{X}} \times_1 \mathbf{U}_1^T \times_2 ... \times_{N_d} \mathbf{U}_{N_d}^T$

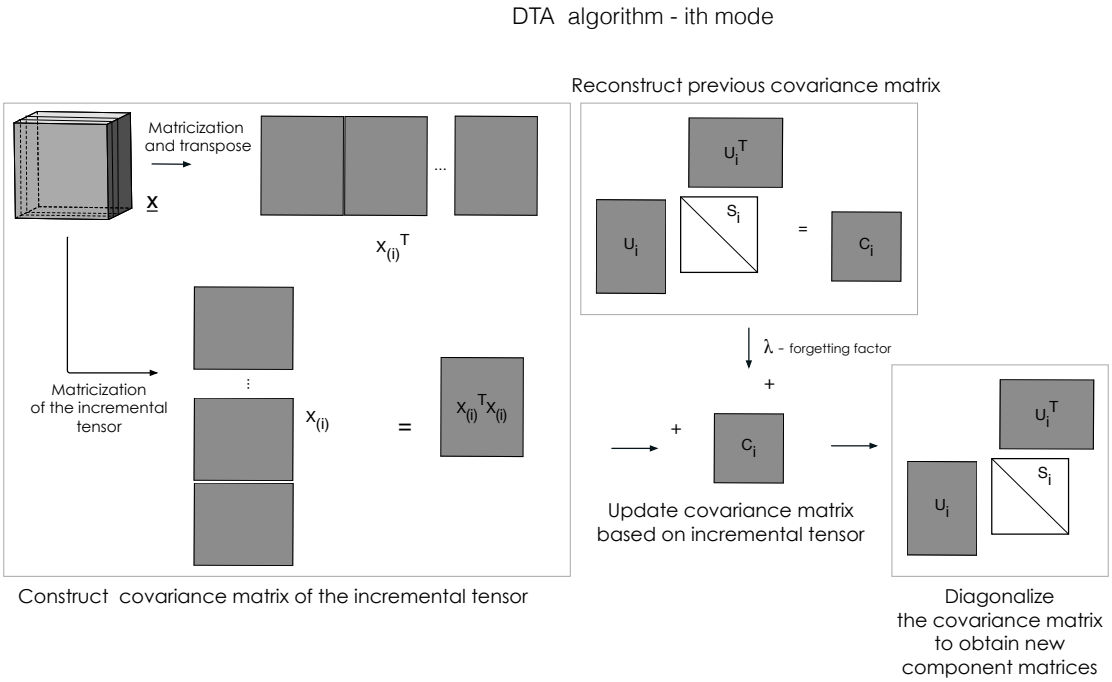


Figure 5.12 – DTA decomposition. The new tensor is matricized along the *ith* way. The new factor matrices are obtained through the weighted sum of two contributions (reconstruction of the previous covariance matrix and contribution from the new tensor).

5.3.2 STF tensor analysis using DTA

The identification of abnormalities in stream data (such as long term EEG recordings) is one possible application of DTA. It involves the identification of sudden variations in the stream data.

To identify abnormal tensors we can use the reconstruction error over time (5.16).

$$E_{r_t} = \left\| \underline{\mathbf{X}}_t - \underline{\mathbf{G}}_t \prod_{i=1}^{N_d} \times_i \mathbf{U}_i \right\|_F \tag{5.16}$$

Considering the final formulation for the core tensor presented in Algorithm 5.2; we can further develop (5.16), and obtain (5.17).

$$E_{r_t} = \left\| \underline{\mathbf{X}}_t - \underline{\mathbf{X}}_t \prod_{i=1}^{N_d} \times_i \mathbf{U}_i^T \mathbf{U}_i \right\|_F^2 \tag{5.17}$$

The identification of sudden variations in the stream data can be performed using a threshold and prior knowledge. A tensor can be identified as abnormal if

the error E_r is ψ standard deviations away from the mean error so far, as in (5.18).

$$E_{r_T} \geq \mu(E_{r_t}) + (\psi)\sigma(E_{r_t}), \;\; t = 1,...,(T-1) \tag{5.18}$$

Additionally, it is possible to find correlated dimensions within the same way and across different ways of a tensor sequence.

We hypothesize that the pre-ictal stage presents variations in the space-time-frequency structure. Our objective is to determine the error over time and identify periods where the error varies above the threshold.

Using $\psi = 2$, we determined the instants where the error surpasses the threshold defined by (5.18). A seizure is considered correctly predicted if the error is above the threshold in the 20 minutes pre-ictal period. If the threshold is exceeded outside this period we consider a false positive.

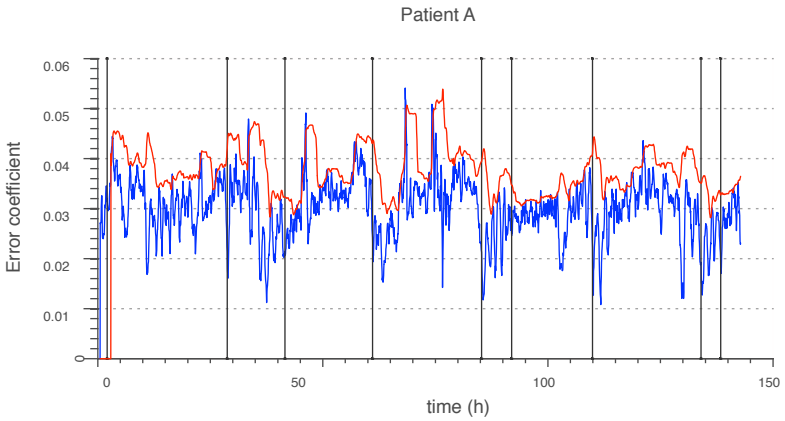
The development of the framework is based on the tensor toolbox for MATLAB (Bader and Kolda, 2012). The toolbox provides a set of classes for manipulating dense, sparse and structured tensors. The iterative decomposition procedures (ITA) are based on the MATLAB source code developed by Sun et al. (2008).

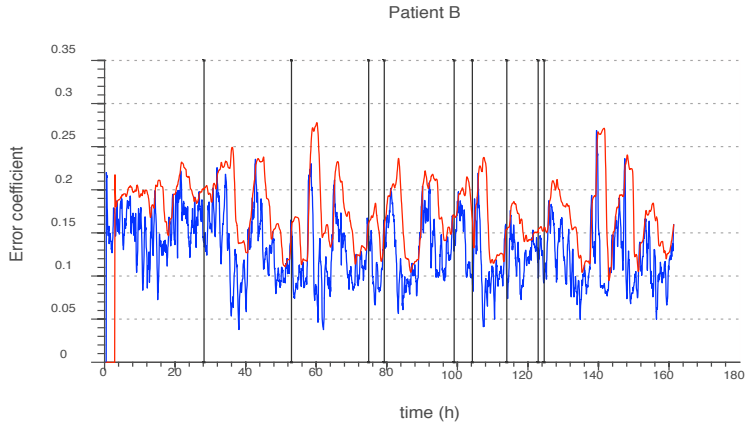
5.3.3 Experiments on a set of four patients

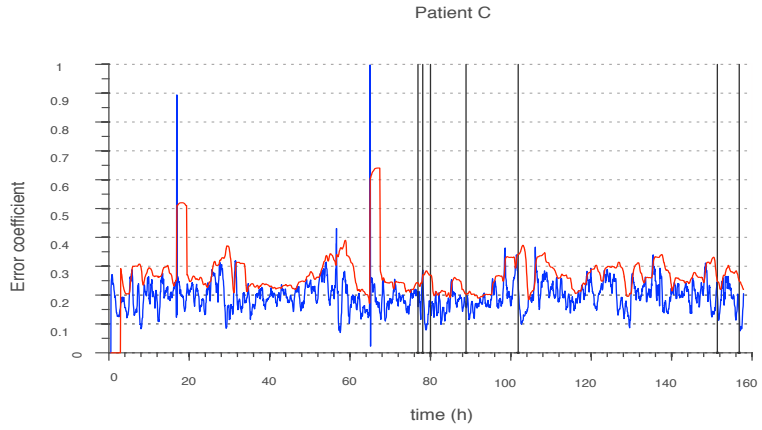
The dataset used in this study is characterized in Section 5.2.1.

The first stage of this approach is the tensor sequence construction and the definition of the ITA parameters. We defined the length of each tensor as 360 (representing 30 minutes of data), and used a shift of 10% (3 minutes) between consecutive STF tensors. The forgetting factor λ defined in Equation (5.15) was set as 0.1. Preliminary experiments demonstrated that this represents a good trade-off between actual and previous information.

The reconstruction error (5.17) for patients A, B, C and D, is illustrated in Figure 5.13.


Patient A
Error coefficient
0.06
0.05
0.04
0.03
0.02
0.01
0
0
50
100
150
time (h)


Patient B
Error coefficient
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
20
40
60
80
100
120
140
160
180
time (h)


Patient C
Error coefficient
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
20
40
60
80
100
120
140
160
time (h)

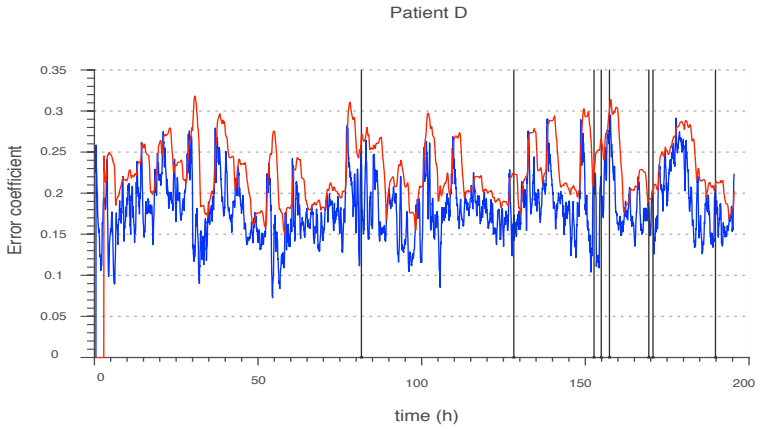


Figure 5.13 – The blue lines represent the time profiles over time while the red lines represent the threshold defined by equation (5.18) for the four patients considered; the black vertical lines represent the seizures onset.

The classification results are summarized in Table 5.2. The best patient in terms of sensitivity was patient B (four of the nine seizures analyzed were correctly predicted). Patient A presented the lower FPR.h^{-1}, however only one of the nine seizure analyzed was correctly predicted.

