KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Engineering Science

UNIVERSIDAD de ORIENTE
Faculty of Engineering in Telecommunications, Informatics, and Biomedical Engineering

Methods for Nonconvulsive Epileptic Seizure Monitoring

Yissel Rodríguez Aldana

Supervisors:
Prof. dr. ir. E. J. Marañón Reyes
(Universidad de Oriente)
Prof. dr. ir. S. Van Huffel
(KU Leuven)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

September 2018

Methods for Nonconvulsive Epileptic Seizure Monitoring

Yissel RODRÍGUEZ ALDANA

Examination committee:
Mr. De Voorzitter, chair
Prof. dr. ir. E. J. Marañón Reyes, supervisor
(Universidad de Oriente)
Prof. dr. ir. S. Van Huffel, supervisor
(KU Leuven)
Jurylid 1
Jurylid 2
External jury member
(External affiliation)

Dissertation presented in partial fulfillment of the requirements for the degree of Doctor of Engineering Science (PhD): Electrical Engineering

September 2018

© 2018 KU Leuven – Faculty of Engineering Science
Uitgegeven in eigen beheer, Yissel Rodríguez Aldana, Celestijnenlaan 200A box 2402, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

Preface

I would like to thank Dr. Valia Rodríguez Rodríguez enormously, for being the person who introduced me to this topic. For being the one who after much struggle to obtain data for the research, with a lot of kindness and disinterestedly, facilitated the data she had. She helped in the data labeling and had the patience to teach me how a non-convulsive seizure looked like in the EEG, how it evolves, how different it was from one patient to another. Actually, without her help, I do not think this research could have been carried out.

My most sincere thanks also go to Dr. Borbála Hunyadi, who was a great support and a good friend during the development of the research. She helped me to take my first steps in the analysis with tensors. She understood what I wanted to say even before I realized it myself. Her criticism always forced me to see beyond what it was evident to me and to study.

I also want to thank Prof. Sabine Van Huffel. Thank you for trusting me, for giving me the opportunity to work with you. Thank you for helping me to improve my work methodology, which was undeveloped before work with you. Thank you very much for your support, for your patience with my unstructured way of working and lack of rigor.

I am grateful to Prof. Enrique J. Marañón Reyes, who welcomed me as a daughter and has always been by my side guiding my steps on the right path. Thank you very much for your patience towards my changing and capricious character. Excuse me for being a daughter that causes headaches.

Many thanks to Prof. Frank Sanabria Macias, for his infinite patience to teach and his enormous willingness to help. Thank you for being a good friend.

Thanks to all the members of BIOMED, present ones and previous ones. Thank you very much for supporting me, for making my life far from home more enjoyable, thank you very much for your friendship.

I also want to thank the VLIR project not only for financing my research but

also for giving me the opportunity to meet wonderful people, inside and outside my country. For giving me the opportunity to approach those people who even when working in the same university I did not have the privilege of knowing. For allowing me to foster friendships that I hope will last for many years.

Last but not least, I would like to thank the most important people in my life, my family, for giving me all the support of the world. Mammy and daddy who have done everything for me, and who always feel proud of their daughter.

Abstract

Nonconvulsive status epilepticus (NCSE) is a condition where the patient is exposed to abnormally prolonged nonconvulsive epileptic seizures (NCES)(epileptic seizures without evident physical symptoms). Hence, the diagnosis can only be stated by means of EEG monitoring. NCES could have long-term consequences as neuronal death, neuronal injury, and alteration of neuronal networks. In general NCES are associated with poor outcome, mostly in those patients with recurrent seizures. Since these continuous seizures may cause permanent brain damage, the prompt recognition of patients at risk of suffering NCSE is necessary in order to manage them properly and to prevent further brain injury. However, despite the clinical efforts to manage NCES and NCSE, and improve the patient's outcome, monitoring this pathology in real-time is very difficult.

In this thesis a patient-specific method is presented and evaluated for real-time NCES detection from scalp EEG applying a combination of tensor analysis and machine learning algorithms. The automated NCES detection facilitates alerting the doctors about the patient risk to evolve to NCSE. In particular, this proposal attempts to identify the NCES based on their similarity to the first NCES detected by the physician on the EEG. The analysis is performed by expanding the EEG signal in third-order tensors, with dimensions "$frequency \times time \times channels$". The components extracted for each dimension are used as input to a binary classification problem to detect the NCES. This approach is shown to outperform several methods proposed in the literature for NCES displaying accuracy values over 98%. Yet, a decrease in performance was observed during evaluation when EEG pattern morphology underwent changes.

To deal with the EEG morphology changes, an incremental learning algorithm is added to the initial patient specific method. The incremental learning addition is shown to improve the overall performance of the algorithm, thereby increasing the sensitivity and accuracy values above 99%.

The proposed methods proved to have an acceptable performance as they detect

seizures with high sensitivity and specificity, while the computational cost is low enough for the real-time implementation.

Resumen

El status epiléptico no convulsivo (NCSE, por sus siglas en inglés) es una condición en la que el paciente está expuesto a crisis epilépticas no convulsivas anormalmente prolongadas (NCES, por sus siglan en inglés) (crisis epilépticas sin síntomas físicos evidentes). Por lo tanto, el diagnóstico solo puede establecerse por medio del monitoreo del EEG. Las NCES podría tener consecuencias a largo plazo como muerte neuronal, lesión neuronal y alteración de las redes neuronales. En general, las NCES se asocian con altos índices de mortalidad y morbilidad, principalmente en aquellos pacientes con crisis recurrentes. Debido a que estas crisis continuas pueden causar daño cerebral permanente, es necesario reconocer rápidamente a los pacientes con riesgo de sufrir una NCES para manejarlos adecuadamente y prevenir lesiones cerebrales adicionales. Sin embargo, a pesar de los esfuerzos clínicos para controlar NCES y NCSE, y mejorar el resultado del paciente, monitorear esta patología en tiempo real es muy difícil.

En esta tesis, se presenta y evalúa un método paciente-específico para la detección de NCES en tiempo real. Para realizar la detección se utiliza EEG superficial al que se le aplica una combinación de análisis tensorial y algoritmos de aprendizaje automático. La detección automática de NCES facilita alertar a los médicos sobre el riesgo del paciente de evolucionar a NCSE. En particular, esta propuesta intenta identificar la NCES en función de su similitud con la primera NCES detectada por el médico en el EEG. El análisis se realiza expandiendo la señal EEG en tensores de tercer orden, con dimensiones "$frecuencia \times tiempo \times canales$". Los componentes extraídos para cada dimensión se utilizan como entrada a un problema de clasificación binaria para detectar el NCES. En el desarrollo de la investigación, se demuestra que este enfoque supera a varios métodos propuestos en la literatura para detección de NCES obteniendo valores de precisión superiores al 98%. Sin embargo, se observó una disminución en el rendimiento durante la evaluación cuando la morfología del patrón EEG sufre cambios en el tiempo.

Para lidiar con los cambios en la morfología del EEG, se introduce un algoritmo

de aprendizaje incremental a la propuesta inicial. La adición de aprendizaje incremental mejora el rendimiento general del método inicial, elevando los valores de sensibilidad y precisión sobre 99%.

Los métodos propuestos demostraron tener un rendimiento aceptable para su aplicación al monitoreo de NCES en tiempo real ya que detectan las crisis con alta sensibilidad y especificidad, y bajo costo computacional.

Beknopte samenvatting

Non-convulsieve status epilepticus (NCSE) is een toestand waarin de patiënt abnormaal lange non-covulsieve epileptische aanvallen (NCES) ondergaat (i.e. aanvallen zonder duidelijke fysieke symptomen). Dergelijke aanvallen kunnen alleen gediagnosticeerd worden door gebruik te maken van EEG monitoring. NCES's gaan gepaard met langetermijngevolgen zoals het afsterven van neuronen, neuronschade, en het wijzigen van neuronale netwerken. In het algemeen worden NCES's geassocieerd met een negatieve prognose, meestal bij patiënten met wederkerende aanvallen. Aangezien deze continue aanvallen permanente hersenschade kunnen berokkenen, is het nodig om patiënten met een risico op NCSE snel te herkennen zodat ze correct opgevolgd kunnen worden en verdere hersenschade vermeden kan worden. Ondanks klinische inspanningen om NCES's en NCSE te controleren en het vooruitzicht voor de patiënt te verbeteren, is het bijzonder moeilijk om dit ziektebeeld in real-time op te volgen.

In deze thesis wordt een patiëntspecifieke methode voorgesteld en geëvalueerd voor het detecteren van NCES's in real-time, gebruik makend van scalp-EEG waarop een combinatie van tensoranalyse en machine learning algoritmes wordt toegepast. De geautomatiseerde NCES-detectie faciliteert het waarschuwen van artsen over het patiëntenrisico om te evolueren naar NCSE. De voorgestelde methode identificeert de NCES's door ze te vergelijken met eerdere NCES's die door een arts gedetecteerd werden in het EEG-signaal. De analyse wordt uitgevoerd door de EEG-data te expanderen tot een derde-orde tensor met dimensies "$frequentie \times tijd \times kanalen$". De verkregen componenten voor elke dimensie worden gebruikt als input voor een binaire classifier voor het detecteren van NCES. Deze aanpak werkt aantoonbaar beter dan andere methoden die reeds beschreven staan in de NCES-literatuur. Onze aanpak haalt een accuraatheid van meer dan 98%. Toch werd er een daling in performantie vastgesteld tijdens de evaluatie van onze methode, dit als gevolg van een variabel EEG-patroon onder invloed van een veranderende morfologie.

Om om te kunnen gaan met dit veranderend EEG signaal werd een incrementeel learning algoritme toegevoegd aan de initiële patiënt specifieke methode. De toevoeging van de incrementele learning verbetert de performantie van het algoritme, waardoor een stijging in sensitiviteit en accuraatheid tot 99% wordt vastgesteld.

De voorgestelde methode heeft een aantoonbare acceptabele performantie, omdat aanvallen gedetecteerd worden met een hoge sensitiviteit en specificiteit, desondanks blijft de computationele kost laag genoeg voor het bewerkstelligen van een real-time implementatie.

List of Abbreviations

ACNS	American Clinical Neurophysiology Society
BTD	Block Term Decomposition
CES	Convulsive Epileptic Seizures
CPD	Canonical Polyadic Decomposition
CSE	Convulsive Status Epilepticus
CWT	Continuous Wavelet Transform
DP	Discrimination Power
ECG	Electrocardiogram
EEG	Electroencephalogram
EMD	Empirical Mode Decomposition
EMG	Electromyogram
EOG	Electrooculogram
ES	Epileptic seizures
FPR	False Positive Rate
HHT	Hilbert Huang Transform
ICU	Intensive Care Unit
ILAE	International League Against Epilepsy
IMF	Intrinsic Mode Function
IQR	Interquartile Range
KNN	K-Nearest Neighbors
LDA	Linear Discriminant Analysis
LMLRA	Low Multilinear Rank Approximation
MAD	Median Absolute Deviation
MD	Modeling Power
MLSVD	Multilinear Singular Value Decomposition
MRMR	Maximum Relevance and Minimum Redundancy

NCES	Nonconvulsive Epileptic Seizures
NCSE	Nonconvulsive Status Epilepticus
n-NCES	Seizure-free EEG
PC	Principal Component
PCA	Principal Component Analysis
PCR	Principal Component Regression
PLS-DA	Partial Least Squares Discriminant Analysis
PLS	Partial Least Squares
PPV	Positive Predictive Value
SE	Status Epilepticus
SVD	Singular Value Decomposition
SVM	Support Vector Machine
SVMRB	Support Vector Machine and Radial Basis Functions

Contents

Abstract iii

Resumen v

Beknopte samenvatting vii

List of Abbreviations ix

Contents xi

List of Figures xv

List of Tables xx

1 Introduction 1

1.1 Objectives 2

1.2 Contributions 3

1.3 Thesis outline 4

1.4 Collaborations 5

2 Nonconvulsive Epileptic Seizure Monitoring: An Overview 7

2.1 Introduction 7

2.2 Epilepsy 7
2.3 Nonconvulsive Status Epilepticus 9
2.4 Electroencephalography (EEG) 11
2.5 NCSE monitoring 13
2.6 Automatic Detection of Epileptic Seizures 16
2.7 Automatic Detection of Nonconvulsive Epileptic Seizures 17
2.8 Conclusion 20

3 Feature Selection Techniques and Machine Learning 22

3.1 Introduction 22
3.2 Signal Analysis Techniques 22
3.2.1 Continuous Wavelet Transform (CWT) 22
3.2.2 Hilbert Huang Transform (HHT) 25
3.3 Blind Source Separation Techniques 29
3.3.1 Principal Component Analysis (PCA) 29
3.3.2 Canonical Polyadic Decomposition 30
3.3.3 Multilinear Singular Value Decomposition(MLSVD) and Low Multilinear Rank Approximation(LMLRA) 31
3.3.4 (L,M,O)- Block Term Decomposition 32
3.4 Feature selection techniques 33
3.4.1 Modeling and Discrimination Power 34
3.4.2 Maximum relevance and minimum redundancy (MRMR) 36
3.4.3 Bhattacharyya Distance 37
3.5 Machine Learning Techniques 39
3.5.1 Linear Discriminant Analysis (LDA) 39
3.5.2 Support Vector Machine (SVM) 40
3.5.3 K-Nearest Neighbors(KNN) 41
3.5.4 Partial Least Squares discriminant analysis (PLS) . . . 42