The time profile of the feature presents a significant variability not always related to pre-ictal periods; the measure rises frequently above the threshold in inter-ictal periods (resulting in a high number of false positives). The low sensitivity results associated to the high FPR.h^{-1} lead us to conclude that the method does not consistently identify variations in the pre-ictal period.

Table 5.2 – Results obtained based on the error reconstruction.

Patient	**Inter-ictal duration (hours)**	**Seiz. Correctly pred. (total number of seiz.) - Alarm sensitivity (%)**	**FPR.h^{-1}**
A	132.47	1(9) - 11%	0.2265
B	145.96	4(9) - 44%	0.2877
C	145.23	1(7) - 14%	0.2548
D	178.45	2(8) - 25%	0. 2746

Different aspects may be considered for a tentative explanation of the poor predictive behavior of this measure. On the one hand, the method may not identify

the variations related to the pre-ictal stage. For example, the variations related to the pre-ictal stage may be gradual, and not detectable by this measure. The features used may contain insufficient information about the variations related to the pre-ictal stage. On the other hand, there are many confounding variables (Mormann et al., 2007) that influence the features extracted from the EEG data. In this case, the variations related to seizure generation may not be observable using this approach.

5.4 Discussion

The characterization of the pre-ictal stage in terms of spatial and spectral properties has some constraints. The two-dimensional representation of data or "flat-world view" might be insufficient to accommodate higher dimensional datasets. A more natural representation of the multidimensional EEG data is based on tensors. In this chapter, we approached the identification and characterization of the pre-ictal state using multi-way analysis.

The first approach based on STF code tensors presented promising results. The analysis of the temporal way revealed that for 32 of the 33 seizures analyzed, at least one component presented significant correlations. Concerning the spatial way, the variability among the patients in terms of number of areas highlighted and proximity to the seizure onset zone is another important observation. The characterization of the pre-ictal stage in terms of temporal, spatial and spectral signatures may represent an important step to narrow the extensive set of combinations feature-channel in seizure prediction studies. For instance, we can use the spatial signatures an appropriate (and reduced) set of channels.

For real-time applications, the analysis of streaming data is a fundamental aspect. We hypothesized that using ITA we could identify variations related to the pre-ictal period.

However, using a tensor stream based on the STF tensor, no consistent pre-ictal changes were detected for most of the seizures analyzed. Moreover, a high $FPR.h^{-1}$ was associated to the low alarm sensitivity. For some seizures, abnormalities were detected in the pre-ictal period, but further studies and extensive testing are necessary in order to conclude about the utility of this technique for seizure prediction.

6 Conclusions

This thesis concerns with the problem of predicting epileptic seizures based on long-term EEG data. The research focuses on two main aspects: (i) development of a seizure prediction algorithm based on machine learning, relying on a high dimensional feature set obtained from multichannel EEG data, (ii) novel approaches to characterize the pre-ictal stage, based on the spatial and temporal behavior of the features extracted from the EEG in order to improve seizure prediction.

6.1 Main Results and Discussion

In Chapter 3, based on a set of 216 epileptic patients from the EPILEPSIAE database, we have investigated an efficient computational framework for epileptic seizure prediction. The classification is made using multiclass SVM and different optimization approaches were investigated to improve the results.

The performance of the method achieved statistical significance in 24 patients (11% of the patients). In fact, the results support the hypothesis that, at least for a small subset of patients, the existence of an identifiable pre-ictal period exists.

Nevertheless, the performance above chance level should be considered as a lower bound for clinical applicability of a seizure prediction algorithm, and 11% represents a very small subset of patients. Future research should focus the improvement of these results, for instance by the development of new features (see Section 6.3).

One of the conclusions is that either scalp or intracranial data may be used for seizure prediction (the results presented by both were very similar). This is rather intriguing due to the different nature of scalp EEG and iEEG. However, these differences were not reflected in the results. On the one hand, the features may not be adequate to retrieve information about the local networks (available in

iEEG) while on the other hand the information about the global behavior may contain additional predictive information.

We have also analyzed different optimization strategies in order to improve the classification results (such as mRMR, RFE, and Kalman filtering). The improvements were minimal and not statistically significant. Nevertheless, the results lead us to conclude that there are specific combinations for each patient, which ultimately confirms the need for individually optimized seizure predictors.

To improve our understanding of the pre-ictal period in terms of the spatio-temporal characteristics and ultimately improve the seizure predictors, we developed two different approaches. The first is based on the spatio-temporal characteristics of the relative spectral power (Chapter 4) and the second is based on the multi-way analysis of STF tensors (Chapter 5).

In Chapter 4, a novel methodology for analyzing the spatio-temporal behavior of the brain electrical activity was proposed. The method is based on the segmentation of topographic mapping of spectral features (and estimation of points of interest), and shows good results for the identification of the four brain states considered – inter-ictal, pre-ictal, ictal, and post-ictal. The results, obtained using a statistical Markov model, suggest that the spatio-temporal dynamics (captured by the trajectories estimated) are related to the epileptic brain states and their transitions.

Chapter 5 focused in a multi-way framework to characterize the pre-ictal period. The method re-arranges multichannel scalp EEG data as a three-way STF tensor with ways representing time samples, frequency bands, and channels. Using PARAFAC model and ALS decomposition algorithm, we modeled the data and identified the rank-one tensor presenting the highest correlation coefficient with the pre-ictal state. We observed that the spatial signature of the components analyzed might be related to the ictal onset zone. Using incremental tensor analysis (ITA) we developed a framework to identify variations in a data stream based on STF tensors. However the results showed poor predictive behavior of the measure proposed.

6.1.1 EPILAB

Some of the main steps of this study were developed using the EPILAB framework (Teixeira et al., 2010; Direito et al., 2011a; Teixeira et al., 2011). EPILAB is a MATLAB® based software package that has been developed to

allow a rapid design and training of several predictors using long-term EEG datasets and to compare their performance. The software includes a variety of feature extraction methods and prediction algorithms with extensive graphical capabilities. It has been conceived as an open architecture where new features, classifiers and general predictors can be easily introduced. This study represents an important contribution to the development of this software.

The development of efficient, and low complexity, seizure predictors plays a major role in the creation of transportable devices for real-time seizure prediction. The development of the "Brainatic", one of the main objectives of the EPILEPSIAE project, is the result of different contributions, including the study presented in this thesis[7].

6.2 Limitations and Shortcomings

Despite of the promising results presented, unavoidable limitations were identified in our research.

One of the limitations encountered is related to the data and variables that could not be controlled. These include light exposure, sleep schedule, activity level and drug administration. The contribution of these confounding variables to the high $FPR.h^{-1}$ obtained is not measurable and should be further explored in future research.

The existence of specific rhythms in the occurrence of seizures (for example circadian dependencies) may have contributed to the low sensitivities obtained. For example, if the training set is based on a specific vigilance state, the model may not be adequate to all vigilance states.

Finally, despite of the characterization of the spatio-temporal dynamics related to the pre-ictal period, we do not have a neurophysiological model to explain these observations, which ultimately would allow a better understanding of mechanisms leading to epileptic seizures.

[7] Teixeira, C.A., Direito, B., Bandarabadi, M., Feldwitch-Drentrup, H., Witton, A., Alvarado-Rojas, C., Le Van Qyen, M., Schelter, B., Favaro, G., Dourado, A., 2012. Brainatic: A system for real-time seizure prediction, was one of the ten nominees for the Annual BCI Research Award 2012.

The results of Chapter 3 show also the limitation of pure data-driven methods of data mining. The consideration of physiological knowledge related to the epileptogenesis phenomena may allow classifiers to increase performance. However this physiological knowledge needs to be further developed.

6.3 Future work

The approach presented in Chapter 3 is based on feature vectors that are formed based on univariate measures. The development of new features able to improve the distinction between the pre-ictal and non pre-ictal states, hence minimizing the effect of confounding variables, is a key aspect of the development of seizure predictors.

Future work will investigate if the integration of novel features improves the performance of the predictors. Bivariate measures, such as phase synchrony, and mean phase coherence, that reflect the complex spatio-temporal connections between different areas of the brain, have been the focus of a number of studies (Garcia Dominguez et al., 2005; Winterhalder et al., 2006; Hegde et al., 2007; Feldwitch-Drentrup et al., 2010; Kuhnert et al., 2010). It is important to evaluate their contribution on the predictors' performance.

Different sources of information, such as ECG, can also be considered to improve the discriminative power of the feature vector. ECG events are known to be associated to epileptic discharges (Novak et al., 1999). Valderrama et al. (2010) suggested that the combination of features derived from EEG and ECG might be useful to detect pre-ictal changes.

Future work should also address the variability associated to the occurrence of seizures. The differences between vigilance states are documented (Quigg, 2000; Schelter et al., 2006b) and can be associated to false predictions. Adaptive classification schemes (Cauwenberghs and Poggio, 2001; Vidaurre et al., 2006) or different thresholds are possibilities that should be further studied.

Apart from epilepsy, we can apply the methods proposed in other research areas where the analysis of the spatio-temporal dynamic of long-term EEG data is important such as the estimation of anesthesia depth.

The continuous development of EPILAB, and the integration of new features and classification capabilities is also an important future contribution.

REFERENCES

Aarabi, A., Fazel-Rezai, R. and Aghakhani, Y. (2009), EEG seizure prediction: Measures and challenges, *in*: 'Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society', pp. 1864–1867.

Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R. and Yener, B. (2007), 'Multiway Analysis of Epilepsy Tensors', *Bioinformatics* 23, 10–18.

Acar, E., Aykut-Bingol, C., Bingol, H. and Yener B. (2006), Computational Analysis of Epileptic Focus Localization, *in*: 'Proceedings of the 4th IASTED International Conference on Biomedical Engineering, pp. 317-322.

Acar, E. and Yener, B. (2009), 'Unsupervised Multiway Data Analysis: A Literature survey', *IEEE Transactions on Knowledge and Data Engineering* 21, 6–20.

Akin, M. (2002), 'Comparison of wavelet transform and FFT methods in the analysis of EEG signals', *Journal of Medical Systems* 26, 241–247.

Alarcon, G., Binnie, C. D., Elwes, R. D. and Polkey, C. E. (1995), 'Power spectrum and intracranial EEG patterns at seizure onset in partial epilepsy', *Electroencephalography and Clinical Neurophysiology* 94, 326–337.

Alba, F.A., Marroquín, J. L., Arce-Santana, E. and Harmony, T. (2010), 'Classification and interactive segmentation of EEG synchrony patterns', *Pattern Recognition* 43, 530–544.