3.6 Performance Metrics 43
3.7 Conclusions 45

4 EEG Data 46

4.1 Introduction 46
4.2 EEG Data description 46
4.3 EEG training and testing data 52

5 NCES Detection using scalp EEG 54

5.1 Introduction 54
5.2 Tensor Formulation 55
5.3 Tensor Decomposition 55
5.3.1 Tensor Rank Estimation 57
5.4 Hypothesis Test 60
5.5 Classifiers 65
5.6 Results and Discussion 65
5.7 Conclusion 73

6 Nonconvulsive epileptic seizure monitoring with incremental learning 75

6.1 Introduction 75
6.2 Incremental learning 77
6.2.1 Alternative batch approaches 83
6.3 Results and Discussion 85
6.3.1 Incremental Learning Approaches Comparison 85
6.3.2 MRMR_IL and Batch methods Comparison 88
6.4 Conclusion 92

7 Conclusions 93

7.1 Summary of obtained results 93

7.2 Future work and challenges 95

Bibliography **97**

Curriculum Vitae **107**

List of publications **108**

List of Figures

1.1 Chapters organization chart. 6

2.1 Focal seizures (A) originate in one hemisphere of the brain. Focal seizures are either simple or complex, depending on the patient's awareness. If if he remains fully conscious during the seizure, it is classified as simple. In case the patient is unconscious the seizure is classified as complex. Generalized onset seizures (B) affect both hemispheres of the brain simultaneously [74]. 9

2.2 Framework for epilepsy classification [70]. 10

2.3 Electrode placement according to the International 10-20 system [44]. 12

2.4 ICU patient with confirmed NCES presenting periodic discharges 14

2.5 Rhythmic Delta activity recorded from an ICU patient with altered mental status and having NCES. 14

2.6 Spike-and-wave EEG pattern recorded from an ambulatory patient diagnosed as having absence seizures. 15

2.7 Sharp-slow wave EEG pattern recorded from an ambulatory patient having NCES after a head trauma. 15

3.1 a)Mexican hat mother wavelet definition and representation. b) Examples of scaling and shifting of the Mexican hat wavelet. Each row corresponds to a scaling value a and each column corresponds to a shift value b. 23

3.2 EEG tensorization process with CWT. 24

3.3 EMD of an EEG segment. The first plot displays the raw EEG. The plots from 2 to 5 display the IMFs obtained from the EEG. Finally the last plot show series tendency or residue $r(t)$. . . . 26

3.4 EEG tensorization process with HHT. 28

3.5 PCA of a matrix X in R rank-1 matrices. 29

3.6 CPD of a third order ($N = 3$) tensor X in R rank-1 terms were $a_r = d_r^{(1)}$, $b_r = d_r^{(2)}$, $c_r = d_r^{(3)}$, $r = 1, \ldots, R$. 30

3.7 BTD of a tensor X in R (L,M,O)-rank terms. 33

3.8 General structure to build the confusion matrix from the classifier output. 44

4.1 Modified 10-20 electrode montage settings used in the EEG recording process. Setups of A)8, B)13, C)14, D)17, and E)19 electrodes. 48

4.2 Absence seizure recorded from an ambulatory patient (delimited by red lines). In the EEG segment, a clear evolution in the EEG pattern can be observed in frequency and morphology during the seizure. In contrast no changes can be noticed in the EMG signal registered. The absence of changes in the EMG confirms the lack of motor signs during the episode. 49

4.3 NCES from a ICU stuporous patient. As can be noticed the EEG evolution in this patient seizure is not as clear as it is for the absence seizure from Fig. 4.2. The absence of muscular activity can be confirmed by the lack of changes in the EMG during the seizure. 50

4.4 Description of an EEG data training set. The figure illustrates the EEG recording from the beginning of the recording process until the end of the first NCES detected. The red bars indicate the beginning and end of the seizure. The blue box indicates an example of the EEG data selected as training set. 52

4.5 Description of an EEG data test set. The figure illustrates the EEG recording from the end of the first NCES detected to the beginning of the second seizure. The two red bars indicate the ending of the first NCES and the beginning of the second NCES respectively. The blue box indicates an example of the EEG data selected as test set. 53

5.1 General block diagram of the proposed NCES detector using scalp EEG. 56

5.2 LMLRA analysis Frequency, Time and Space mode histograms. Axis X and Y represent the number of components and tensors respectively. 57

5.3 Obtained MP and DP for H_Tensors and W_Tensor CPD signatures computed for several numbers of factors. The displayed MP values are computed for the feature matrices of the first NCES. The MP and DP values displayed on the charts are averaged over the number of features of each mode. The red boxes indicate the selected decomposition. 58

5.4 Obtained MP and DP values computed for all tensor representations considered for BTD in the W_Tensor *ml-rank* analysis. The displayed MP values are computed for the feature matrices of the first NCES. The MP and DP values displayed on the charts are averaged over the number of features of each mode.The red boxes indicate the selected decomposition. 59

5.5 Obtained MP and DP values computed for all tensor representations considered for BTD in the H_Tensors ml-rank analysis. The displayed MP values are computed for the feature matrices of the first NCES. The MP and DP values displayed on the charts are averaged over the number of features of each mode.The red boxes indicate the selected decomposition. 60

5.6 p-values resulting from performing Wilcoxon signed-rank test applied to the two performed hypothesis tests. The results in the red boxes correspond to the hypothesis test performed between the first seizure and the subsequent seizures. The blue boxes show the results from the hypothesis test performed between the first NCES and the n-NCES segments. High average p-value (close to 1) means that H0 is accepted. 61

5.7 Bhattacharyya distance computed for each signature between the first seizure and the subsequent seizures (blue boxes), and between the first NCES and the n-NCES segments (red boxes). These distance values were calculated for all signatures obtained from all tensors in all training sets of all cases specified in Table 4.2 decomposed with CPD. 63

5.8 Histograms generated by the Space signature distribution of the first seizure, subsequent NCES and n-NCES data for two different cases described in Table 4.1(Case3 (A) and Case10 (B)). The Space signatures were obtained for the H_Tensor decomposed with CPD. 64

5.9 Specificity, Sensitivity, and Accuracy scores obtained in the classification of the test set for KNN, SVMRB and LDA. The results for the different signatures are displayed in the following order: A) Frequency, B)Time, C)Space, and D) Assembled signatures. The columns in the graph for each classifier stand for specificity, sensitivity, and accuracy respectively. 67

5.10 FPR values displayed by the KNN, SVMRB and LDA classifiers during the different trainings. 69

5.11 H_Tensors CPD: false positive classification example. Misclassified n-NCES epoch shows a short paroxysm of approximately 1 s of seizure-like activity in the EEG. This activity is captured by highly similar signatures as the first seizure's time and frequency signatures. This leads to the detection of the paroxysm as a seizure. Signatures are represented in the following order: Space, Frequency and Time. 70

5.12 H_Tensors CPD: false negative classification example. Seizure epoch classified as a non-seizure. The EEG seizure patterns from the misclassified epoch are dissimilar to the patterns from the first seizure. The space signature from the misclassified epoch shows an inverted activation pattern compared to the one shown by the first seizure. This inverted pattern could be interpreted as a lighter seizure or a non-seizure EEG segment. The frequency component from the misclassified epoch is steeper than the first seizure frequency component. The time component from the misclassified epoch is smoother than the one from the first seizure. However, the time component from both epochs are quite similar. The dissimilarities found between the computed components could explain the classifier failure. Signatures are represented in the following order: Space, Frequency and Time. . 71

6.1 Hard_IL execution time (left) and accuracy (right) values obtained for different values of tr. The test was conducted for the recording number 10 which is the one that completes more training rounds. 82

6.2 Flow chart of the NCES detection method with incremental learning. Section (A) describes the Batch method (see Fig. 5.1 for details). The three implementations of incremental learning approaches tested: HardIL, Cross_IL, and MRMR_IL are described in sections (B),(C), and (D) of the diagram. For the graph simplicity, $\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_t, y_t\}$ refers to the samples resulting from the double labeling. SV stands for support vector. 84

6.3 Misclassified epoch by The Incremental Learning approach from Case 6. The red box indicates seizure activity. The blue boxes mark the preictal and postictal EEG patterns. As can be observed the patterns from both preictal and postictal activities are quite similar. Since the classifier considers the postictal activity as seizure, this similarity could cause the classifier failure. 90

List of Tables

2.1 Reviewed publications, database characteristics, summary of features, classifiers, performance and drawbacks. N/R: Not reported. 21

4.1 Description of the database of EEG recordings. The Table describes the protocol and type of medical setting for each recording, as well as the patient etiology. As can be seen, the etiology of the patients and the duration of the recordings is very different. Regarding the recording's duration, a clear difference can be appreciated between the ones from the Ambulatory setting and the ones from ICU. In the table 'x' and 'o' indicate whether the signal was recorded or not for the patient respectively. 51

4.2 Available Data for NCES and seizure free (n-NCES) class and the resulting training sets for each EEG recording given in epochs of 3s duration 53

5.1 Bhattacharyya distance computed between the first seizure and the subsequent NCES and n-NCES segments. These distance values were calculated for the Space signature obtained from the H_Tensor decomposed with CPD. 64

5.2 Rough comparison of the proposed method performance with other existing methods in terms of accuracy, sensitivity and specificity. N/R stands for "Not Reported". 72

6.1 Average Specificity, Sensitivity, PPV, and FP per hour values obtained for the Batch Method and the three implemented incremental learning algorithms. In the table Spec, and Sen stand for Specificity and Sensitivity metrics respectively. 87

6.2 Number of false positive and negative detections of the MRMR_IL method with and without double labeling, compared to the erroneous detections made by the Batch Method. 88

6.3 Specificity, Sensitivity, and PPV, scores obtained in the classification for MRMR-IL, Kollialil(1 and 2), Sharma ,and Fatma. In the table Spec, and Sen stand for Specificity and Sensitivity metrics respectively. 91

Chapter 1

Introduction

This work has been supported by the Belgian foreign Affairs-Development Cooperation through VLIR-UOS (2013-2019) (Flemish Interuniversity Council-University Cooperation for Development) in the context of the Institutional University Cooperation program with Universidad de Oriente. The thesis is subscribed to the project number two (P2,"Research and applications in biomedical images and signal processing") within the framework of the project. The P2 aims to improve the applications and social impact of the University manufactured medical technology in Cuba, as well as the training and evaluation of this technology, to increase the level of the technical assistance and to contribute to the technological development of Cuba.

This project supports the development of new medical technology by Universidad de Oriente and its application to improve the medical service in the eastern region of Cuba. It focuses on the more relevant pathologies prevalent in Cuba, thereby maximizing the impact of the potential developments. The pathologies of interest include oncology, sickle cell anemia, pathologies associated with neonates, cardiac, and brain disorders. The area addressed by this thesis is related to the field of brain disorder recognition, more specifically to the nonconvulsive status epilepticus detection.

Status epilepticus (SE) is a condition where the patient is exposed to abnormally prolonged epileptic seizures. This condition could have long-term consequences as neuronal death, neuronal injury, and alteration of neural networks, depending on the seizure type and duration [83]. SE can be divided into two major categories: convulsive SE (CSE) and nonconvulsive SE (NCSE). CSE is characterized by muscular spasms typical of convulsive epileptic seizures (CES). On the other hand, NCSE symptoms are too subtle to be noticed since

this status is characterized by nonconvulsive seizures, which show persistent epileptic activity in the electroencephalogram (EEG) without visible physical manifestations.

Nonconvulsive epileptic seizures (NCES) can be found associated with coma/stupor or not (e.g., absence status, focal SE). Patients without coma/stupor could show symptoms such as automatisms, facial twitching, eye deviation or jerking that might help NCES diagnosis. On the other hand, patients with coma/stupor treated at intensive care units (ICU) are usually unconscious or unresponsive. Consequently, these NCES have no observable physical symptoms at all. Hence, the diagnosis can only be accomplished using EEG monitoring. As explained before, these seizures are often associated with severe brain damage with very poor prognosis [89]. The prompt recognition of patients at risk of suffering NCSE is necessary in order to manage them properly and to prevent further brain injury. Therefore, it is crucial to monitor the EEG in the ICU and emergency units, to allow doctors to detect seizures when it is still possible to prevent the patient from undergoing NCSE. In case that it is not possible to identify the NCSE before it occurs, it should at least be detected within 24 hours after the onset, when it is still possible to avoid permanent brain damage.

Due to the long-term nature of EEG recordings from patients at risk of suffering NCSE, its analysis is a time-consuming task. This task can be performed by either clinical doctors or seizure detection software. The visual inspection conducted by clinicians might lead to intra-observational inconsistencies in the results (the same expert can annotate the seizures differently when (s)he reviews the data multiple times) [71]. Software approaches are more time efficient, more consistent and objective.