Andersson, C. (2000), 'The N-way Toolbox for MATLAB', *Chemometrics and Intelligent Laboratory Systems* 52, 1–4.

Andrzejak, R. G., Chicharro, D., Elger, C. E. and Mormann, F. (2009), 'Seizure prediction: Any better than chance?', *Clinical Neurophysiology* 120, 1465–1478.

Andrzejak, R. G., Mormann, F., Kreuz, T., Rieke, C., Kraskov, A., Elger, C. E. and Lehnertz, K. (2003), 'Testing the null hypothesis of the nonexistence of a preseizure state', *Physical Review E* 67: 010901.

Bader, B., Kolda, T. (2012), MATLAB Tensor Toolbox Version 2.5.

Barandela, R., Sánchez, J. S., García, V. and Rangel, E. (2003), 'Strategies for learning in class imbalance problems', *Pattern Recognition* 36, 849–851.

Barlow, J. S. (1993) *The Electroencephalogram: Its patterns and Origins,* MIT Press, Massachusetts.

Bazil, C. W., Short, D., Crispin, D. and Zheng, W. (2000), 'Patients with intractable epilepsy have low melatonin, which increases following seizures', *Neurology* 55, 1746–1748.

Bishop, C. M., (2006), *Pattern Recognition and Machine Learning*, Springer.

Blanco, S., Garcia, H., Quiroga, R. Q., Romanelli, L. and Rosso, O. A. (1995), 'Stationarity of the EEG Series', *IEEE Engineering in Medicine and Biology Magazine* 14, 395–399.

Blum, D. E. (1998), 'Computer-based electroencephalography: technical basics, basis for new applications, and potential pitfalls', *Electroencephalography and Clinical Neurophysiology* 106, 118–126.

Box, G. E. P. and Jenkins, G. (1990), *Time Series Analysis, Forecasting and Control*. Holden-Day, Inc.

Burges, C. J. (1998), 'A Tutorial on Support Vector Machines for Pattern Recognition', *Data Mining and Knowledge Discovery* 2, 121–167.

Bro, R. (1997), 'PARAFAC. Tutorial and applications', *Chemometrics and Intelligent Laboratory Systems* 38(2), 149–171.

Byvatov, E., Fechner, U., Sadowski, J. and Schneider, G. (2003), 'Comparison of support vector machine and artificial neural network systems for drug/nondrug classification', *Journal of Chemical Information and Computer Sciences* 43, 1882–1889.

Cao, L. (1997), 'Practical method for determining the minimum embedding dimension of a scalar time series', *Physica D: Nonlinear Phenomena* 110, 43–50.

Carlson, C., Arnedo, V., Cahill, M. and Devinsky, O. (2009), 'Detecting nocturnal convulsions: efficacy of the MP5 monitor', *Seizure* 18(3), 225–7.

Carreño, M., and Lüders, H. O. (2001), General Principles of Presurgical Evaluation, *in* Lüders, H. O., ed., 'Textbook of Epilepsy surgery', Lippincott Williams & Wilkins, Philadelphia.

Cauwenberghs, G. and Poggio, T. (2001), 'Incremental and Decremental Support Vector Machine Learning', *Learning* 13, 409.

Chang, C.C. and Lin, C.-J. (2011), 'LIBSVM: a library for support vector machines', *ACM Transactions on Intelligent Systems and Technology* 2, 27.

Chang, Y. and Lin, C.-J. (2008), Feature Ranking Using Linear SVM, *in*: 'Journal of Machine Learning Research Workshop and Conference Proceedings', pp. 53–64.

Chaovalitwongse, W., Iasemidis, L. D., Pardalos, P. M., Carney, P. R., Shiau, D.-S. and Sackellares, J. C. (2005), 'Performance of a seizure warning algorithm based on the dynamics of intracranial EEG', *Epilepsy Research* 64, 93–113.

Chawla, N. V., Japkowicz, N. and Kolcz, A. (2004), 'Editorial: Special Issue on Learning from Imbalanced Data Sets', *ACM SIGKDD Explorations Newsletter* 6, 1–6.

Chen, Y.-W. and Lin, C.-J. (2006), 'Combining SVMs with Various Feature Selection Strategies', *Studies in Fuzziness and Soft Computing* 207, 315–324.

Chisci, L., Mavino, A., Perferi, G., Sciandrone, M., Anile, C., Colicchio, G. and Fuggetta, F. (2010), 'Real-time epileptic seizure prediction using AR models and support vector machines', *IEEE Transactions on Biomedical Engineering* 57, 1124–1132.

Chiu, A. W. L., Derchansky, M., Cotic, M., Carlen, P. L., Turner, S. O., and Bardakjian, B. L. (2011), 'Wavelet-based Gaussian-mixture hidden Markov model for the detection of multistage seizure dynamics: A proof-of-concept study', *BioMedical Engineering Online* 10.

Cichocki, A., Mørup, M., Smaragdis, P., Wang, W. and Zdunek, R. (2008a), 'Advances in nonnegative matrix and tensor factorization', *Computational Intelligence and Neuroscience* 2008, 852187.

Cichocki, A., Washizawa, Y., Rutkowski, T., Bakardjian, H. and Phan, A.-H. (2008b), 'Noninvasive BCIs: Multiway Signal-Processing Array Decompositions', *Computer* 41, 34–42.

Cooper, G. (2000), *The Cell: A Molecular Approach*, ASM Press.

Cortes, C. and Vapnik, V. (1995), 'Support-vector networks', *Machine Learning* 20, 273–297.

Cour, T., Bénéezit, F. and Shi, J. (2005), Spectral Segmentation with Multiscale Graph Decomposition, *in*: 'Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition', pp. 1124 – 1131.

Cunha, J. P., Paula, L. M., Bento, V. F., Bilgin, C., Dias, E. and Noachtar, S. (2012), 'Movement quantification in epileptic seizures: a feasibility study for a new 3D approach', *Medical engineering & physics* 34(7), 938–45.

Daubechies, I. (1992), *Ten Lectures on Wavelets*, Society for Industrial and Applied Mathematics, Philadelphia.

De Vos, M., De Lathauwer, L., Vanrumste, B., Van Huffel, S. and Van Paesschen, W. (2007a), 'Canonical decomposition of ictal scalp EEG and accurate source localisation: principles and simulation study', *Computational Intelligence and Neuroscience* 2007, 58253.

De Vos, M., Vergult, a, De Lathauwer, L., De Clercq, W., Van Huffel, S., Dupont, P., Palmini, A. and Van Paesschen, W. (2007b), 'Canonical decomposition of ictal scalp EEG reliably detects the seizure onset zone', *NeuroImage* 37, 844–854.

Delamont, R. S., Julu, P. O. and Jamal, G. A. (1999) 'Changes in a measure of cardiac vagal activity before and after epileptic seizures', *Epilepsy Research* 35, 87–94.

Delorme, A. and Makeig, S. (2004), 'EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis', *Journal of Neuroscience Methods* 134, 9–21.

Dhar, S., and Cherkassky, V. (2010), 'Visualization and Interpretation of SVM Classifiers', *Wiley Interdisciplinary Reviews, Data Mining and Knowledge Discovery.*

Ding, C. and Peng, H. (2005), 'Minimum redundancy feature selection from microarray gene expression data', *Journal of Bioinformatics and Computational Biology* 3, 185–205.

Direito, B., Costa, R., Feldwitch-Dentrup, H., Valderrama, M., Nikolopoulos, S., Schelter, B., Jachan, M., Teixeira, C. A., Aires, L., Timmer, J., Le Van Quyen,

M. and Dourado, A. (2011a), EPILAB: A MATLAB platform for multifeature and multialgorithm seizure prediction, *in*: Osorio, I., Zaveri, H. P., Frei, M. G., Arthurs, S., eds., 'Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics', CRC Press, Boca Raton, Florida, pp. 489-500.

Direito, B., Dourado, A., Vieira, M. and Sales, F. (2008), Combining Energy and Wavelet Transform for Epileptic Seizure Prediction in an Advanced Computational System, *in*: 'Proceedings of the International Conference on Biomedical Engineering and Informatics', pp. 380–385.

Direito, B., Ventura, F., Teixeira, C. A. and Dourado, A. (2011b), Optimized feature subsets for epileptic seizure prediction studies, *in*: 'Proceedings of the 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society', pp. 1636–1639.

Drongelen, W., Nayak, S., Frim, D., Kohrman, M., Towle, V., Lee, H., McGee, A., Chico, M. and Hecox, K. (2003), 'Seizure anticipation in pediatric epilepsy: use of kolmogorov entropy', *Pediatric Neurology* 29, 207–213.

Drummond, J. C., Brann, C. A., Perkins, D. E. and Wolfe, D. E. (1991), 'A comparison of median frequency, spectral edge frequency, a frequency band power ratio, total power, and dominance shift in the determination of depth of anesthesia', *Acta Anaesthesiologica Scandinavica* 35(8), 693–699.

Duncan, J. S., Sander, J. W., Sisodiya, S. M. and Walker, M. C. (2006), 'Adult epilepsy'. *Lancet* 367, 1087–1100.

Durazzo, T. S. and Zaveri, H. P. (2011), Seizure prediction and the circadian rhythm, *in*: Osorio, I., Zaveri, H. P., Frei, M. G., Arthurs, S., eds., 'Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics', CRC Press, Boca Raton, Florida, pp. 489-500.

Durbin, R., Eddy, S., Krogh, A. and Mitchison, G. (1998), *Biological Sequence Analysis - Probabilistic models of proteins and nucleic acids,* Cambridge University Press.

Elger, C. E. and Lehnertz, K. (1998), 'Seizure prediction by non-linear time series analysis of brain electrical activity', *European Journal of Neuroscience* 10, 786–789.

Engel, J. (2006), 'ILAE classification of epilepsy syndromes', *Epilepsy Research* 70, 5–10.

Esteller, R., Echauz, J., D'Alessandro, M., Worrell, G., Cranstoun, S., Vachtsevanos, G. and Litt, B. (2005), 'Continuous energy variation during the seizure cycle: towards an on-line accumulated energy', *Clinical Neurophysiology* 116, 517–526.

Federico, P., Abbott, D. F., Briellmann, R. S., Harvey, A. S. and Jackson, G. D. (2005) 'Functional MRI of the pre-ictal state', *Brain* 128, 1811–1817.

Feldwitch-Drentrup, H., Schelter, B., Jachan, M., Nawrath, J., Timmer, J. and Schulze-Bonhage, A. (2010), 'Joining the benefits: combining epileptic seizure prediction methods', *Epilepsia* 51, 1598–1606.

Feldwitch-Drentrup, H., Staniek, M., Schulze-Bonhage, A., Timmer, J., Dickten, H., Elger, C.E., Schelter, B. and Lehnertz, K. (2011), 'Identification of Preseizure States in Epilepsy: A Data-Driven Approach for Multichannel EEG Recordings', *Frontiers in Computational Neuroscience* 5, 1–9.

Fisher, R. S. (2012), 'Therapeutic devices for epilepsy', *Annals of Neurology* 71, 157–168.

Fisher, R. S., Boas, W. V. E., Blume, W., Elger, C. E., Genton, P., Lee, P. and Engel, J. (2005), 'Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)', *Epilepsia* 46, 470–472.

Fisher, R. S. and Engel, J. J. (2010), 'Definition of the postictal state: when does it start and end?' *Epilepsy behavior EB* 19(2), 100–104.

Fisher, R. S., Salanova, V., Witt, T., Worth, R., Henry, T., Gross, R., Oommen, K., *et al.* (2010), 'Electrical stimulation of the anterior nucleus of thalamus for treatment of refractory epilepsy', *Epilepsia*, 51(5), 899–908.

Garcia Dominguez, L., Wennberg, R., Gaetz, W., Cheyne, D., Snead, O. C. and Perez Velazquez, J. L. (2005), 'Enhanced synchrony in epileptiform activity? Local versus distant phase synchronization in generalized seizures', *The Journal of neuroscience* 25, 8077–8084.

Gigola, S., Ortiz, F., D'Attellis, C. E., Silva, W. and Kochen, S. (2004), 'Prediction of epileptic seizures using accumulated energy in a multiresolution framework', *Journal of Neuroscience Methods* 138, 107–111.

Gotman, J. (1982), 'Automatic recognition of epileptic seizures in the EEG', *Electroencephalography and Clinical Neurophysiology* 54(5), 530–540.

Grewal, S. and Gotman, J. (2005), 'An automatic warning system for epileptic seizures recorded on intracerebral EEGs', *Clinical neurophysiology* 116(10), 2460–72.

Guyon, I. and Elisseeff, A. (2003), 'An Introduction to Variable and Feature Selection', *Journal of Machine Learning Research* 3, 1157–1182.

Guyon, I., Weston, J., Barnhill, S. and Vapnik, V. (2002), 'Gene Selection for Cancer Classification using Support Vector Machines', *Machine Learning* 46, 389–422.

Harrison, M. A., Frei, M. G. and Osorio, I. (2005a), 'Accumulated energy revisited', *Clinical Neurophysiology* 116, 527–531.

Harrison, M. A., Osorio, I., Frei, M. G., Asuri, S. and Lai, Y.-C. (2005b), 'Correlation dimension and integral do not predict epileptic seizures', *Chaos* 15, 33106.

Harshman, R. A. (1970), 'Foundations of the PARAFAC procedure: models and conditions for an "explanatory" multimodal factor analysis', *UCLA Working Papers in Phonetics* 16, 1–84.

Hegde, A., Erdogmus, D., Shiau, D.S., Principe, J.C. and Sackellares, J.C. (2007), 'Clustering approach to quantify long-term spatio-temporal interactions in epileptic intracranial electroencephalography', *Computational intelligence and neuroscience* 2007.

Hilborn, R. C. (2000), *Chaos and Nonlinear Dynamics: An Introduction for Scientists and Engineer,* Oxford University Press.

Hitchcock F. L. (1927), 'Multiple invariants and generalized rank of a p-way matrix or tensor', *Journal of Mathematics and Physics* 7, 39-79.

Hjorth, B. (1970), 'EEG analysis based on time domain properties', *Electroencephalography and Clinical Neurophysiology* 29, 306–310.

Hsu, C.W. and Lin, C.-J. (2002), 'A comparison of methods for multi-class support vector machines', *IEEE Transactions on Neural Networks* 13, 415–425.

Iasemidis, L. D., Shiau, D.-S., Pardalos, P. M., Chaovalitwongse, W., Narayanan, K., Prasad, A., Tsakalis, K., Carney, P. R. and Sackellares, J. C.

(2005), 'Long-term prospective on-line real-time seizure prediction', *Clinical Neurophysiology* 116, 532–544.

Iasemidis, L.D., Shiau, D.-S., Sackellares, J.C., Pardalos, P.M. and Prasad, A. (2004), 'Dynamical resetting of the human brain at epileptic seizures: application of nonlinear dynamics and global optimization techniques', *IEEE Transactions on Biomedical Engineering* 51, 493–506.

Iasemidis, L.D., Pardalos, P., Sackellares, J.C. and Shiau, D.-S. (2001), 'Quadratic Binary Programming and Dynamical System Approach to Determine the Predictability of Epileptic Seizures', *Journal of Combinatorial Optimization* 5, 9–26.

Iasemidis, L.D., Principe, J.C. and Sackellares J. C. (1999), Measurement and quantification of spatiotemporal dynamics of human epileptic seizures, *in*: Akay, M. (ed.), 'Nonlinear Biomedical Signal Processing: Dynamic Analysis and Modeling', John Wiley & Sons.

Iasemidis, L. D. and Sackellares, J. C. (1996), 'Chaos Theory and Epilepsy', *Neuroscientist*, 118–126.

Iasemidis, L. D., Sackellares, J. C., Zaveri, H. P. and Williams, W. J., (1990), 'Phase space topography and the Lyapunov exponent of electrocorticograms in partial seizures', *Brain Topography* 2, 187–201.

Ihle, M., Feldwitch-Drentrup, H., Teixeira, C. A., Witon, A., Schelter, B., Timmer, J. and Schulze-Bonhage, A. (2012), 'EPILEPSIAE - A common database for research on seizure prediction', *Computer Methods and Programs in Biomedicine* 106, 127–138.

Jarque, C. M. and Bera, A. K. (1987), 'A Test for Normality of Observations and Regression Residuals', *International Statistical Review* 55, 163–172.

Jeung, H., Shen, H.T. and Zhou, X. (2007), 'Mining Trajectory Patterns Using Hidden Markov Models', *Data Warehousing and Knowledge Discovery*.

Jouny, C. C., Franaszczuk, P. J. and Bergey, G. K. (2005), 'Signal complexity and synchrony of epileptic seizures: is there an identifiable preictal period?', *Clinical Neurophysiology* 116, 552–558.

Kaibara, M. and Blume, W. T. (1988), 'The postictal electroencephalogram', *Electroencephalography and Clinical Neurophysiology* 70(2), 99–104.

Kalman, R. E. (1960), 'A New Approach to Linear Filtering and Prediction Problems', *Transactions of the ASME – Journal of Basic Engineering*, 82, 35–45.

Klatt, J., Feldwitch-Drentrup, H., Ihle, M., Navarro, V., Neufang, M., Teixeira, C. A., Adam, C., Valderrama, M., Alvarado-Rojas, C., Witon, A., Le Van Quyen, M., Sales, F., Dourado, A., Timmer, J., Schulze-Bonhage, A. and Schelter, B. (2012), 'The EPILEPSIAE database - An extensive electroencephalography database of epilepsy patients', *Epilepsia* 53, 1669–1676.

Kolda, T. G. and Bader, B. W. (2009), 'Tensor Decompositions and Applications', *SIAM Review* 51(3), 455–500.

Kruskal, J. B. (1964), 'Multidimensional Scaling by Optimizing Goodness of Fit to a Nonmetric Hypothesis', *Psychometrika* 29(1).

Kruskal, J. B. and Hart, R. E. (1966), 'A geometric interpretation of diagnostic data from a digital machine: based on a study of the Morris, Illinois Electronic Central Office', *Bell System Technical Journal* 45, 1299–1338.

Kuhnert, M.-T., Elger, C. E. and Lehnertz, K. (2010), 'Long-term variability of global statistical properties of epileptic brain networks', *Chaos* 20.

Kwan, P. and Brodie, M. J. (2000), 'Early identification of refractory epilepsy', *The New England Journal of Medicine*, 342, 314–319.

Lai, Y.-C., Harrison, M.A., Frei, M. and Osorio, I. (2003), 'Inability of Lyapunov Exponents to Predict Epileptic Seizures', *Physical Review Letters* 91, 8–11.

Lai, Y.-C., Osorio, I., Harrison, M. A. and Frei, M.G. (2002), 'Correlation-dimension and autocorrelation fluctuations in epileptic seizure dynamics', *Physical Review E* 65, 1–5.

Lange, H. H., Lieb, J. P., Engel, J. and Crandall, P. H. (1983), 'Temporo-spatial patterns of pre-ictal spike activity in human temporal lobe epilepsy', *Electroencephalography and Clinical Neurophysiology* 56(6), 543–555.

Le Van Quyen, M., Adam, C., Martinerie, J., Baulac, M., Clémenceau, S. and Varela, F. (2000), 'Spatio-temporal characterizations of non-linear changes in intracranial activities prior to human temporal lobe seizures', *European Journal of Neuroscience* 12, 2124–2134.

Le Van Quyen, M., Martinerie, J., Baulac, M. and Varela, F. (1999), 'Anticipating epileptic seizures in real time by a non-linear analysis of similarity between EEG recordings', *NeuroReport* 10, 2149–2155.

Le Van Quyen, M., Martinerie, J., Navarro, V., Boon, P., D'Havé, M., Adam, C., Renault, B., Varela, F. and Baulac, M. (2003), 'Anticipation of epileptic seizures from standard EEG recordings', *Lancet* 357, 183–188.

Lee, H. and Choi, S. (2009), Group nonnegative matrix factorization for EEG classification, *in*: 'Proceedings of the 12th International Conference on Artificial Intelligence and Statistics', pp. 320–327.

Lee, H., Kim, Y.-D., Cichocki, A. and Choi, S. (2007), 'Nonnegative tensor factorization for continuous EEG classification', *International Journal of Neural Systems* 17, 305–317.

Lehnertz, K. and Elger, C. (1998), 'Can Epileptic Seizures be Predicted? Evidence from Nonlinear Time Series Analysis of Brain Electrical Activity', *Physical Review Letters* 80, 5019–5022.

Lehnertz, K. and Litt, B. (2005), 'The First International Collaborative Workshop on Seizure Prediction: summary and data description', *Clinical Neurophysiology* 116, 493–505.

Lehnertz, K., Mormann, F., Kreuz, T., Andrzejak, R. G., Rieke, C., David, P. and Elger, C.E. (2003),' Seizure prediction by nonlinear EEG analysis', *IEEE Engineering in Medicine and Biology Magazine* 22, 57–63.