Most of the available automated seizure detection software packages have been extensively trained on seizures obtained from patients with seizure disorders without brain injuries. Nevertheless, these seizures differ from those of comatose/brain-injured patients. The ictal activity from brain-injured/comatose patients has a longer duration, is less defined in time (unclear on-set /off-set), less organized and lower in maximum frequency [71]. Therefore, the performance of the algorithms proposed to detect convulsive epileptic seizures, is lower when applied to detect NCES.

1.1 Objectives

In this context, given the limitations of the EEG analysis techniques proposed for monitoring and diagnosis of patients at risk of suffering from NCSE in ICUs

and others medical settings, this research aims to develop methods based on scalp EEG analysis that allow accurate detection of the NCES and hence early diagnosis of NCSE. For this purpose, the natural multichannel structure of the EEG is exploited using tensor analysis of the data and machine learning techniques. The study objectives include:

- Develop a NCES detection algorithm based on scalp EEG.
- Implement an alarm system for early detection of NCSE.
- Assess the comparison of the proposed methods with the ones available in the literature with the same aim.

1.2 Contributions

As stated before, the automated NCES detection facilitates multiple alerting and therapeutic applications, such as the diagnosis of the NCSE. One of the disadvantages of the current methods is that they rely on thresholds (the number of channels affected by the seizure and the seizure duration are the most common ones). In fact, they are not typically robust and result in lower sensitivities and specificities. In this thesis, a patient-specific approach for real-time NCES detection is presented and evaluated. This approach is based on scalp EEG and it combines tensor analysis and machine learning algorithms.

We propose an alternative combination for tensor expansion based on the Hilbert Huang Transform and tensor analysis. This combination enables to represent the EEG data as a third order tensor from which the more powerful features for the NCES data classification are extracted.

An optimal combination of features and classifier is found. The performance of such combination guarantees an optimal separation between the NCES and the seizure-free EEG thereby achieving a very low average false positive rate.

The proposed algorithm implements an alarm system that permits early identification of the nonconvulsive status epilepticus.

Finally, we introduce an incremental learning algorithm to reinforce the method performance in long-term EEG monitoring. The addition of the Incremental Learning approach enables the proposed method to adapt to changes in the patient's EEG pattern thereby improving the proposed method performance.

1.3 Thesis outline

The thesis is organized as follows:

Chapter 2 provides an overview of the background of epilepsy, NCSE, and NCES from a neurological point of view. The fundamentals of the use of the EEG signal for NCSE and NCES monitoring and diagnosis are discussed. Also, a review of the available methods for automatic NCES detection is presented. Finally, possible improvements that could be introduced in the present research are discussed.

Chapter 3 introduces the methods applied for feature selection and classification for NCES detection. Blind source separation (BSS) techniques are described for tensor analysis. Numerous feature selection methods used in the thesis are presented. Several classifiers are introduced as binary classification problem solvers.

Chapter 4 describes the database used during the thesis to develop the methods proposed. The procedure for handling the data during the training and testing phases is detailed.

Chapter 5 proposes a novel NCES detection method based on tensor analysis. Some performance metrics are defined and the proposed algorithm is evaluated using a database containing scalp EEG from fourteen patients with NCES. A comparison between tensor representation methods, features and classifier combinations is presented, to select the best combination for NCES data classification.

Chapter 6 introduces an incremental learning algorithm to the method proposed in Chapter 5. The incremental learning approach updates the classifier training over time, enabling the improved method to adapt to changes in the patient EEG. The results of the system improved with Incremental Learning are compared with other algorithms proposed in the literature for NCES to prove its suitability for this task.

Chapter 7 summarizes the conclusions of this research and provides a short discussion about the NCES detection methods which require further study and investigation. Potential ways to enhance the results and algorithms are presented as possible future work.

The chapters' organization is presented in Fig. 1.1.

1.4 Collaborations

This Ph.D. research was conducted within the Center of Neuroscience and Signals and Image Processing, Faculty of Telecommunications, Informatics and Biomedical Engineering, Universidad de Oriente, and the Biomedical Data Processing research group (BioMed), Stadius Center for Dynamical Systems(Stadius), Signal Processing and Data Analytics, Department of Electrical Engineering (ESAT), KU Leuven, under the supervision of Prof. Enrique J. Marañón Reyes and Prof. Sabine Van Huffel respectively.

The work presented in this thesis was done in collaboration with Dr. Valia Rodríguez Rodríguez (Aston University, United Kingdom) who provided the NCES EEG data and trained me in the recognition of NCES patterns in the EEG recordings.

Additionally, this research was carried out in close collaboration with Dr. Borbála Hunyadi (ESAT-Stadius, KU Leuven) whose research and advice influenced in a significant manner the methodological choices of this thesis.

Finally, the research benefited from the collaboration with Prof. Lieven De Lathauwer and Martijn Boussé from the BIOTENSORS team (BIOTENSORS (n^o339804) is an ERC advanced grand funded by the European Research Council under the European Union's Seventh Framework Program (FP7/2007-2013), grant holder: Sabine Van Huffel), who provided their advice in the tensor interpretation and analysis.

This thesis has been supported by the Belgian foreign Affairs-Development Cooperation through VLIR-UOS (2013-2019) (Flemish Interuniversity Council-University Cooperation for Development) in the context of the Institutional University Cooperation program with Universidad de Oriente.

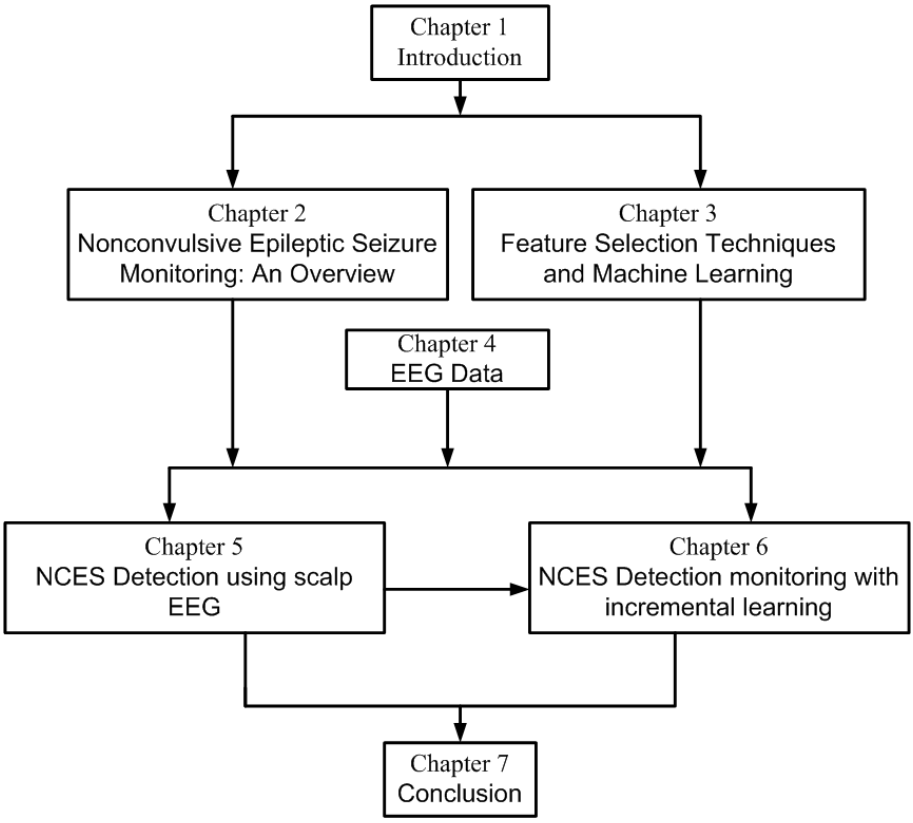


Figure 1.1: Chapters organization chart.

Chapter 2

Nonconvulsive Epileptic Seizure Monitoring: An Overview

2.1 Introduction

The electroencephalogram (EEG) has transformed the epilepsy detection and diagnosis. When epilepsy is suspected, the clinicians use the EEG to identify patterns that might confirm the diagnosis (e.g., interictal epileptiform discharges). Despite the vast technological advances in recent decades in the assessment of neurological disorders (e.g., neuroimaging technology), the EEG is still the primary tool for diagnosing patients with epilepsy, and it prevails as the tool with the highest time resolution for the evaluation of the brain physiological function. This chapter reviews the EEG role in epilepsy and nonconvulsive status epilepticus diagnosis. We offer an overview of the available algorithms for automatic NCES detection and introduce possible improvements in this field.

2.2 Epilepsy

Epilepsy is the second most common nervous system disease affecting about 50 million people worldwide [56]. It is a brain disease characterized by a predisposition to generate unprovoked epileptic seizures (ES) that may be

etiologically different. According to the International League Against Epilepsy (ILAE), an ES is defined as "a transient occurrence of signals and symptoms due to abnormal excessive or synchronous neural activity in the brain" [31]. In practice, epilepsy can be diagnosed following at least one ES [30]. A seizure results from abnormal synchronous discharges of cortical neurons recruiting neighboring cells into a critical mass that abruptly and temporarily alter cerebral function [74].

In a systematic review about the prevalence and incidence of epilepsy worldwide, conducted over 222 population-based studies,the annual cumulative incidence of epilepsy was found to be 67.77 per 100,000 persons while the incidence rate was 61.44 per 100,000 person-years [29]. The prevalence of epilepsy did not differ by age group, sex, or study quality [29, 58]. On the other hand, the incidence varies among different age groups. There are five age groups defined for the epilepsy analysis: (a) neonatal (0 to 30 days of life), (b) infancy (1 month to 2 years of life), (c) childhood (> 2 to 12 years of life), (d) adolescence and adulthood (> 12 to 59 years of life), and (e) elderly (> 60 years of life). Among these groups, the lowest incidence of epilepsy is found for people between 30 to 40 years, and the highest is seen in infancy (0-2 years of life) and the elderly (> 60 years of life) [58].

ES can be caused by numerous etiologies, including genetic disorders, infectious disease, metabolic disorders, cerebrovascular disease, traumatic brain injury, toxins, and degenerative nerve diseases. Generally, congenital and perinatal factors are the most common causes with onset in infancy and childhood. Meanwhile for adults, the causes of the epilepsy are more likely to be external and non-genetic. However, this discrimination is not absolute [58]. The primary etiology of the epilepsy is commonly localized in the brain, but metabolic-toxic changes associated with failure of important organ systems may cause chronic seizure disorders, and thus epilepsy [71].

According to ILAE and based on the clinical manifestations (semiology), ES can be divided as those with focal onset, generalized onset, or unknown onset (see Fig. 2.1) [70, 32]. Seizures with focal onset originate in one hemisphere of the brain, can be very localized or they can spread to the other hemisphere later in the seizure. Seizures generalized from onset originate in both brain hemispheres and spread rapidly to all neuronal networks. A seizure is classified as having an unknown onset when it is not possible to define whether epilepsy is focal or generalized [70, 32].

Fig. 2.2 shows a more general classification framework introduced in [70]. The new framework has a multilevel structure (seizure types, epilepsy types, epilepsy syndromes) that combines the previously defined seizure types and epilepsy schemes into a single mechanism. The earliest prominent manifestation defines

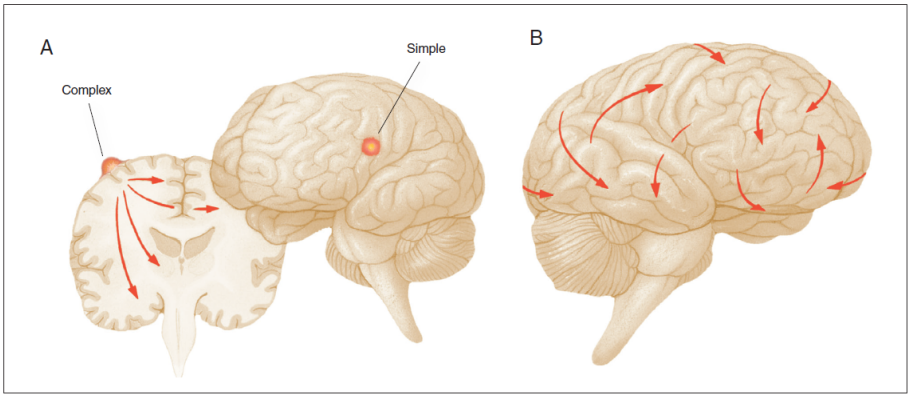


Figure 2.1: Focal seizures (A) originate in one hemisphere of the brain. Focal seizures are either simple or complex, depending on the patient's awareness. If if he remains fully conscious during the seizure, it is classified as simple. In case the patient is unconscious the seizure is classified as complex. Generalized onset seizures (B) affect both hemispheres of the brain simultaneously [74].

the seizure type, which might then progress to other signs and symptoms, e.g., focal seizures can become bilateral tonic-clonic. Tonic-clonic activity can be present from the onset of a generalized seizure or emerge in the course of a focal-onset seizure [32].

2.3 Nonconvulsive Status Epilepticus

All epileptic seizure types can evolve into prolonged status epilepticus episodes [74]. Status Epilepticus (SE) has been described by the ILAE as "a single epileptic seizure of $>$ 30-min duration or a series of epileptic seizures during which function is not regained between ictal events in a 30-minute period"[41]. Lowenstein et al. argue this definition, claiming that it was impractical for the patient management since irreversible neurological damage could appear after 30 min of ongoing seizure activity. Instead, Lowenstein proposed to divide the generalized, convulsive SE characterization in two: an operational one ("5 min of (a) continuous seizures or (b) two or more discrete seizures between which there is incomplete recovery of consciousness"), and a mechanistic one (conceptual one). "Generalized, convulsive status epilepticus refers to a condition in which there is a failure of the 'normal' factors that serve to terminate a typical GTCS",[GTCS: generalized tonic-clonic seizures]) [55]. This approach

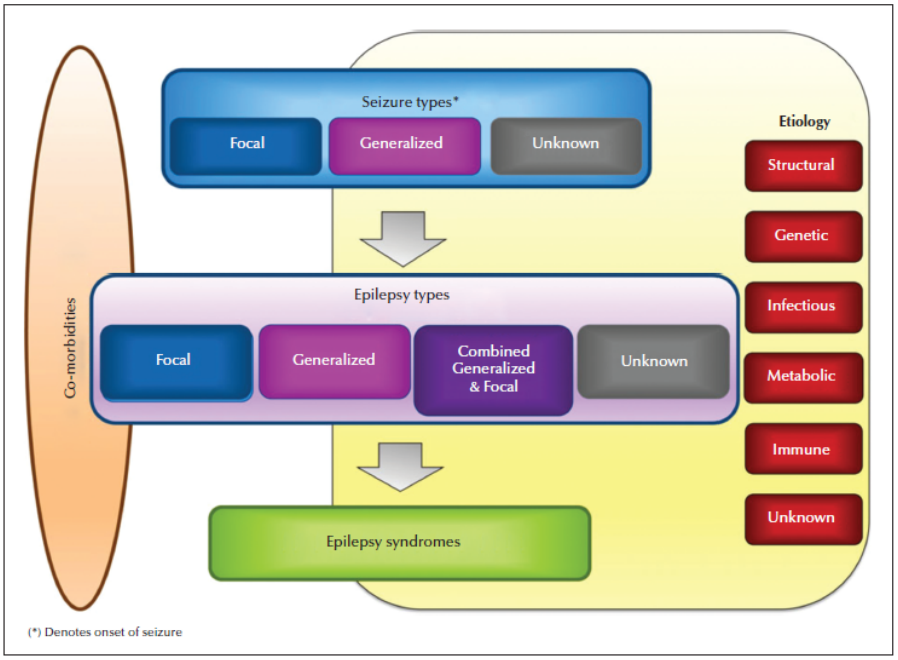


Figure 2.2: Framework for epilepsy classification [70].

has been proposed for adults and older children (>5 years old), and only refers to generalized convulsive SE. Other types of SE are not covered by this criterion.

In the 2015 ILAE special report mentioning the SE definition, a more general concept was offered by Trinka et. al.The SE is defined as "a condition resulting either from the failure of the mechanisms responsible for seizure termination or from the initiation of mechanisms, which lead to abnormally, prolonged seizures (after time point t_1). It is a condition, which can have long-term consequences (after time point t_2), including neuronal death, neuronal injury, and alteration of neuronal networks, depending on the type and duration of seizures"[83]. This definition gives a conceptual approach, adding two operational dimensions, the length of the seizure and the time points t_1(beyond which the seizure should be regarded as "continuous seizure activity") and t_2 (time of ongoing seizure activity after which there is a risk of long-term consequences).

Similarly to the rest of the epileptic disorders, SE can be classified regarding the seizure onset as focal or generalized. According to the 2016 American

Epilepsy Society report, given the signs and symptoms the SE can be divided in three main categories (1) convulsive SE, (2) nonconvulsive SE and (3) repeated partial seizure [34]. A more extensive classification is presented in [83], which proposes to classify the seizures in terms of two physical features, (1) the presence or absence of prominent motor symptoms and (2) the degree (qualitative or quantitative) of impaired consciousness. The SE with prominent motor symptoms and consciousness impairment can be defined as convulsive SE (CSE) as opposed to nonconvulsive SE (NCES).

CSE and NCSE are associated with severe irreversible brain damage and poor outcome. Delaying the diagnosis, and in consequence, the treatment may increase the mortality rate [38]. The prompt recognition of the seizure activity will determine the treatment effectiveness; the faster the diagnosis is made and the shorter the duration of the seizures, the more effective the treatment and the better the outcome [38]. The CSE recognition is evident given the recurrent overt convulsive seizures that characterize it. On the other hand, recognizing NCES is harder given the lack of physical signs. However, if CSE is untreated or inadequately treated, it progresses from overt convulsions to nonconvulsive [82].

NCES was initially described in ambulatory patients with chronic epilepsy, but it becomes frequently recognized in critically ill and intensive care unit (ICU) patients. Seizures are very common in the ICU, especially in patients with acute or chronic brain injuries. Most patients diagnosed as having NCSE come from this medical setting [77]. The prevalence of seizures in patients with undergoing continuous EEG monitoring ranges from 8% in patients without prior seizures history to 48% after convulsive status epilepticus to more than 50% in comatose patients from a neurological ICU [38]. Symptoms of NCSE vary and can be deceptive, as they can appear in other epileptic disorders different from nonconvulsive epileptic seizures (NCES). It has been shown from several studies that the majority of seizures in the critically ill are nonconvulsive and can therefore only be diagnosed via electroencephalography (EEG) [77]. Hence, the NCES diagnosis is based fundamentally on the recognition of ongoing nonconvulsive seizures in the EEG, and clinical and electroencephalographical responses to antiepileptic drugs (AEDs).

2.4 Electroencephalography (EEG)

The electroencephalogram (EEG) is the recording of the electrical potentials by multiple electrodes (placed on the surface of the scalp) from the neurons of the most external layer of the cerebral cortex. The electrical activity of

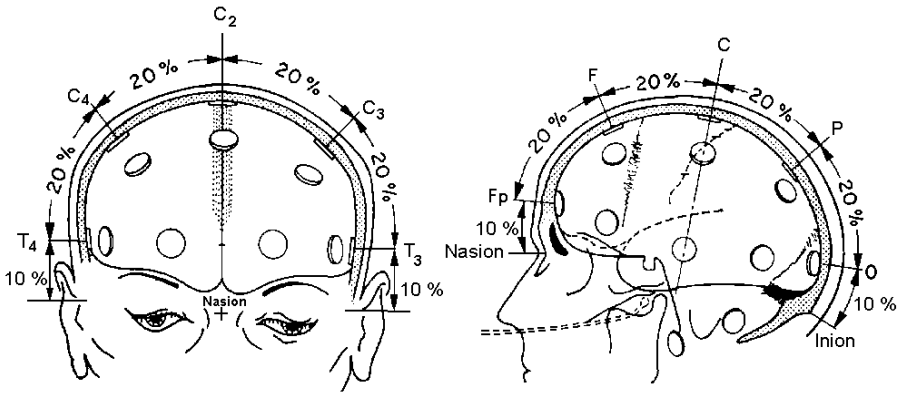


Figure 2.3: Electrode placement according to the International 10-20 system [44].

a single neuron is too small to be measured from the scalp surface. Hence, the electrical activity recorded by the EEG is an attenuated measure of the summed extracellular potentials of many neurons discharging synchronously. It is assumed that each EEG electrode could record the summed activity of roughly $6cm^2$ of underlying cortex [11].

The EEG electrodes are usually placed on the scalp according to some positioning standards, one of them is the 10-20 system (Fig. 2.3). The 10-20 system is a standardized method, where the location on the scalp determines the name of the electrodes (frontopolar (FP), frontal (F), central (C), parietal (P), temporal (T), occipital (O), and auricular (A). Additionally, the electrodes over the left hemisphere are numbered with odd numbers; even numbers are given to the electrodes placed over the right hemisphere, and 'z' is referring to the midline. The order in which the EEG data is displayed is called the montage. Montages may be bipolar or referential. Bipolar montages involve a comparison of voltages recorded from (usually adjacent) active electrodes in a chain-like fashion, (e.g., front-to-back or side-to-side). By contrast, referential recordings involve a comparison of each electrode to an (ideally) inactive electrode. Examples of reference sites include the ear, the mastoid, the vertex, or an average of many active leads ("average reference") [11].

The human EEG is characterized by the presence of wave trains with different frequencies, phases, and amplitudes. The signal shows frequency activity over the range of 1 to 30 Hz, with amplitudes spanned from 20 to 300 μV

[71]. The EEG correlates clinically with the brain function, dysfunction, and diseases. Traditionally EEG waves are classified according to the dominant frequencies measured in Hz. The relationship between EEG rhythms and physiological changes of brain states (alertness vs. sleep, sleep phases, coma and epileptic seizures), as well as the relationship between EEG rhythms and the pharmacodynamic effects, have already been well established and analyzed [71]. There are five frequency bands of clinical interest: Delta band ($0.1 - 3.5$ Hz), theta band ($4 - 7.5$ Hz), the alpha band ($8 - 13$ Hz), the beta band ($14 - 25$ Hz) and gamma band (> 25 Hz) [71].

The synchrony registered by the EEG may be physiological, as seen in the alpha rhythm over the posterior channels. However, when the cortex becomes excessively synchronized, some pathological EEG morphologies associated to epilepsy disorders may be detected (e.g., spikes and sharp waves).

2.5 NCSE monitoring

The role of the EEG in the diagnosis and classification of epileptic disorders and seizure types is well-established [11, 71, 51, 79]. The most common use of EEG monitoring is to determine whether a patient with abnormal synchronous brain activity is having epileptic seizures. The interictal EEG plays a primary role in providing a definitive diagnosis of epilepsy when seizures are recorded [80]. Furthermore, the EEG may indicate electrographic evidence of seizures even if these are nonconvulsive.

Monitoring SE during treatment as well as monitoring EEG in intensive care, critical care, or emergency settings may provide clinical evidence relative to the presence, amount, and response to treatment when seizures or status epilepticus are considered [80]. Several studies have found that following seizures or SE, especially with mental status impairment, many patients still have NCES or NCSE, even after being treated [16].

There are no specific EEG patterns representative of NCES or NCSE that allow to differentiate them as convulsive or nonconvulsive. EEG patterns encountered in NCES (e.g., triphasic waves) can be recorded in patients with various epileptic disorders. This non-specificity creates confusion for diagnosing NCSE, thus hindering this task. The main EEG patterns for NCSE can be summarized as periodic discharges (PD) Fig. 2.4, rhythmic delta activity (RDA), Fig. 2.5, and spike-and-wave/sharp-and-wave (plus subtypes),Fig. 2.6 and 2.7. Into these patterns, the morphologic characteristics such as the sharpness, the number of phases, the triphasic morphology, absolute and relative amplitude, and the polarity can be analyzed. Likewise, it is important to examine the seizure



Figure 2.4: ICU patient with confirmed NCES presenting periodic discharges



Figure 2.5: Rhythmic Delta activity recorded from an ICU patient with altered mental status and having NCES.

location, determine whether the seizure is lateralized, bilateral, multifocal or generalized. This will help to establish the type of seizure causing the NCSE [7, 83]. Other parameters relative to time like the pattern dynamics (evolving, fluctuating, or static) as well as the prevalence, frequency, and duration are considered important prognostic factors in NCSE monitoring [74].

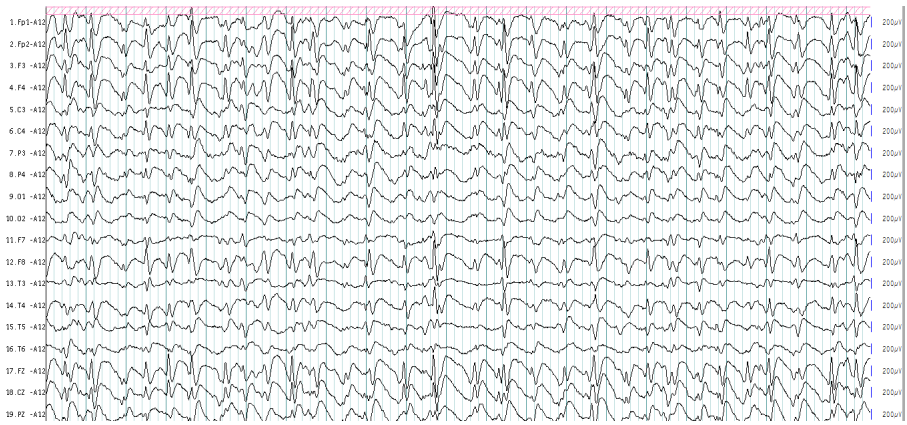


Figure 2.6: Spike-and-wave EEG pattern recorded from an ambulatory patient diagnosed as having absence seizures.