Li, Z., Silva, A. M. and Cunha, J. P. (2002), 'Movement quantification in epileptic seizures: a new approach to video-EEG analysis', *IEEE transactions on Biomedical Engineering* 49, 565-573.

Litt, B., D'Alessandro, M., Esteller, R., Echauz, J. and Vachtsevanos, G. (2003), Translating seizure detection, prediction and brain stimulation into implantable devices for epilepsy, *in*: 'Proceedings of the 1st International IEEE EMBS Conference on Neural Engineering', pp. 485–488.

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

Lopes da Silva, F. (2008), 'The Impact of EEG/MEG Signal Processing and Modeling in the Diagnostic and Management of Epilepsy', *IEEE Reviews in Biomedical Engineering* 1, 143–156.

Lopes da Silva, F., Blanes, W., Kalitzin, S. N., Parra, J., Suffczynski, P. and Velis, D. N. (2003), 'Epilepsies as dynamical diseases of brain systems: basic models of the transition between normal and epileptic activity', *Epilepsia* 44, 72–83.

Lüders, H. O., Najm, I., Nair, D., Widdess-Walsh, P. and Bingman, W. (2006), 'The epileptogenic zone: general principles', *Epileptic disorders*, 8(2), S1–9.

Maiwald, T., Winterhalder, M., Aschenbrenner-Scheibe, R., Voss, H. U., Schulze-Bonhage, A. and Timmer, J. (2004), 'Comparison of three nonlinear seizure prediction methods by means of the seizure prediction characteristic', *Physica D* 194, 357–368.

Malow, B. (2004), 'Sleep deprivation and epilepsy', *Epilepsy Currents* 4(5), 193–5.

Mammone, N., Morabito, F. C. and Principe, J. C. (2006), 'Visualization of the short term maximum Lyapunov exponent topography in the epileptic brain', *in*: 'Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society', pp. 4257–4260.

Mammone, N., Principe, J. C., Morabito, F. C., Shiau, D.-S. and Sackellares, J. C. (2010), 'Visualization and modelling of STLmax topographic brain activity maps', *Journal of Neuroscience Methods* 189, 281–294.

Marques de Sá, J. P. (2001), *Pattern Recognition. Concepts, Methods and Applications, Methods*, Springer, Berlin.

Martinerie, J., Adam, C., Le Van Quyen, M., Baulac, M., Clemenceau, S., Renault, B. and Varela, F. (1998), 'Epileptic seizures can be anticipated by non-linear analysis', *Nature Medicine* 4, 1173–1176.

Massey, F. J. (1951), 'The Kolmogorov-Smirnov test for goodness of fit', *Journal of the American Statistical Association* 46(253), 68–78.

Meyer, D., Leisch, F. and Hornik, K. (2003), 'The support vector machine under test' *Neurocomputing* 55(1-2), 169–186.

Mirowski, P., Madhavan, D., Lecun, Y. and Kuzniecky, R. (2009), 'Classification of patterns of EEG synchronization for seizure prediction', *Clinical Neurophysiology* 120, 1927–1940.

Miwakeichi, F., Martínez-Montes, E., Valdés-Sosa, P., Nishiyama, N., Mizuhara, H. and Yamaguchi, Y. (2004), 'Decomposing EEG data into space-time-frequency components using Parallel Factor Analysis', *NeuroImage* 22, 1035–1045.

Montgomery, D. C., and Runger, G. C. (2003). *Applied Statistics and Probability for Engineers*, John Wiley & Sons, Inc., New York.

Morrell, M. (2006), 'Brain stimulation for epilepsy: can scheduled or responsive neurostimulation stop seizures?' *Current Opinion in Neurology*, 19(2), 164–168.

Mormann, F. (2000), 'Mean phase coherence as a measure for phase synchronization and its application to the EEG of epilepsy patients', *Physica D: Nonlinear Phenomena* 144, 358–369.

Mormann, F., Andrzejak, R. G., Elger, C. E. and Lehnertz, K. (2007), 'Seizure prediction: the long and winding road', *Brain* 130, 314–333.

Mormann, F., Kreuz, T., Rieke, C., Andrzejak, R. G., Kraskov, A., David, P., Elger, C.E. and Lehnertz, K. (2005), 'On the predictability of epileptic seizures', *Clinical Neurophysiology* 116, 569–587.

Murray, J. (1993), 'Coping with uncertainty of uncontrolled epilepsy', *Seizure* 2, 168–178.

Netoff, T., Park, Y. and Parhi, K.K. (2009), Seizure prediction using cost-sensitive support vector machine, *in*: 'Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society', pp. 3322–3325.

Ngugi, A. K., Kariuki, S. M., Bottomley, C., Kleinschmidt, I., Sander, J. W. and Newton, C. R., (2011), 'Incidence of epilepsy: a systematic review and meta-analysis', *Neurology* 77, 1005–1012.

Novak, V., Reeves, L., Novak, P., Low, A. and Sharbrough, W. (1999), 'Time-frequency mapping of R-R interval during complex partial seizures of temporal lobe origin', *Journal of the Autonomic Nervous System* 77, 195–202.

Nunez, P. L. and Srinivasan, R. (2006), *Electric Fields of the Brain: The Neurophysics of EEG*, Oxford University Press.

Osorio, I., Frei, M. G. and Wilkinson, S. B. (1998), 'Real-Time Automated Detection and Quantitative Analysis of Seizures and Short-Term Prediction of Clinical Onset', *Epilepsia* 39, 615–627.

Ouyang, G., Li, X., Li, Y. and Guan, X. (2007), 'Application of wavelet-based similarity analysis to epileptic seizures prediction', *Computers in Biology and Medicine* 37, 430–437.

Pardalos, P. M., Iasemidis, L. D., Sackellares, J. C., Shiau, D.-S., Carney, P. R., Prokopyev, O. A. and Yatsenko, V. A. (2004), 'Seizure warning algorithm based on optimization and nonlinear dynamics', *Mathematical Programming 101*, 365–385.

Pardey, J., Roberts, S. and Tarassenko, L. (1996), 'A review of parametric EEG analysis modelling techniques for EEG analysis', *Medical Engineering & Physics* 18(1), 2–11.

Park, Y., Luo, L., Parhi, K.K. and Netoff, T. (2011), Seizure prediction with spectral power of EEG using cost-sensitive support vector machines, *Epilepsia* 52, 1761–1770.

Pearson, K. (1901), 'On lines and planes of closest fit to systems of points in space', *Philosophical Magazine Series* 6, 2(11), 559–572.

Peng, H., Long, F. and Ding, C. (2005), 'Feature Selection Based on Mutual Information : Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy', *IEEE transactions on pattern analysis and machine intelligence* 27(8), 1226–1238.

Pikovsky, A., Rosenblum, M., and Kurths, J. (2001), S*ynchronization: A Universal Concept in Nonlinear Sciences*, Cambridge University Press, Cambridge.

Qiao, X. and Liu, Y. (2009), 'Adaptive weighted learning for unbalanced multicategory classification', *Biometrics* 65, 159–168.

Quigg, M. (2000), 'Circadian rhythms: interactions with seizures and epilepsy', *Epilepsy research* 42(1), 43–55.

Raghunathan, S., Gupta, S. K., Markandeya, H. S., Irazoqui, P. P. and Roy, K. (2011), 'Ultra Low-Power Algorithm Design for Implantable Devices: Application to Epilepsy Prostheses', *Journal of Low Power Electronics and Applications* 1(3), 175–203.

Rajdev, P., Ward, M. P., Rickus, J., Worth, R. and Irazoqui, P. P., (2010), 'Real-time seizure prediction from local field potentials using an adaptive Wiener algorithm', *Computers in Biology and Medicine* 40, 97–108.

Rajna, P., Clemens, B., Csibri, E., Dobos, E., Geregely, A., Gottschal, M., György, I., Horváth, A., Horváth, F., Mezöfi, L., Velkey, I., Veres, J. and Wagner, E. (1997), 'Hungarian multicentre epidemiologic study of the warning and initial symptoms (prodrome, aura) of epileptic seizures', *Seizure: European Journal of Epilepsy* 6, 361–368.

Roerdink, J. and Meijster, A. (2001), 'The Watershed Transform: Definitions, Algorithms and Parallelization Strategies', *Fundamenta Informaticae* 41, 1–40.

Rogowski, Z., Gath, I. and Bental, E. (1981), 'On the Prediction of Epileptic Seizures', *Biological Cybernetics* 15, 9–15.

Rothman, S. M., Smyth, M. D., Yang, X.-F. and Peterson, G. P. (2005), 'Focal cooling for epilepsy: an alternative therapy that might actually work', *Epilepsy & behavior* 7(2), 214–21.

Sackellares, J.C. (2008), 'Seizure prediction', *Epilepsy Currents* 8, 55–59.

Sandini, G., Romano, P., Scotto, A. and Traverso, G. (1983), 'Topography of brain electrical activity: a bioengineering approach', *Medical Progress Through Technology* 10, 5–18.

Schelter, B., Winterhalder, M., Maiwald, T., Brandt, A., Schad, A., Schulze-Bonhage, A. and Timmer, J. (2006a), 'Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction', *Chaos* 16, 013108.

Schelter, B., Winterhalder, M., Maiwald, T., Brandt, A., Schad, A., Timmer, J. and Schulze-Bonhage, A. (2006b), 'Do false predictions of seizures depend on the state of vigilance? A report from two seizure-prediction methods and proposed remedies', *Epilepsia* 47, 2058–2070.

Schomer, D. L. and Lopes da Silva, F. (2010), *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, Lippincott Williams & Wilkins, Philadelphia.

Schulze-Bonhage, A., Feldwitch-Drentrup, H. and Ihle, M. (2011), 'The role of high-quality EEG databases in the improvement and assessment of seizure prediction methods', *Epilepsy & Behavior* 22, S88–S93.

Schulze-Bonhage, A., Sales, F., Wagner, K., Teotonio, R., Carius, A., Schelle, A. and Ihle, M. (2010), 'Views of patients with epilepsy on seizure prediction devices', *Epilepsy & Behavior* 18, 388–396.

Semmlow, J. L. (2004), *Biosignal and Biomedical Image Processing*, CRC Press.

Shahwan, A., Bailey, C., Maxiner, W. and Harvey, A. S. (2009), 'Vagus nerve stimulation for refractory epilepsy in children: More to VNS than seizure frequency reduction', *Epilepsia* 50(5), 1220–1228.

Shi, J. and Malik, J. (2000), 'Normalized cuts and image segmentation', *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 888–905.

Shoeb, A. (2009), 'Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment', Massachusetts Institute of Technology.

Shoeb, A., Pang, T., Guttag, J. and Schachter, S. (2011), 'Vagus Nerve Stimulation Triggered by Machine-Learning Based Seizure Detection: Initial Implementation and Evaluation', *in*: Osorio, I., Zaveri, H. P., Frei, M. G., Arthurs, S., eds., 'Epilepsy: The Intersection of Neurosciences, Biology, Mathematics, Engineering, and Physics', CRC Press, Boca Raton, Florida, pp. 385–396.

Shoeb, A., Pang, T., Guttag, J., and Schachter, S. (2009), 'Non-invasive computerized system for automatically initiating vagus nerve stimulation following patient-specific detection of seizures or epileptiform discharges', *International journal of neural systems*, 19(3), 157–72.

Singla, R. (2011), 'Comparison of SVM and ANN for classification of eye events in EEG', *Journal of Biomedical Science and Engineering* 4, 62–69.

Smart, O., Firpi, H. and Vachtsevanos, G. (2007), 'Genetic Programming of Conventional Features to Detect Seizure Precursors', *Engineering Applications of Artificial Intelligence* 20, 1070–1085.

Snyder, D. E., Echauz, J., Grimes, D. B. and Litt, B. (2008), 'The statistics of a practical seizure warning system', *Journal of Neural Engineering* 5, 392–401.

Song, S., Zhan, Z., Long, Z., Zhang, J. and Yao, L. (2011), 'Comparative study of SVM methods combined with voxel selection for object category classification on fMRI data', *PloS one* 6.

Stacey, W., Le Van Quyen, M., Mormann, F. and Schulze-Bonhage, A. (2011), 'What is the present-day EEG evidence for a preictal state?' *Epilepsy Research* 97, 243–251.

Stacey, W. and Litt, B. (2008), 'Technology Insight: neuroengineering and epilepsy—designing devices for seizure control', *Nature Clinical Practice Neurology* 4(4), 190–201.

Suffczynski, P., Kalitzin, S.N. and Lopes da Silva, F. (2004), 'Dynamics of non-convulsive epileptic phenomena modeled by a bistable neuronal network', *Neuroscience* 126, 467–484.

Sun, J., Tao, D. and Faloutsos, C. (2006), 'Beyond Streams and Graphs: Dynamic Tensor Analysis', *in*: 'Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining'.

Sun, J., Tao, D., Papadimitriou, S., Yu, P. S. and Faloutsos, C. (2008), 'Incremental tensor analysis', *ACM Transactions on Knowledge Discovery from Data* 2, 1–37.

Świderski, B., Osowski, S. and Cichocki, A. (2007), Epileptic Seizure Prediction Using Lyapunov Exponents and Support Vector Machine, *in*: 'Proceedings of the 8th International Conference on Adaptive and Natural Computing Algorithms', pp. 373 – 381.

Świderski, B., Osowski, S. and Rysz, A. (2005), Lyapunov exponent of EEG signal for epileptic seizure characterization, *in*: 'Proceedings of the European Conference on Circuit Theory and Design', pp. 153–156.

Takens, F. (1981), Detecting strange attractors in turbulence, in: Rand, D., Young, L. eds., 'Dynamical Systems and Turbulence', Springer, New York.

Tate, R. F. (1954), 'Correlation Between a Discrete and a Continuous Variable. Point-Biserial Correlation', *The Annals of Mathematical Statistics* 25(3), 603–607.

Teixeira, C. A., Direito, B., Costa, R. P., Valderrama, M., Feldwitch-Drentrup, H., Nikolopoulos, S., Le Van Quyen, M., Schelter, B. and Dourado, A. (2010), A computational Environment for Long-Term Multi-Feature and Multi-

Algorithm Seizure Prediction, *in*: 'Proceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society'.

Teixeira, C. A., Direito, B., Feldwitch-Drentrup, H., Valderrama, M., Costa, R. P., Alvarado-Rojas, C., Nikolopoulos, S., Le Van Quyen, M., Timmer, J., Schelter, B. and Dourado, A. (2011), 'EPILAB: A software package for studies on the prediction of epileptic seizures', *Journal of Neuroscience Methods* 200, 257–271.

ten Caat, M., Maurits, N.M. and Roerdink, J. (2008), 'Data-driven visualization and group analysis of multichannel EEG coherence with functional units', *IEEE Transactions on Visualization and Computer Graphics* 14, 756–771.

Temko, A., Thomas, E., Marnane, W., Lightbody, G. and Boylan, G. B. (2011), 'Performance assessment for EEG-based neonatal seizure detectors', *Clinical Neurophysiology* 122, 474–482.

Trosset, M. and Priebe, C. (2008), 'The out-of-sample problem for classical multidimensional scaling', *Computational Statistics & Data Analysis* (10), 4635–4642.

Tucker, L. R. (1966), 'Some mathematical notes on three-mode factor analysis', *Psychometrika* 31(3), 279–311.

Ulrich, T. J. (1972), 'Maximum entropy power spectrum of truncated sinusoids', *Journal of Geophysical Research* 77(8), 1396–1400.

Valderrama, M., Nikolopoulos, S., Adam, C., Navarro, V. and Le Van Quyen, M. (2010), Patient-specific seizure prediction using a multi-feature and multi-modal EEG-ECG classification, *in*: 'Proceedings of the 12th Mediterranean Conference on Medical and Biological Engineering and Computing'.

Vapnik, V. (1999), 'An overview of statistical learning theory', *IEEE Transactions on Neural Networks* 10(5), 988–99.

Vapnik, V. and Chervonenkis, A. (1974), 'Ordered risk minimization', *Automation and Remote Control* 34, 1226–1235.

Vidaurre, C., Schlögl, A., Cabeza, R., Scherer, R. and Pfurtscheller, G. (2006), 'A fully on-line adaptive BCI', *IEEE Transactions on Biomedical Engineering* 53, 1214–1219

Viglione, S. and Walsch, G. (1975), 'Epileptic seizure prediction', *Electroencephalography and Clinical Neurophysiology* 39, 435–436.

Welch, G. and Bishop, G. (2006), 'An Introduction to the Kalman Filter', University of North Carolina.

Wendling, F., Hernandez, A., Bellanger, J.-J., Chauvel, P. and Bartolomei, F. (2005), 'Interictal to ictal transition in human temporal lobe epilepsy: insights from a computational model of intracerebral EEG', *Journal of clinical neurophysiology* 22(5), 343–56.

Wilcoxon, F. (1945), 'Individual Comparisons by Ranking Methods', *Biometrics Bulletin* 1, 80–83.

Winterhalder, M., Maiwald, T., Voss, H. U., Aschenbrenner-Scheibe, R., Timmer, J. and Schulze-Bonhage, A. (2003), 'The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods', *Epilepsy & Behavior* 4, 318–325.

Winterhalder, M., Schelter, B., Maiwald, T., Brandt, A., Schad, A., Schulze-Bonhage, A. and Timmer, J. (2006), 'Spatio-temporal patient-individual assessment of synchronization changes for epileptic seizure prediction', *Clinical Neurophysiology* 117, 2399–2413.

Wolf, A., Swift, J. B., Swinney, H. L. and Vastano, J. A. (1985), 'Determining Lyapunov exponents from a time series', *Physica*, 285–317.

Wong, S., Gardner, A. B., Krieger, A. M. and Litt, B. (2007), 'A stochastic framework for evaluating seizure predictions algorithms using Hidden Markov Models', *Journal of Neurophysiology* 97, 2525–2532.

Appendix A

Summary of the results obtained by the best models in the testing set data for scalp EEG recordings. The statistically significant cases are highlighted in the last column.

Pat. ID is the patient identification; *Loc.* is the localization of the epileptic focus (c. = central, f = frontal, o = occipital, p = parietal, t = temporal, and h = hemisphere); *Lat.* is the lateralization (b = bilateral, l = left, and r = right), *Elect. array* is the electrode array of the best model (*random* = random selection, *focal* = selection based on seizure onset zone and propagation, *10-20* = selection based on the discretization of the patient scalp); *SOP* is the seizure occurrence period (or pre-ictal time); *#Seiz.* is the number of seizures in the testing data; *Train F-mea.* is the *F-measure* determined in the training set; *Test F-mea.* is the *F-measure* determined in the test set; *Sens. (%)* is the ratio of predicted seizures over all seizures in the training set; $FPR.h^{-1}$ is the false positive rate raised by the predictor; *Crit. S. (%)* is the critical sensitivity of the random predictor; and *p-value* is the *p*-value of the statistical test.

Pat. ID	**Loc.**	**Lat.**	**Elect. array**	**SOP**	**# seiz.**	**Train *F-mea.***	**Test *F-mea.***	**Sens. (%)**	**FPR.h^{-1}**	**Crit. S. (%)**	***p*-value**
102	c	l	*10-20*	40	8	0,74	0,07	25,00	0,09	37,50	0,62
202	t	r	*random*	30	5	0,76	0,12	60,00	0,22	60,00	0,11
300	f	b	*random*	10	19	0,78	0,19	20,00	0,25	21,05	0,08
402	t	l	*random*	20	2	0,72	0,08	50,00	0,06	50,00	0,38
500	t	r	*10-20*	40	4	0,74	0,11	25,00	0,12	50,00	0,98
600	t	b	*random*	30	8	0,58	0,09	25,00	0,21	50,00	0,91
700	t	l	*focal*	40	9	0,66	0,11	22,22	0,31	55,56	1,00
800	t	l	*random*	20	4	0,68	0,03	25,00	0,15	50,00	0,90
1200	t	r	*10-20*	10	21	0,73	0,11	18,18	0,25	19,05	0,10
1300	-	-	*focal*	20	3	0,56	0,05	33,33	0,00	0,00	0,00
1600	t	l	*10-20*	20	4	0,70	0,04	25,00	0,23	50,00	0,97
2000	t	r	*random*	30	3	0,53	0,05	100,00	0,46	100,00	0,10