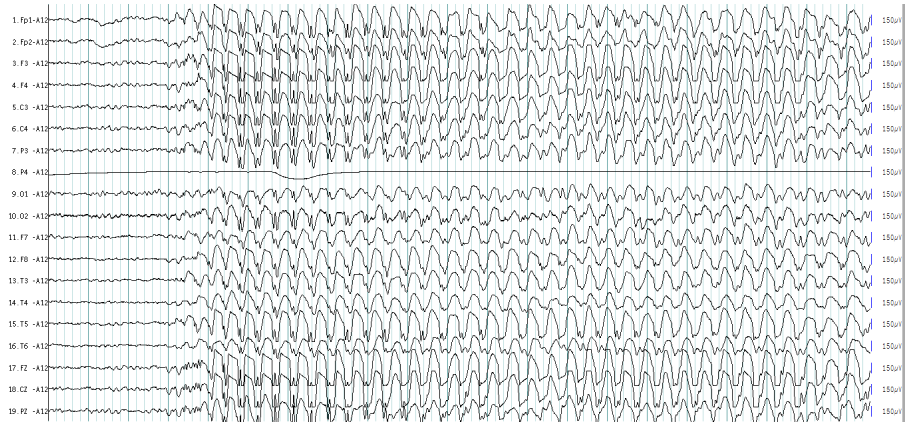


Figure 2.7: Sharp-slow wave EEG pattern recorded from an ambulatory patient having NCES after a head trauma.

The NCSE frequency ranges from $0.2 - 5\ Hz$. The occurrence of generalized or lateralized periodic discharges frequently below 2.5 Hz (about 0.3 Hz) or rhythmic discharges greater than 0.5 Hz (ranging from $0.2 - 3\ Hz$, usually at $0.5 - 2\ Hz$) are considered as activity preceding NCSE [47, 84, 53]. The occurrence of rhythmic delta-theta activity below 4 Hz is also considered a possible sign of NCSE [84]. The NCSE without coma/stupor is characterized by an EEG pattern of spike-and-wave complexes with frequencies from $3Hz$ to 3.5 Hz [47]. This pattern can also be found in patients with coma/stupor at NCSE risk [84]. NCSE in a coma/stupor patient is diagnosed after the emergence of epileptiform discharges above 2.5 Hz [47, 84].

2.6 Automatic Detection of Epileptic Seizures

The introduction of automated spike and seizure detection algorithms has transformed epilepsy monitoring and diagnosis. Originally, long-term EEG monitoring was used to identify the electrographic activity of clinically recognizable seizures. Therefore, for the monitoring to be successful, the events had to have symptoms that could be noticed by the patient or the medical staff. Hence, it was not possible to diagnose events with minor symptoms (e.g., NCES) with the EEG. To detect subtle events, the EEG data was recorded for 24 hours or more. Then, it was manually reviewed by the clinician. The manual review of such amount of data is tedious and time-consuming labor. Usually, a trained technologist performed a first review labeling events of interest for the doctor to complete the final review [16]. Nowadays, a seizure detection algorithm makes the preliminary screening of the data, making the process more time efficient and accurate.

Most of the existing automated seizures detection algorithms, designed for epilepsy monitoring, have been extensively trained and validated on seizures obtained from patients with convulsive seizure disorders, without other brain damage. However, seizures from brain-injured patients are considerably different. Ictal activity from comatose or brain-injured patients is often less organized, slower, with longer duration and without clear on- and off-set [71].

Many efforts have been made in the seizure detection field [86, 60], however, very few of them for NCES detection [5]. Orosco et al. in [60] review 36 seizure detection algorithms, published until 2013. In this study, Orosco compares the algorithms taking into account the processing technique and classifiers used, pointing out the Wavelet Transform and the Entropy as the most popular processing techniques among the seizure detectors. Regarding the classifiers,

Artificial Neural Networks and Support Vector Machines were the more frequent ones.

2.7 Automatic Detection of Nonconvulsive Epileptic Seizures

As stated before, the amount of methods proposed for NCES detection is small compared to those proposed for CES. The first technical solution proposed for NCES detection was presented in 2003 by Adeli et al. [3]. In this study, Adeli used wavelet analysis (WA) to analyze the EEG of a patient with absence seizures to detect the spike wave events. No statistical results are reported from the experiment. Yet, the usefulness of the WA for absence seizure detection was stated. Jacquin et al. in [42] developed an algorithm to detect the spike wave pattern of the absence seizure. In his proposal, Jacquin besides WA, uses the fractal dimension (FD) as features for the NCES detection. The data used in this study consisted of four different data sets: 93 recordings from NYU Epilepsy Center, 79 from a private practice from Virginia, a database of 90 instances of expert-confirmed seizure segments of type spike-wave or polyspike-wave and a database of 101 EEG recordings of 15 minutes' duration from their clinical research sites. The algorithm divided the EEG into epochs of 2.56 s duration, further subdivided into eight non-overlapping data blocks of a duration of 320 ms. A six scales WA was performed to each epoch, and "candidate" spike-wave activity was detected at the first, second and fourth scales by the wavelet coefficients exceeding thresholds. This algorithm results in a large number of false detections. The author reduces the number of false positives by classifying the "candidate" spike-wave activity as true only if the FD in at least one electrode was below a threshold of 1.3 and the moving standard deviation of the FD was below a threshold of 1.5 in all frontal leads. The final method reported a sensitivity of 83% and a specificity of 96%.

Other methods to detect absence seizures based on WA are introduced in [96] and [50]. Xanthopoulos et al. [96] apply the continuous wavelet transform (CWT) to segments of 200 samples of raw EEG. The variance profile and the sample variance are estimated from the wavelet scales corresponding to the interesting frequencies (2.5-4.5 Hz) of each channel. Then, two thresholds are applied, one for seizure detection (maximum variance value during a seizure), and another for artifact rejection (variance threshold between the seizure on-set and off-set). The algorithm is tested on 100 absence seizures recorded from six children (two seizure free). The algorithm detects 97.25% of the clinically significant absence seizures with minimal false positive detections. On the other hand Kollialil et al. [50] proposed to perform absence seizure detection

using a multi-class Support Vector Machine (SVM) and a single feature vector computed from the fifth detail wavelet coefficients. Features as energy, mean energy, entropy, mean cross-correlation, mean curve length, the coefficient of variation, interquartile range (IQR) and median absolute deviation (MAD) are compared to obtain an optimal single feature. The data for this experiment consisted of normal, epileptic and interictal EEG data from 100 subjects, from a reputed Neurology Clinic. The best performances were found for the energy, entropy, MAD, and IQR with an accuracy above 95%. The best feature was the IQR with an accuracy value of 99.66%.

Another approach to absence seizure detection is offered by Liang et al. in [54]. Liang et al. established a comparison of feature combinations of the approximate entropy (ApEn) with: a) multiband spectral power, b) parameters from an autoregressive (AR) model. They used two linear classifiers (linear least squares and linear discriminant analysis) and two non-linear classifiers (back propagation neural network and an SVM with radial basis). The method was tested in the EEG signals from three Long Evans rats with spontaneous absence seizure. The ApEn combined with the multiband spectral power and the SVMRB resulted in the best combination, yielding the highest average accuracy value, 97.5%. No results for humans are reported in this study.

Other authors focused their efforts to NCES detection in patients from critical care or ICU settings. As was stated before, the background EEG from unconscious or brain/injured subjects is more difficult to differentiate from the ictal activity. Minasyan et al. in [57] presented an algorithm to diagnose NCSE in UCI patients based on several EEG parameters and ANN. This method performs in two stages: 1) detect and classify 1-sec EEG feature vectors (basic rhythms, paroxysmal events, sleep events and artifacts) and 2) the 1-sec feature vectors are accumulated for 1 minute to compute an EEG state vector (ESV). In the first step, Minasyan et al. built a feature library of 58 parameters clustered in four groups: spectral, time-domain, wavelet and complexity (nonlinear measures). For the second step, the built ESV was used as the entry to a multilayer perceptron network that classified the 1-min window as NCSE, slow brain activity, fast brain activity, burst suppression or artifact. The data used for the test consisted of 21 EEG studies, 16 from ICU or emergency (10 with confirmed NCSE), four routine EEG studies and one from a sleeping subject. The algorithm reported sensitivity and specificity values of 71% and 99% respectively.

In 2013, Khan et al. in [48] used the parameters of a 6th order AR model and a linear classifier to detect NCES. The data used for this study collected at All India Institute of Medical Sciences consisted of 13 seizures of different duration obtained from the EEG data of 5 subjects. The algorithm achieved scores of 86.6% for sensitivity and 96.9% for specificity. The method does not detect the

seizures that do not prevail in more than 50% of the channels.

Sackellares et al. in [69] conducted a study where the NCES detection is based on the EEG signal amplitude variation (to detect significant amplitude changes) and regularity statistics (to detect periodic changes). The algorithm performance is assessed in terms of the sensitivity and the mean false detection rate, obtaining values of 90.4% and 0.66/hour respectively. These results are compared with two other commercial seizure detection software tools. These commercial software tools showed poor performance obtaining sensitivity values of 12.9 and 10.1% and false detection rates of 1.036/hour and 0.013/hour respectively. The methods were tested on 453 seizures obtained from 24 EEG recordings. The obtained results indicated that 1) the currently available commercial detection software does not perform adequately for the NCES in case of ICU patients and 2) the method proposed by Sackellares was suitable to be used in a real clinical setting to assist the doctors in the NCES detection.

Meanwhile, Varshney et al. focused their efforts to detect the NCES onset [88]. This method analyzed the EEG in epochs of 1-second' duration which were further subdivided in smaller windows of 5, 10, 15, 20 and 25 samples. The features are estimated in the smallest windows. The classification is performed by a simple linear classifier. To test the algorithm the same database from [48] is used. The method yielded high sensitivity and specificity values of 90.20% and 92.43% respectively. The reported seizure onset latency was 0.78 seconds. If the seizure did not appear on at least 23% of the EEG channels, it remained undetected.

Sharma et al. proposed a method for NCES detection based on wavelet statistical parameters and a linear classifier [72]. The proposed algorithm analyzed the EEG in epochs of 1-second' duration. Each epoch is denoised using WA applying cubical thresholding. The features are the IQR, the MAD, and the Normalized Covariance. The method performance is evaluated in terms of sensitivity and specificity, reporting values of 100% and 99.3% respectively. The dataset used for testing consisted of 24 seizures recorded from the EEG of 9 subjects in the All India Institute of Medical Sciences. Similarly to the method proposed by Khan et al. in [48], Sharma et al. require the seizure activity to be present in at least 50% of the channels. Otherwise, the seizures will be missed.

Fatma et al. [28] propose a reduced complexity NCES detection method based on the modified Wilson amplitude. This algorithm does not use any classifier to discriminate between seizure and normal EEG. Instead, thresholds are used to make the decision. This method reports a median sensitivity and specificity of 100% and 99.21% respectively.

2.8 Conclusion

This chapter reviewed different EEG-based techniques previously proposed in the literature for automatic epileptic seizure detection. Their results, strengths, and drawbacks are reported. Table 2.1 summarizes the reviewed methods.

This review reveals that a large number of published seizure detection methods address CES [86], [60] and only a few of them address NCES [5]. The proposed methods for NCES detection published until 2015 are summarized in the work of Ansari and Sharma [5]. These methods can be split according to their principal aim: those that detect NCES without coma/stupor (i.e. absence seizures) [3, 42, 96, 50, 54] and the ones that perform the NCES detection on ICU patients [57, 48, 69, 72, 88]. All databases used in the studies are private and very dissimilar in size and composition, thereby complicating objective comparison between the diverse methods.

The absence of standardized databases to validate the results and the huge variety of performance metrics used to assess them further complicates the comparison and assessment of the real differences between the seizure detection algorithms. The most widespread measures in the literature are the sensitivity, the specificity and the accuracy (proportion of correctly identified EEG segments). Yet, the use of these measures is far from being standard.

Regarding the feature selection, the most widespread features are Wavelet Transform (WT) scales [3, 42, 96], [72], Entropy [50, 54] and nonlinear parameters [42, 50, 57].

The methodology used in these studies is also very diverse. Classifiers such as SVM, neural networks (NN), and linear discriminant functions (LD) are the most commonly used ones.

The major drawbacks of these methods are related to the arbitrary nature of the selected thresholds for performing the detection. Seizure duration and the number of channels affected by the seizure activity are the most popular criteria for thresholding [96, 48, 88, 72]. Seizures are not detected if they are too short in time or too localized (just a few channels affected by the seizure activity). NCES are heterogeneous across patients, the patterns present in the EEG will depend on the etiology of the seizure. Therefore a threshold which works for one patient will not necessarily work for another. Also, there are more meaningful ways to describe the seizure's spatial localization than the number of channels; that is, the seizure topography.

Table 2.1: Reviewed publications, database characteristics, summary of features, classifiers, performance and drawbacks. N/R: Not reported.