2200	f	r	*10-20*	20	14	0,45	0,10	14,29	0,00	0,00	0,00
2300	t	l	*random*	40	7	0,66	0,09	57,14	0,53	71,43	0,79
2600	t	r	*10-20*	40	2	0,67	0,07	0,00	0,00	0,00	1,00
2800	f	b	*focal*	30	1	0,63	0,03	100,00	0,16	100,00	0,61
2900	-	-	*random*	30	7	0,65	0,09	28,57	0,05	28,57	0,16
3300	-	-	*10-20*	40	6	0,53	0,16	33,33	0,20	50,00	0,89
3500	t	r	*random*	40	2	0,68	0,12	100,00	0,14	100,00	0,09
3600	-	-	*focal*	40	24	0,59	0,34	24,00	0,15	29,17	0,25
3700	p	l	*random*	10	2	0,57	0,01	50,00	0,59	100,00	0,90
4000	t	l	*10-20*	20	3	0,76	0,06	66,67	0,27	66,67	0,22
4200	-	-	*random*	10	1	0,46	0,02	100,00	0,23	100,00	0,37
4400	t	r	*10-20*	20	2	0,74	0,07	50,00	0,12	50,00	0,62
4500	t	l	*10-20*	30	1	0,56	0,02	0,00	0,02	100,00	1,00
4700	t	r	*random*	40	1	0,58	0,05	100,00	0,23	100,00	0,84
4900	t	l	*10-20*	40	2	0,70	0,12	100,00	0,25	100,00	0,25
5100	t	r	*focal*	40	4	0,63	0,16	100,00	0,61	100,00	0,14
5200	t	b	*focal*	30	3	0,63	0,05	33,33	0,03	33,33	0,44
5500	t	b	*random*	40	1	0,55	0,01	0,00	0,00	0,00	1,00
5800	t	b	*10-20*	10	6	0,65	0,05	33,33	0,56	50,00	0,69
5900	t	r	*focal*	10	4	0,36	0,08	25,00	0,01	25,00	0,10
6000	t	l	*focal*	30	4	0,56	0,08	100,00	0,34	75,00	0,01
6100	t	b	*focal*	30	16	0,62	0,32	37,50	0,23	37,50	0,06
6200	t	l	*focal*	30	8	0,59	0,12	50,00	0,17	37,50	0,03
6300	t	b	*10-20*	30	3	0,58	0,03	66,67	0,47	100,00	0,77
6500	-	-	*10-20*	40	13	0,63	0,13	7,14	0,06	23,08	1,00
6600	t	r	*random*	20	2	0,70	0,05	50,00	0,25	100,00	0,86
6700	t	r	*random*	30	3	0,64	0,08	66,67	0,27	66,67	0,41
6800	t	l	*10-20*	20	2	0,74	0,07	100,00	0,58	100,00	0,31
6900	-	-	*random*	30	2	0,66	0,03	100,00	0,45	100,00	0,40
7000	t	b	*10-20*	40	2	0,80	0,00	0,00	0,13	100,00	1,00
7200	t	b	*focal*	20	1	0,75	0,04	100,00	0,07	100,00	0,24
7300	-	-	*focal*	30	8	0,55	0,05	25,00	0,21	50,00	0,91
7500	-	-	*10-20*	10	1	0,67	0,02	100,00	0,31	100,00	0,47
7700	-	-	*10-20*	40	4	0,62	0,06	50,00	0,51	100,00	0,99
7800	-	-	*10-20*	10	3	0,54	0,03	33,33	0,26	66,67	0,79
8100	-	-	*random*	30	3	0,65	0,03	33,33	0,16	66,67	0,95
11002	t	r	*10-20*	40	5	0,51	0,08	20,00	0,13	60,00	0,99
16202	-	-	*random*	30	5	0,78	0,05	20,00	0,12	40,00	0,98
19202	-	-	*10-20*	40	9	0,58	0,03	9,09	0,03	22,22	0,90
21602	f	l	*focal*	30	8	0,63	0,08	50,00	0,38	62,50	0,36

21902	t	l	*random*	30	3	0,65	0,07	33,33	0,00	0,00	0,00
22602	f	l	*random*	40	8	0,66	0,14	37,50	0,30	62,50	0,89
23902	t	l	*focal*	20	2	0,70	0,02	0,00	0,07	50,00	1,00
26102	t	l	*10-20*	30	5	0,75	0,05	40,00	0,12	40,00	0,30
30802	t	r	*focal*	40	6	0,83	0,15	66,67	0,11	50,00	0,00
32202	o	l	*10-20*	10	9	0,56	0,10	11,11	0,11	22,22	0,85
32502	f	r	*focal*	30	4	0,73	0,11	50,00	0,28	75,00	0,67
32702	t	l	*random*	30	3	0,62	0,07	33,33	0,12	66,67	0,89
46702	t	r	*random*	40	2	0,62	0,05	50,00	0,13	100,00	0,88
50802	t	l	*10-20*	30	2	0,77	0,05	0,00	0,15	100,00	1,00
53402	t	r	*focal*	30	5	0,59	0,07	40,00	0,10	40,00	0,21
55202	t	r	*random*	30	6	0,62	0,03	16,67	0,07	33,33	0,91
56402	t	l	*random*	40	1	0,70	0,04	0,00	0,02	100,00	1,00
58602	t	l	*random*	20	19	0,74	0,17	31,58	0,19	26,32	0,01
60002	t	l	*10-20*	20	5	0,66	0,02	20,00	0,23	40,00	0,99
70202	t	l	*focal*	30	3	0,72	0,12	33,33	0,11	66,67	0,85
71002	t	r	*10-20*	10	7	0,00	0,00	0,00	0,00	0,00	1,00
71102	t	r	*10-20*	30	5	0,77	0,19	20,00	0,08	40,00	0,90
71202	t	l	*10-20*	30	5	0,77	0,19	20,00	0,08	40,00	0,90
71402	-	-	*10-20*	10	2	0,45	0,00	0,00	0,02	50,00	1,00
71502	t	r	*10-20*	40	2	0,59	0,02	0,00	0,06	50,00	1,00
71602	t	l	*focal*	40	11	0,62	0,07	0,00	0,05	27,27	1,00
71702	-	-	*random*	40	6	0,56	0,02	0,00	0,09	33,33	1,00
72302	t	l	*focal*	40	8	0,88	0,12	25,00	0,06	25,00	0,40
72402	-	-	*focal*	40	14	0,43	0,12	7,69	0,13	28,57	1,00
72702	-	-	*random*	40	8	0,67	0,13	0,00	0,05	25,00	1,00
72802	-	-	*random*	30	5	0,78	0,20	50,00	0,13	40,00	0,03
73102	c	l	*focal*	20	1	0,91	0,01	0,00	0,10	100,00	1,00
75202	t	r	*10-20*	10	5	0,77	0,03	40,00	0,31	40,00	0,24
79502	p	l	*10-20*	10	6	0,50	0,01	16,67	0,28	33,33	0,97
80602	f	l	*random*	20	5	0,61	0,04	40,00	0,10	40,00	0,11
81102	t	r	*random*	20	10	0,67	0,05	11,11	0,07	20,00	0,94
81402	f	l	*random*	30	12	0,74	0,04	8,33	0,10	25,00	1,00
85202	t	l	*focal*	40	7	0,80	0,16	14,29	0,04	28,57	0,89
92102	p	r	*10-20*	40	3	0,64	0,05	0,00	0,07	66,67	1,00
93402	t	l	*10-20*	10	4	0,83	0,03	25,00	0,27	50,00	0,89
93902	t	r	*random*	30	6	0,69	0,01	33,33	0,20	50,00	0,74
95202	t	l	*random*	30	9	0,78	0,10	44,44	0,28	44,44	0,24
96002	t	r	*10-20*	40	5	0,77	0,05	40,00	0,16	60,00	0,64
100002	-	-	*random*	20	18	0,49	0,11	17,65	0,07	16,67	0,09

101702	t	l	*focal*	40	3	0,66	0,07	33,33	0,06	66,67	0,77
102202	t	l	*10-20*	10	25	0,63	0,04	16,00	0,19	16,00	0,09
103002	o	r	*random*	10	5	0,45	0,03	40,00	0,17	40,00	0,09
103802	-	-	*10-20*	10	7	0,56	0,07	14,29	0,27	28,57	0,98
104602	t	l	*random*	40	1	0,62	0,03	0,00	0,16	100,00	1,00
109202	f	l	*focal*	30	2	0,59	0,01	50,00	0,23	100,00	0,93
109502	t	r	*random*	10	7	0,66	0,10	14,29	0,01	14,29	0,15
110602	t	r	*random*	10	5	0,53	0,04	20,00	0,14	40,00	0,74
111902	c	l	*10-20*	30	5	0,69	0,01	20,00	0,09	40,00	0,93
112402	f	r	*focal*	40	11	0,72	0,19	27,27	0,13	36,36	0,50
112802	t	l	*focal*	20	3	0,52	0,02	33,33	0,11	33,33	0,74
113902	t	r	*random*	20	19	0,53	0,15	31,58	0,44	36,84	0,34
114702	t	r	*10-20*	20	22	0,68	0,12	9,09	0,08	13,64	0,76
115102	f	r	*focal*	30	12	0,47	0,07	8,33	0,09	25,00	1,00
115202	o	r	*random*	30	6	0,66	0,05	33,33	0,13	50,00	0,47
123902	t	l	*random*	40	5	0,73	0,20	50,00	0,03	20,00	0,00
136402	-	-	*random*	30	3	0,66	0,07	33,33	0,02	33,33	0,36
1233803	-	-	*focal*	10	1	0,57	0,00	100,00	0,58	100,00	0,69
1234303	t	l	*random*	20	4	0,78	0,06	66,67	0,23	50,00	0,02
1235003	t	l	*random*	40	2	0,80	0,05	0,00	0,08	50,00	1,00
1299403	t	l	*10-20*	10	1	0,78	0,00	100,00	0,46	100,00	0,60
1300003	t	r	*10-20*	30	3	0,72	0,01	0,00	0,03	33,33	1,00
1305803	h	l	*10-20*	40	1	0,71	0,01	0,00	0,14	100,00	1,00
1306003	t	r	*random*	40	4	0,68	0,06	75,00	0,09	50,00	0,01
1306203	-	-	*10-20*	10	11	0,64	0,03	27,27	0,24	27,27	0,09
1306903	t	l	*random*	30	8	0,54	0,02	12,50	0,05	25,00	0,89
1307003	t	r	*10-20*	40	1	0,82	0,04	0,00	0,03	100,00	1,00
1307103	t	l	*10-20*	10	7	0,57	0,01	28,57	0,20	28,57	0,21
1307403	t	l	*focal*	20	3	0,66	0,05	33,33	0,07	33,33	0,55
1307503	h	r	*random*	20	14	0,57	0,11	14,29	0,19	28,57	0,95
1307803	-	-	*10-20*	10	1	0,95	0,00	0,00	0,06	100,00	1,00
1308403	t	b	*10-20*	40	1	0,72	0,02	100,00	0,17	100,00	0,74
1308503	t	l	*10-20*	40	2	0,71	0,04	0,00	0,53	100,00	1,00
1308603	t	r	*random*	10	1	0,56	0,01	100,00	0,27	100,00	0,42
1309803	t	l	*10-20*	30	2	0,66	0,01	50,00	0,27	100,00	0,96
1310803	t	l	*focal*	30	1	0,63	0,02	0,00	0,19	100,00	1,00
1311003	t	l	*random*	40	1	0,66	0,01	100,00	0,17	100,00	0,74
1312603	h	r	*focal*	30	1	0,78	0,07	0,00	0,00	0,00	1,00
1312703	t	r	*focal*	20	3	0,76	0,14	33,33	0,16	66,67	0,85
1312803	t	l	*10-20*	40	1	0,83	0,00	0,00	0,01	100,00	1,00

1312903	f	l	*10-20*	10	2	0,72	0,01	50,00	0,20	50,00	0,55
1313003	t	l	*focal*	10	5	0,49	0,00	20,00	0,37	40,00	0,97
1313403	t	l	*focal*	30	2	0,70	0,01	0,00	0,41	100,00	1,00
1313903	t	l	*10-20*	40	1	0,66	0,01	100,00	0,06	100,00	0,40
1314103	t	l	*focal*	10	3	0,52	0,02	33,33	0,33	66,67	0,86
1314703	t	l	*random*	10	3	0,87	0,02	66,67	0,22	33,33	0,04
1314803	f	l	*focal*	40	5	0,72	0,05	40,00	0,07	40,00	0,22
1315003	t	r	*random*	40	2	0,78	0,13	0,00	0,29	100,00	1,00
1315203	f	r	*focal*	40	1	0,60	0,01	0,00	0,14	100,00	1,00
1315403	-	-	*random*	20	3	0,77	0,05	66,67	0,41	66,67	0,42
1316003	f	b	*random*	10	1	0,66	0,00	100,00	0,19	100,00	0,32
1316303	t	l	*random*	40	1	0,77	0,01	0,00	0,13	100,00	1,00
1316403	h	b	*10-20*	40	8	0,81	0,10	0,00	0,05	25,00	1,00
1316503	t	r	*focal*	30	4	0,71	0,13	25,00	0,10	50,00	0,90
1317003	t	l	*10-20*	40	2	0,79	0,02	0,00	0,14	100,00	1,00
1317203	t	l	*10-20*	20	2	0,73	0,02	100,00	0,68	100,00	0,40
1317303	t	r	*random*	30	2	0,57	0,10	50,00	0,38	100,00	0,99
1317403	-	-	*10-20*	40	1	0,67	0,02	100,00	0,31	100,00	0,92
1317903	t	l	*10-20*	20	2	0,67	0,00	0,00	0,20	50,00	1,00
1318803	t	r	*focal*	10	3	0,72	0,04	66,67	0,85	66,67	0,44
1319103	t	r	*random*	40	1	0,79	0,07	0,00	0,00	0,00	1,00
1319203	-	r	*10-20*	40	5	0,56	0,03	40,00	0,27	60,00	0,92
1320303	t	r	*focal*	30	1	0,81	0,00	0,00	0,03	100,00	1,00
1320503	t	l	*focal*	30	1	0,66	0,05	100,00	0,13	100,00	0,53
1320903	t	l	*focal*	30	2	0,65	0,04	0,00	0,16	100,00	1,00
1321003	t	r	*10-20*	40	1	0,72	0,04	0,00	0,00	0,00	1,00
1321103	t	l	*focal*	20	1	0,78	0,02	100,00	0,35	100,00	0,75
1321803	t	r	*10-20*	20	4	0,75	0,09	50,00	0,62	75,00	0,87
1321903	t	l	*random*	20	8	0,58	0,21	42,86	0,95	75,00	1,00
1322703	t	l	*10-20*	30	1	0,72	0,02	100,00	0,47	100,00	0,94
1322803	t	r	*focal*	40	1	0,66	0,03	100,00	0,74	100,00	1,00
1323803	t	r	*10-20*	10	2	0,57	0,00	50,00	0,86	100,00	0,97
1324103	p	r	*random*	40	3	0,72	0,09	100,00	0,07	66,67	0,00
1324803	c	r	*random*	40	6	0,67	0,05	0,00	0,04	33,33	1,00
1324903	c	r	*10-20*	20	10	0,70	0,19	30,00	0,12	30,00	0,06
1325003	-	-	*random*	20	1	0,77	0,02	100,00	0,64	100,00	0,92
1325103	t	r	*random*	10	5	0,74	0,06	60,00	0,47	40,00	0,04
1325403	t	l	*focal*	40	2	0,72	0,08	50,00	0,00	0,00	0,00
1325603	-	-	*focal*	40	4	0,66	0,02	0,00	0,17	75,00	1,00
1325903	f	r	*focal*	20	6	0,91	0,04	16,67	0,03	16,67	0,54

1326003	f	b	*focal*	10	16	0,64	0,07	35,29	0,34	25,00	0,00
1326103	t	r	*random*	40	1	0,62	0,00	0,00	0,21	100,00	1,00
1326503	p	r	*random*	20	1	0,81	0,03	100,00	0,25	100,00	0,64
1327403	t	l	*10-20*	40	9	0,74	0,03	30,00	0,13	44,44	0,34
1327903	f	r	*10-20*	30	6	0,86	0,12	66,67	0,11	33,33	0,00
1328603	t	l	*random*	20	16	0,69	0,20	29,41	0,52	43,75	0,71
1328803	f	r	*random*	30	2	0,68	0,06	50,00	0,13	50,00	0,79
1328903	t	b	*random*	20	1	0,73	0,02	100,00	0,88	100,00	0,97
1329303	t	l	*random*	20	1	0,81	0,00	100,00	0,61	100,00	0,91
1329503	f	r	*10-20*	30	4	0,63	0,32	100,00	0,13	50,00	0,00
1330103	t	l	*random*	10	1	0,76	0,01	100,00	0,25	100,00	0,40
1330203	t	r	*focal*	40	3	0,77	0,03	100,00	0,17	66,67	0,01
1330903	h	l	*focal*	40	6	0,69	0,07	33,33	0,21	50,00	0,91

Appendix B

Summary of the results obtained by the best models in the testing set data for invasive EEG recordings. The statistically significant cases are highlighted in the last column.

Pat. ID is the patient identification; *Loc.* is the localization of the epileptic focus (c. = central, f = frontal, o = occipital, p = parietal, t = temporal, and h = hemisphere); *Lat.* is the lateralization (b = bilateral, l = left, and r = right), *Elect. array* is the electrode array of the best model (*random* = random selection, *focal* = selection based on seizure onset zone and propagation); *SOP* is the seizure occurrence period (or pre-ictal time); *#Seiz.* is the number of seizures in the testing data; *Train F-mea.* is the *F-measure* determined in the training set; *Test F-mea.* is the *F-measure* determined in the test set; *Sens. (%)* is the ratio of predicted seizures over all seizures in the training set; *FPR.h*$^{-1}$ is the false positive rate raised by the predictor; *Crit. S. (%)* is the critical sensitivity of the random predictor; and *p-value* is the *p*-value of the statistical test.

Pat. ID	**Loc.**	**Lat.**	**Elect. array**	**SOP**	**# seiz.**	**Train *F-mea.***	**Test *F-mea.***	**Sens. (%)**	**FPR.h^{-1}**	**Crit. S. (%)**	***p*-value**
700	t	l	*random*	40	2	0,83	0,01	0,00	0,08	50,00	1,00
3800	p	l	*focal*	40	1	0,65	0,08	0,00	0,34	100,00	1,00
52303	-	-	*random*	40	3	0,91	0,09	33,33	0,16	66,67	0,92
72002	-	-	*focal*	10	3	0,78	0,00	0,00	0,10	33,33	1,00
79502	-	-	*random*	40	6	0,63	0,06	16,67	0,15	50,00	0,99
92202	f	l	*focal*	40	22	0,68	0,27	4,55	0,06	18,18	1,00
95802	t	l	*random*	10	11	0,47	0,01	18,18	0,97	45,45	1,00
97002	t	r	*focal*	20	16	0,68	0,10	62,50	0,39	37,50	0,00
107302	t	r	*focal*	20	15	0,65	0,03	20,00	0,32	33,33	0,81
107702	t	r	*random*	30	6	0,62	0,05	33,33	0,20	50,00	0,59
107802	f	r	*random*	20	19	0,56	0,18	5,00	0,00	0,00	0,00
108402	f	l	*focal*	10	56	0,70	0,27	55,00	0,29	12,50	0,00
112502	t	r	*random*	40	11	0,54	0,03	18,18	0,54	63,64	1,00

114602	o	r	*random*	20	11	0,69	0,06	36,36	0,26	36,36	0,08
115002	f	r	*focal*	30	5	0,65	0,07	20,00	0,18	40,00	0,97
1235903	t	r	*random*	30	4	0,77	0,02	25,00	0,42	75,00	1,00
1236703	f	l	*random*	20	21	0,49	0,05	14,29	0,36	33,33	0,99
1241003	f	l	*focal*	20	2	0,65	0,01	50,00	0,27	100,00	0,76
1242703	t	r	*random*	20	2	0,67	0,01	0,00	0,17	50,00	1,00
1243603	t	r	*focal*	40	3	0,60	0,02	0,00	0,32	100,00	1,00
1244803	t	l	*random*	40	4	0,69	0,17	0,00	0,06	50,00	1,00
1245203	t	r	*random*	40	6	0,70	0,03	20,00	0,13	50,00	0,99
1256303	o	-	*random*	10	2	0,73	0,00	50,00	0,27	50,00	0,52
1259203	-	l	*random*	40	1	0,91	0,11	100,00	0,05	100,00	0,24
1270403	t	l	*random*	40	5	0,79	0,04	20,00	0,22	60,00	1,00
1272703	t	l	*10-20*	40	8	0,69	0,24	14,29	0,00	0,00	0,00
1310403	-	-	*focal*	30	1	0,80	0,09	100,00	0,06	100,00	0,20
1311203	-	-	*random*	20	9	0,54	0,02	22,22	0,26	33,33	0,77
1322903	-	-	*random*	40	3	0,82	0,12	66,67	0,35	100,00	0,62
1324503	-	-	*random*	30	9	0,69	0,06	33,33	0,08	22,22	0,03
1327603	-	-	*focal*	40	7	0,77	0,06	14,29	0,16	42,86	1,00