Method	BD			Features	Classification	Performance				Drawbacks
	No. of Patient	No. Seizures	Seizure Type			Specificity(%)	Sensitivity(%)	Accuracy(%)	FPR(/h)	
Jacquin. et al. [42]	180	90	absence seizure	WA, Fractal dimension	Thresholds	96	83	89.5	N/R	N/R
Xhathopoulos et al. [96]	6	100	absence seizure	WA, Variance	Thresholds	N/R	97.25	N/R	N/R	Seizures lasting less than 3 seconds are not detected by this algorithm. The sensitivity is related to the length of the analysis window. Smaller window length lead to higher sensitivity and higher chance for false positive detection whereas larger window lengths result in low resolution and lower chance for false positive detection.
Kollialil et al. [50]	300	100	absence seizure	IQR	SVM	N/R	N/R	96.66	N/R	N/R
Liang et al. [54]	3(rats)	1758 (segments)	absence seizure	approximate entropy, spectral power	SVM	N/R	N/R	97.5	N/R	No results reported for humans
Minasyan et al. [57]	21	422 (segments)	NCES	spectral,time domain, WA, nonlinear measures	ANN	99	71	85	N/R	N/R
Khan et al. [48]	5	13	NCES	AR parameters	LD	96.90	86.60	91.85	N/R	Seizure should be present in at least 50% of the channels for the final detection.
Sackellares et al. [69]	24	453	NCES	amplitude, regularity	Thresholds	N/R	90.4	N/R	0.66	N/R
Varshney et al. [88]	5	13	NCES	average absolute difference of adjacent samples	LD	92.43	90.2	N/R	N/R	If the seizure is not spread over 23% of the EEG channels, it is not detected
Sharma et al. [72]	9	24	NCES	IQR, MAD, covariance	LD	99.3	100	N/R	N/R	If the seizure is not spread over 50% of the channels it is not detected
Fatma et al. [28]	9	23	NCES	Wilson amplitude	Thresholds	88.02	100	94.01	N/R	The threshold values selected show poor generalization capability. (Experimental results)

Chapter 3

Feature Selection Techniques and Machine Learning

3.1 Introduction

The previous chapter reviewed state of the art in automated detection of epileptic seizures. This chapter describes several techniques and methods used for NCES detection during the research conducted in this thesis. After a brief review of the Continuous Wavelet Transform (WT), and the Hilbert Huang Transform (HHT) some feature selection methods, blind source separation and machine learning techniques are introduced. The performance metrics used to assess the methods developed in this work are described. Finally, the EEG data used in the training and testing of the algorithms is explained.

3.2 Signal Analysis Techniques

3.2.1 Continuous Wavelet Transform (CWT)

The Continuous WT (CWT) performs a multi-resolution analysis of a signal x(t) by decomposing it into sub-band signals representing activity at different time scales. The WT has been reported in the literature as one of the preferred techniques by the authors to perform epileptic seizure detection [60]. This signal decomposition is defined as follows,

$$x_w(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{3.1}$$

in which the signal x(t) is convolved with scaled and shifted versions of the mother wavelet. $()^*$ indicates the complex conjugate, ψ is the analyzing wavelet, a (> 0) is the scale parameter and b is the shift parameter [36].

To expand the EEG into a three-way tensor expressed as $X \in \mathbb{R}^{(F \times T \times Ch)}$ the CWT is applied to each channel. The CWT transforms each EEG channel in a matrix, where the rows are the wavelet scales (F), the columns time instants (T) and the elements are coefficients. The Ch dimension represents the spatial distribution among the channels. As mother wavelet a "Mexican Hat" (see Fig. 3.1) is used as reported in [2, 40]. Fig. 3.2 illustrates the tensorization process with CWT.

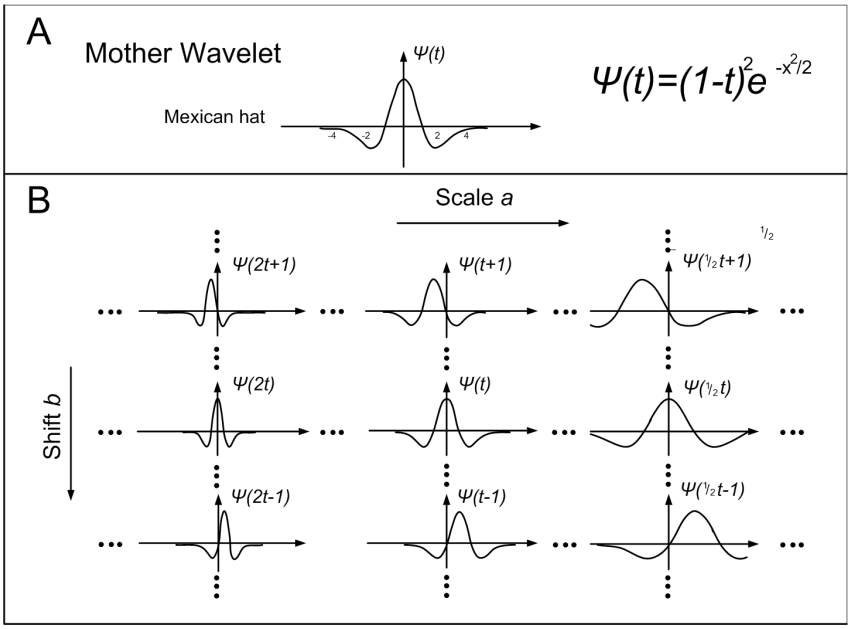


Figure 3.1: a)Mexican hat mother wavelet definition and representation. b) Examples of scaling and shifting of the Mexican hat wavelet. Each row corresponds to a scaling value a and each column corresponds to a shift value b.

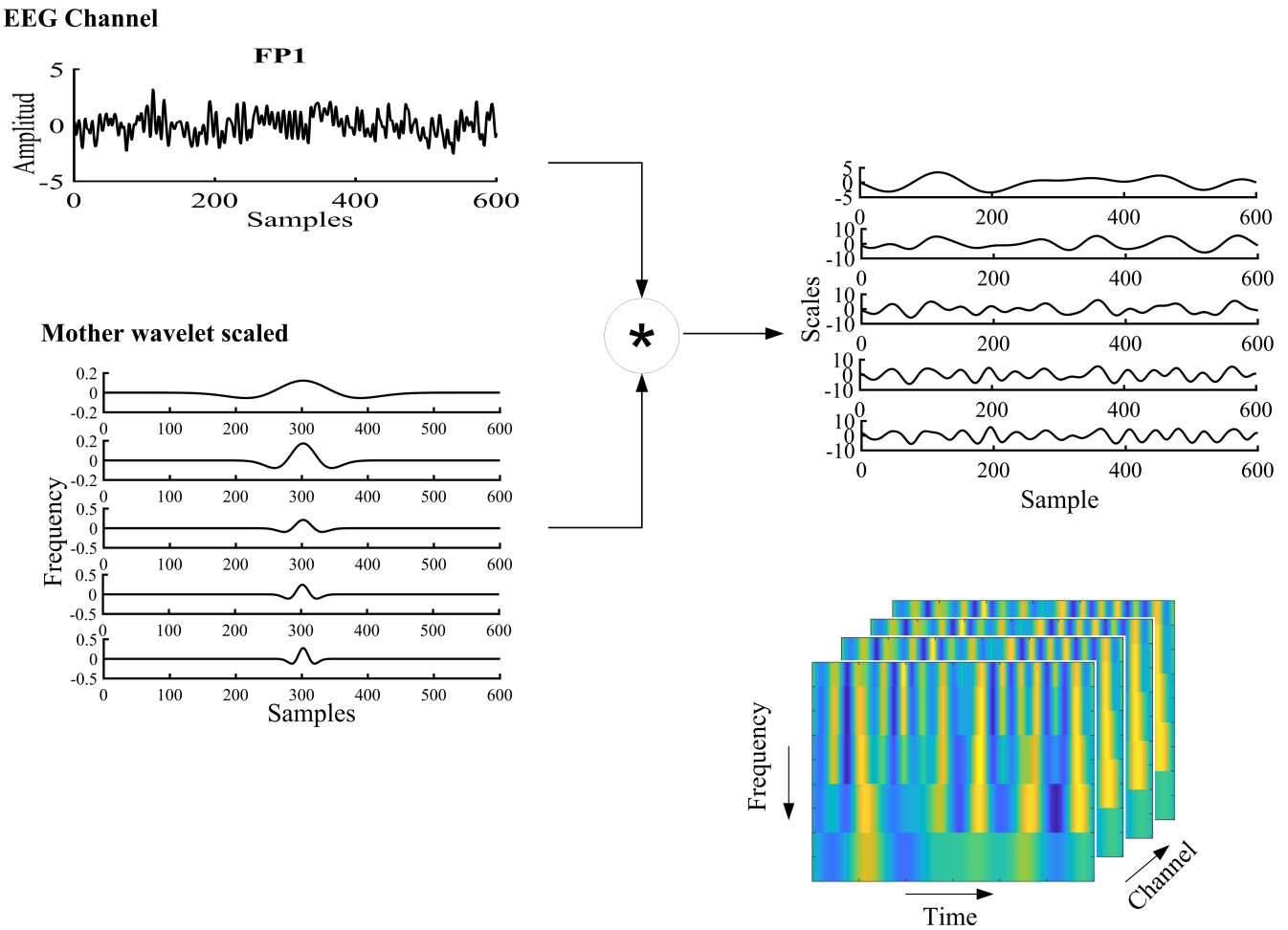


Figure 3.2: EEG tensorization process with CWT.

3.2.2 Hilbert Huang Transform (HHT)

The HHT is a method that uses two steps to analyze the data. First, the data are decomposed into a number of components and a residual using the Empirical Mode Decomposition (EMD). Second, the Hilbert transform is applied to those components to construct a $frequency \times time \times energy$ distribution, designated as the Hilbert spectrum. In this spectrum, all events in the time domain will be preserved by the instantaneous frequency computed by the Hilbert transform [39].

The main characteristic of EMD is to decompose a signal into a sum of so-called Intrinsic Mode Functions (IMF) plus a residue which is conventionally defined as the temporal trend of the series. By definition, an IMF satisfies two conditions: (i) for the entire data set, the number of extrema and the number of zero crossings must be either equal or differ at most by one and (ii) the mean at any point of the contour defined by interpolating the local maxima (upper envelope) and the contour defined by interpolating the local minima (lower envelope) is zero [17]. For any one-dimensional discrete signal, EMD can be presented with the following equation,

$$x(t) = \sum_{k=1}^{n} imf_k(t) + r(t) \tag{3.2}$$

where n is the number of IMFs, imf_k is the kth IMF of the signal, and r is the residue [39]. Fig. 3.3 illustrate the decomposition of an EEG segment.

Once the IMFs are computed each of them is transformed with the Hilbert Transform defined as,

$$y_{imf_k}(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{imf_k(t)}{t - t_0} dt \tag{3.3}$$

where P indicates the Cauchy principal value. $y_{imf_k}(t)$ and $imf_k(t)$ form a complex conjugate pair that describes an analytic signal $z_{imf_k}(t)$,

$$z_{imf_k}(t) = imf_k(t) + jy_{imf_k}(t) = a_{imf_k}(t)e^{j\theta_{imf_k}(t)} \tag{3.4}$$

where a_{imf_k} is the instantaneous amplitude and θ_{imf_k} is the instantaneous phase defined as,

$$a_{imf_k}(t) = \sqrt{[imf_k(t)]^2 + [y_{imf_k}(t)]^2} \tag{3.5}$$

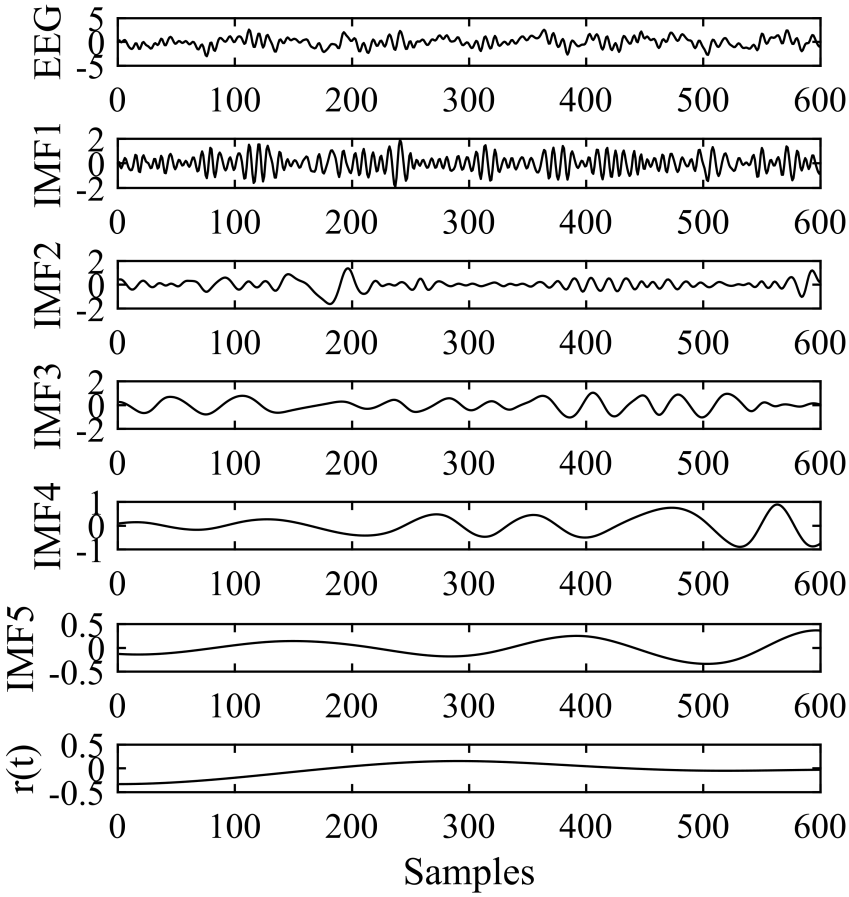


Figure 3.3: EMD of an EEG segment. The first plot displays the raw EEG. The plots from 2 to 5 display the IMFs obtained from the EEG. Finally the last plot show series tendency or residue $r(t)$.

and

$$\theta_{imf_k}(t) = \arctan\left(\frac{y_{imf_k}(t)}{imf_k(t)}\right) \tag{3.6}$$

respectively.

Then the instantaneous frequency is computed as

$$w_{imf_k}(t) = \frac{d\theta_{imf_k}(t)}{dt} \tag{3.7}$$

The $x_{HH}(w,t)$ matrices are now constructed, representing the energy at each time point t and at each instantaneous frequency w in the following way,

$$x_{HH}(w,t) = \sum_{\substack{k=1 \\ w=w_{imf_k}(t)}}^{n} s_{imf_k}(t) \tag{3.8}$$

where n is the number of IMFs and the energy s_{imf_k} is defined as

$$s_{imf_k}(t) = a_{imf_k}(t)^2 \tag{3.9}$$

To expand the EEG into a three-way tensor, the HHT is applied to every EEG channel. The $x_{HH}(w,t)$ matrices are arranged as frontal slices of a tensor $X \in \mathbb{R}^{(F \times T \times Ch)}$, where F indicates frequency values, T the number of time samples and Ch the number of EEG channels. Fig. 3.4 shows the EEG tensorization process with HHT.

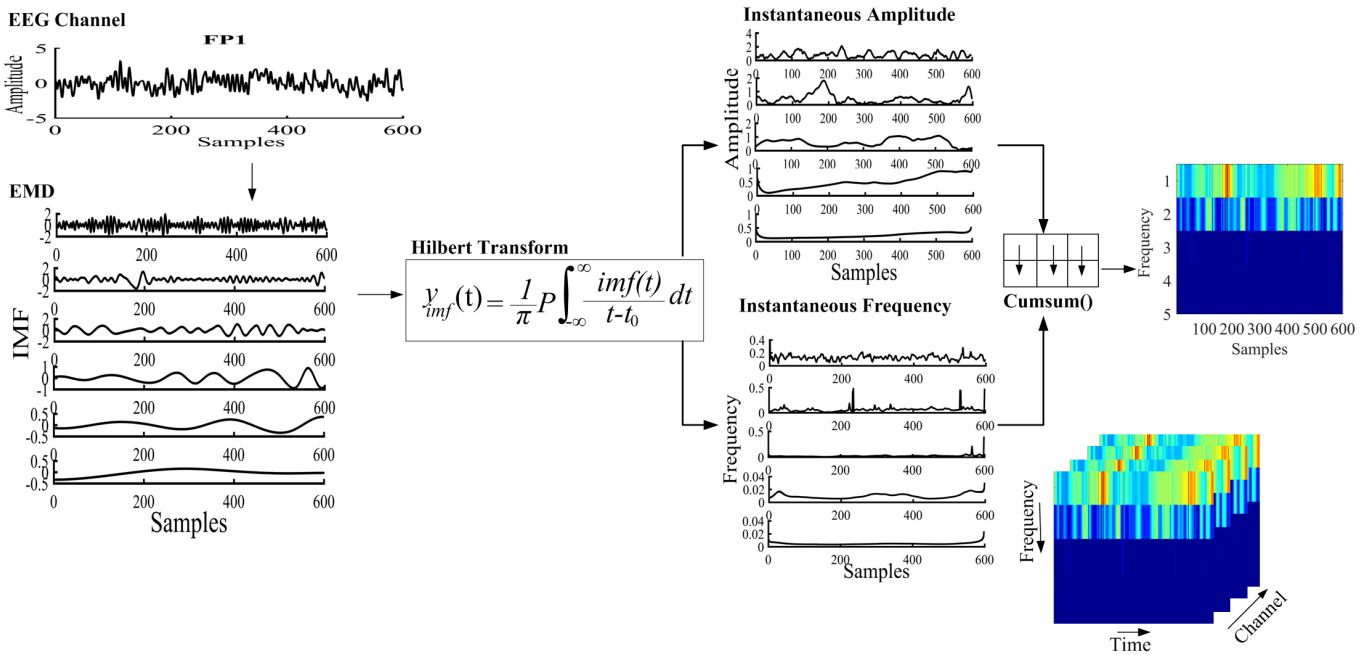


Figure 3.4: EEG tensorization process with HHT.

3.3 Blind Source Separation Techniques

3.3.1 Principal Component Analysis (PCA)

Principal component analysis (PCA) breaks down a matrix X of rank R in a sum of rank-one matrices (see Fig. 3.5) which can be written as the outer product (represented as '$\circ$') of two vectors, a score vector p_r, and a loading vector q_r for $r = 1, \ldots, R$. The outer product of two vectors is equivalent to a matrix multiplication $p_r q_r^T$ (provided p_r is a column vector and q_r^T is a row vector). Then PCA can be written in the following way,

$$X = p_1 \circ q_1^T + p_2 \circ q_2^T + \cdots + p_R \circ q_R^T \tag{3.10}$$

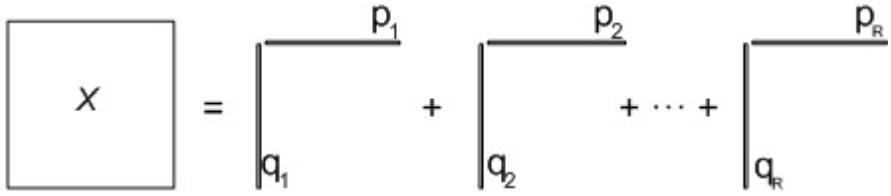


Figure 3.5: PCA of a matrix X in R rank-1 matrices.

or equivalently:

$$X = PQ^T \tag{3.11}$$

Here P is composed by the p vectors as columns and Q^T by the q^T vectors as rows. To obtain P and Q the matrix X is decomposed by means of the singular value decomposition (see [90] for more details about the singular value decomposition (SVD)) as follows:

$$X = U\Sigma V^T \tag{3.12}$$

with

$$U^T U = V^T V = I \tag{3.13}$$

where U and V are the matrices of the left and right singular vectors respectively, and Σ is a diagonal matrix with the singular values as diagonal elements. The variance explained by each singular vector corresponds to the squared magnitude of its corresponding singular value. The singular values and corresponding singular vectors are ordered in decreasing order according to the amount of variance of X that explains each one. From this decomposition we have $P = U\Sigma$ and $Q^T = V^T$ [68].

3.3.2 Canonical Polyadic Decomposition

The Canonical Polyadic Decomposition (CPD) represents an Nth-order tensor $X \in \mathbb{R}^{(I_1 \times I_2 \times I_3 \times \cdots \times I_N)}$ as the outer product of rank-1 tensors in the following way,

$$X \approx \sum_{r=1}^{R} d_r^{(1)} \circ d_r^{(2)} \circ \cdots \circ d_r^{(N)} \tag{3.14}$$

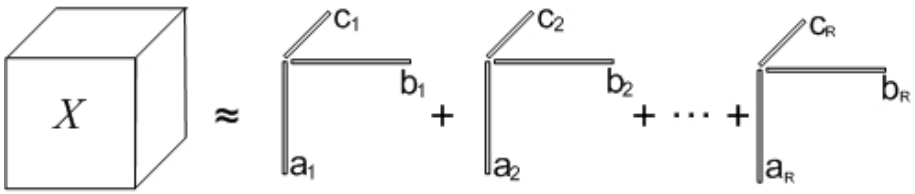


Figure 3.6: CPD of a third order ($N = 3$) tensor X in R rank-1 terms were $a_r = d_r^{(1)}$, $b_r = d_r^{(2)}$, $c_r = d_r^{(3)}$, $r = 1, \ldots, R$.

where R is a positive integer and $D^{(n)} \in \mathbb{R}^{I_n \times R}$ is the factor matrix whose elements $d_r^{(1)} \in \mathbb{R}^{I_1}, d_r^{(2)} \in \mathbb{R}^{I_2}, \ldots, d_r^{(N)} \in \mathbb{R}^{I_N}$ are nonzero vectors that define the mode-n signature, $n = 1, \ldots, N$. The outer product of two vectors $d_r^{(1)} \circ d_r^{(2)}$, is the matrix $d_r^{(1)} {d_r^{(2)}}^T$ whose i, j entry is the product ${d_r^{(1)}}_i {d_r^{(2)}}_j$.

The outer product X of the vectors $a_r \in \mathbb{R}^I$, $b_r \in \mathbb{R}^J$ and $c_r \in \mathbb{R}^K$ is defined by $x_{ijk} = a_i b_j c_k$, $\forall i, j, k$.

The tensor rank of X is equal to the minimal number of R components that generates an 'exact' CPD of X, where 'exact' means that the equality in Eq. (3.14) holds [76]. A tensor CPD is unique under mild conditions up to permutation and scaling [76]. Let $k_{D^{(n)}}$ be the *k-rank* of the factor matrix $D^{(n)}$ defined as the maximum value $k_{D^{(n)}}$ ensuring that any subset of $k_{D^{(n)}}$ columns of $D^{(n)}$ are linearly independent, with $k_{D^{(n)}} \leq r_{D^{(n)}}$ ($r_{D^{(n)}}$ rank of the factor matrix $D^{(n)}$). According to Kruskal's theorem [73] a sufficient condition for the uniqueness of CPD in a N-way tensor is given by,

$$\sum_{n=1}^{N} k_{D^{(n)}} \geq 2R + (N-1) \tag{3.15}$$

As an example, the CPD of a third-order tensor $X \in \mathbb{R}^{(F \times T \times Ch)}$ is defined as,

$$X \approx \sum_{r=1}^{R} a_r \circ b_r \circ c_r \tag{3.16}$$

where $a_r \in \mathbb{R}^F$,$b_r \in \mathbb{R}^T$, and $c_r \in \mathbb{R}^{Ch}$ are nonzero vectors that define the mode-n signature, $n = 1, \ldots, N$ (for $N = 3$) with $1 \leq r \leq R$. The uniqueness of the CPD of X up to permutation and scaling uncertainty is expressed as,

$$k_A + k_B + k_C \geq 2R + 2 \tag{3.17}$$

where $A = [a_1 \ldots a_R]$, $B = [b_1 \ldots b_R]$ and $C = [c_1 \ldots c_R]$ are the factor matrices corresponding to each *mode-n*, and k_A, k_B, and k_C their respective *k-rank*. Fig. 3.6 illustrates this CPD.

3.3.3 Multilinear Singular Value Decomposition(MLSVD) and Low Multilinear Rank Approximation(LMLRA)

The Multilinear Singular Value Decomposition(MLSVD) of a third-order tensor $X \in \mathbb{R}^{(F \times T \times Ch)}$ can be written as,

$$X = Z \bullet_1 U^{(1)} \bullet_2 U^{(2)} \bullet_3 U^{(3)} \tag{3.18}$$

where $U^{(n)}$, $n = 1, \ldots, N$, $N = 3$, are the orthonormal basis for the n different subspaces of the *mode-n* vectors (the column vectors of the $U^{(n)}$ matrices) and $Z \in \mathbb{R}^{(L \times M \times O)}$ is a tensor all-orthogonal (all slices of order $N-1$ are mutually

orthogonal to each other) and ordered (the elements $z_{l-1,m-1,o-1} > z_{lmo}$ $\forall$ $l = 1, \ldots, L$; $m = 1, \ldots, M$; $o = 1, \ldots, O$) [18]. And $\bullet_n$ denotes the *mode-n* product of the tensor Z with the matrices $U^{(n)}$, e.g. the *mode-n* product of the tensor Z with the matrices $U^{(1)} \in \mathbb{R}^{F \times L}$ results in a tensor whose entries are [21],

$$(Z \bullet_1 U^{(1)})_{fmo} = \sum_{l=1}^{L} z_{lmo} u_{fl} \tag{3.19}$$

Essentially, the MLSVD considers a given tensor as N sets of *mode-n* vectors and computes the matrix SVD of these sets. The truncation of the MLSVD gives a suboptimal solution to the tensor best low *ml_rank* approximation problem which can be used as initialization to iterative algorithms such as the low multilinear rank approximation (LMLRA). The multilinear rank (*ml_rank*) of a high-order tensor is the *n-tuple* of the *mode-n* ranks. It is important to notice that the *ml_rank* is not the same as the tensor rank since the rank of each *mode-n* can be different (for the matrix case both definitions coincide). The LMLRA can be formulated as MLSVD, the difference between these two algorithms lies in the computation and the optimality of the decomposition. LMLRA refines the MLSVD initial guess using an optimization based method (e.g. nonlinear unconstrained optimization, nonlinear least squares, adaptive cross approximation) [18].

3.3.4 (L,M,O)- Block Term Decomposition

BTD is a generalization of CPD and MLSVD, and decomposes a tensor in a sum of subtensors of lower rank. The (L,M,O)-BTD of a third-order tensor $X \in \mathbb{R}^{(F \times T \times Ch)}$ is written as,

$$X = \sum_{r=1}^{R} D_r \bullet_1 A_r \bullet_2 B_r \bullet_3 C_r \tag{3.20}$$

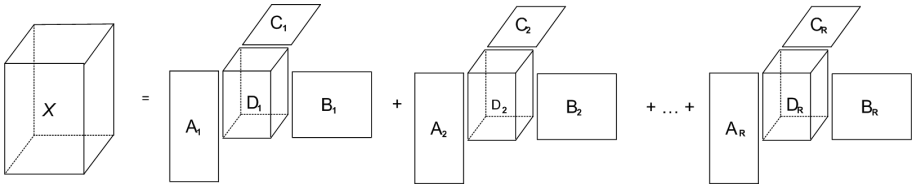


Figure 3.7: BTD of a tensor X in R (L,M,O)-rank terms.

where $\bullet_n$ denotes the mode-n product, $D_r \in \mathbb{R}^{(L\times M\times O)}$ are full rank tensors, $A_r \in \mathbb{R}^{(F\times L)}$ (with $F \geq L$), $B_r \in \mathbb{R}^{(T\times M)}$ (with $T \geq M$), and $C_r \in \mathbb{R}^{(Ch\times O)}$ (with $Ch \geq O$) are factor matrices of full column rank, $1 \leq r \leq R$ (see Fig. 3.7). The mode-n product of the tensor $D_r \in \mathbb{R}^{(L\times M\times O)}$ with the matrices $A_r \in \mathbb{R}^{(F\times L)}$, $B_r \in \mathbb{R}^{(T\times M)}$, and $C_r \in \mathbb{R}^{(Ch\times O)}$ is defined by [21],

$$(D_r \bullet_1 A_r)_{fmo} = \sum_{l=1}^{L} d_{r_{lmo}} a_{r_{fl}} \quad \forall f, m, o \tag{3.21}$$

$$(D_r \bullet_2 B_r)_{lto} = \sum_{m=1}^{M} d_{r_{lmo}} b_{r_{tm}} \quad \forall l, t, o \tag{3.22}$$

$$(D_r \bullet_3 C_r)_{lmch} = \sum_{o=1}^{O} d_{r_{lmo}} c_{r_{cho}} \quad \forall l, m, ch \tag{3.23}$$

respectively.

BTD is unique up to permutation and scaling inherited from CPD under more restrictive conditions. Essential uniqueness conditions for the (L,M,O)-BTD can be found in [22].

3.4 Feature selection techniques

A feature is a measurable property of an observed process [15]. Every day more and new features become available to modeling processes. However, increase in the number of features brings some disadvantages with it, such as the increase in the number of dimensions of the data, including noise and redundancy in the data. The dimensionality increase limits in the first place the capacity to store the amount of data required to model a process, but it also increases the

computational requirements and decreases the accuracy of the model. As the number of features increase, the amount of data needed for accurate modeling is exponentially larger. If the number of features exceed the data available, it is no longer possible to produce a reliable model. On the other hand, even when we can measure a feature from a process it does not mean that it describes this process. Such features represent noise in the feature set. A noisy feature when added to the process model, increases the classification error on new data. Furthermore, if two features are highly correlated the problem of redundancy in the feature set arises. The fundamental problem with redundancy is that the feature set is then no longer a complete representation of the characteristics of the observed process.

Feature selection reduces the dataset dimensionality by selecting the most relevant features and removing noisy and redundant features. Diminishing the number of features in a dataset decreases the required storage, improve the learning performance, reduces the computational complexity, and allows to built better generalizable models [78]. For the classification problem, feature selection intends to find a subset of features that are capable of discriminating samples that belong to different classes. For example, a feature h_i is supposed to be relevant to a class c_j if h_i and c_j are highly correlated. In this section some feature selection techniques used in this research are described.

Before introducing the feature selection techniques, let us establish the notation used in this section to formulate the feature selection for a binary classification problem. Let $H \in (R)^{d \times N}$ be a feature set for a given set of classes $\{C_j\}_{j=1}^{2}$, where d and N are the number of features and available samples respectively. It is the purpose of a feature selection technique to find a feature subset $\hat{N}$, with $\hat{N} < d$, of the input dataset H that maximizes the generalization capabilities of the specified model.

3.4.1 Modeling and Discrimination Power

The Modeling Power (MP) and Discrimination Power (DP) were proposed by Wold et. al. in [93] for feature selection in chemical applications. Subsequently, these measures have been used for feature selection in diverse applications such as image processing [20], DNA samples selection [13], and biomedical modeling [10, 46].The MP describes the variable relevance for the model and the DP expresses how well a variable discriminates between different classes.

To compute the MP and the DP, at first stage, principal component analysis (PCA) is carried out on the feature matrix H(including the features as its columns). Then, the number of principal components (PC) λ^H, that represents most of the variance of the feature matrix H is selected.

Modeling Power (MP)

After selecting the PC's explaining most of the variance, the MP of the jth column of H (H_j) is computed as [93]:

$$MP_j^H = 1 - \frac{\sqrt{\sum_{i=1}^{n^H} \frac{e_{ij}^{H\,2}}{(n^H - \lambda^H - 1)}}}{\sqrt{\sum_{i=1}^{n^H} \frac{(H_{ij} - \mu_j^H)^2}{n^H - 1}}} \tag{3.24}$$

where μ_j^H is the mean of H_j, and n^H is the number of elements in H_j, e_{ij}^H is the $(i,j)^{th}$ entry of the residual matrix E^H obtained from H as follows:

$$E^H = H - P_{\lambda^H}^H {Q_{\lambda^H}^H}^T \tag{3.25}$$

Where $P_{\lambda^H}^H$ are the λ^H selected PCs and $Q_{\lambda^H}^H$ are the loadings of the selected PC respectively after truncated PCA.

This measure takes values between 0 and 1. An MP value close to 1 implies that the variable is relevant for the class description. If this value is close to 0, then it can be concluded that the variable in terms of modeling is less significant [93].

Discriminating Power (DP)

To compute DP, let us define two classes c_1 and c_2 with feature matrices H_{c_1} and H_{c_2}. H_{c_1} and H_{c_2} are analyzed separately with PCA and the $\lambda^{H_{c_1}}$ and $\lambda^{H_{c_2}}$ PCs that represent most of the variance of the features (for each class) are retained. Then, two PCA models are obtained with $\lambda^{H_{c_1}}$ and $\lambda^{H_{c_2}}$ PCs. The DP between the two classes, c_1 and c_2, of the feature j is defined as,

$$DP_j^{H_{c_1},H_{c_2}} = \sqrt{\frac{S_j^{H_{c_2}}(H_{c_1})^2 + S_j^{H_{c_1}}(H_{c_2})^2}{S_j^{H_{c_2}}(H_{c_2})^2 + S_j^{H_{c_1}}(H_{c_1})^2}} - 1 \tag{3.26}$$

where $S_j^{H_{c_2}}(H_{c_2})^2$ and $S_j^{H_{c_1}}(H_{c_1})^2$ are the variance of the residual matrices $E^{H_{c_2}}$ and $E^{H_{c_1}}$ obtained for the PCA models H_{c_2} and H_{c_1} with (3.25), and are defined as,

$$S_j^{H_{c_2}}(H_{c_2})^2 = \sum_{i=1}^{n^{H_{c_2}}} \frac{{e_{ij}^{H_{c_2}}}^2}{(n^{H_{c_2}} - \lambda^{H_{c_2}} - 1)} \tag{3.27}$$

and

$$S_j^{H_{c_1}}(H_{c_1})^2 = \sum_{i=1}^{n^{H_{c_1}}} \frac{{e_{ij}^{H_{c_1}}}^2}{(n^{H_{c_1}} - \lambda^{H_{c_1}} - 1)} \tag{3.28}$$

where $\lambda^{H_{c_2}}$ and $\lambda^{H_{c_1}}$ are the number of selected PCs for the PCA models H_{c_2} and H_{c_1} respectively. The $S_j^{H_{c_2}}(H_{c_1})^2$ and $S_j^{H_{c_1}}(H_{c_2})^2$ terms are the residual variance of H_{c_2} and H_{c_1} when fitting samples from class c_1 onto the PCA model obtained for class c_2 ($\lambda^{H_{c_2}}$) , and when fitting samples from the class c_2 onto the PCA model obtained for class c_1 ($\lambda^{H_{c_1}}$) respectively [93], these terms are defined as,

$$S_j^{H_{c_1}}(H_{c_2})^2 = \frac{J}{n^{H_{c_2}}(J - \lambda^{H_{c_1}})} \sum_{i=1}^{n^{H_{c_2}}} (e_{ij}^{H_{c_2}})^2 \tag{3.29}$$

and

$$S_j^{H_{c_2}}(H_{c_1})^2 = \frac{J}{n^{H_{c_1}}(J - \lambda^{H_{c_2}})} \sum_{i=1}^{n^{H_{c_1}}} (e_{ij}^{H_{c_1}})^2 \tag{3.30}$$

where J is the number of columns of H_{c_2} and H_{c_1}. $n^{H_{c_1}}$ and $n^{H_{c_2}}$ are the number of rows of H_{c_2} and H_{c_1} respectively.

If the DP value is close to 0, low discriminatory power is observed. If it is much bigger than 1, a good discriminatory power is observed [93].

3.4.2 Maximum relevance and minimum redundancy (MRMR)

The maximum relevance and minimum redundancy (MRMR) [25] is a feature selection method that aims to remove redundancy in the feature dataset while maximizing the relevance of the features in the final feature set. The method selects features with high relevance with respect to a class and that are non-redundant [97]. For continuous data the relevance can be estimated using the

F-statistic to compute the feature correlation with the class in the following way [78],

$$\max_{c} R_F \quad \text{with} \quad R_F = \frac{1}{|\hat{N}|^2} \sum_{i} F(i,c) \tag{3.31}$$

where $\hat{N}$ denotes the size of the desired feature subset. c is the target class and $F(i,c)$ is the F-statistic between the feature i and the class c, defined as [98],

$$F_i = \frac{[\sum_{c=1}^{2} n_c(\bar{H}_{.i}^{c} - \bar{H}_{.i})]}{\sigma^2} \tag{3.32}$$

Where $H_{.i}$ denotes the ith column of the feature matrix H, n_c is the number of samples from the cth class. $\bar{H}_{.i}$ is the mean value of the column $H_{.i}$ and $\bar{H}_{.i}^{c}$ is the mean value of the $H_{.i}$ column within the cth class. σ^2 is the pooled variance defined as [98],

$$\sigma^2 = \frac{[\sum_{c}(n_c - 1)\sigma_c^2]}{\hat{N} - 2} \tag{3.33}$$

where σ^2 is the variance of $H_{.i}$ within the class c

The average minimum redundancy is computed as [78],

$$\min_{H \in \hat{N}} R_{coff} \quad \text{with} \quad R_{coff} = \frac{1}{|\hat{N}|^2} \sum_{i,j \in \hat{N}} C(i,j) \tag{3.34}$$

where $C(i,j)$ is the correlation between the ith and j^{th} columns of H.

The features for the final feature set are selected one by one by applying a greedy search to maximize the objective function, based on relevance and redundancy in the following way,

$$\max_{H \in \hat{N}} R \quad \text{with} \quad R = \frac{R_F}{R_{coff}} \tag{3.35}$$

The Algorithm 1 describes the implementation of the MRMR method.

3.4.3 Bhattacharyya Distance

Before proceeding with the classification, a preliminary exploration of how the selected features contribute to the discrimination between the defined classes is

Algorithm 1 Maximum relevance and minimum redundancy feature selection

Input: Feature matrix H, class labels c, number of output features $\hat{N}$, number of features in H is N.
Output: Output feature set F

```
idle_ht = [1 : N]
for i=1:H do
    relevance = Fscore(H(:, i), c)
end for
[R, ind] = max(relevance)
F[1] = ind
idle_ht = idle_ht − F
for i=2:N̂ do
    obj1 = relevance(idle_ht)
    for j=1:|idle_ht| do
        redun(j) = compute_redun(H(:, idle_ht))
    end for
    obj2 = relevance(idle_ht)/redun
    [newind, obj2] = best_subset(obj1, obj2, idle_ht)
    [R, id] = max(obj2)
    F[i] = id
    idle_ht = idle_ht − F
end for
```

needed. Hence, to quantify how separable the two given classes are, a distance measure is required. For such exploration, the Bhattacharyya distance is used in Section 5.4.

The Bhattacharyya distance can be defined taking into account its close relationship with the Bhattacharyya coefficient which quantifies the closeness between two distributions. In case of two identical histograms the Bhattacharyya coefficient value is one ($\sum_{i=1}^{B} \sqrt{(P(b_i)P'(b'_i))} = 1$). Then, the Bhattacharyya distance for two normalized distributions P and P' is defined as [26],

$$d^{bt} = \sqrt{1 - \sum_{i=1}^{B} \sqrt{P(b_i)P'(b'_i)}} \tag{3.36}$$

Where b_i and b'_i are the frequencies coded at the ith bin [4]. The Bhattacharyya distance takes values between 0 and 1. A distance value close to zero means that the distributions are similar, and a distance value close to one indicates that the distributions are apart.

3.5 Machine Learning Techniques

This section introduces the classifiers used in this research. For better understanding of reading, we must clarify that the classifiers are estimated using a supervised learning approach. The classification techniques introduced in this section are described to solve a binary classification problem. However, all of them can be extended to solve multi-class problems.

For supervised learning, the final objective is that the classifier learns from a labeled training set of entries $\{x_i\}_{i=1}^{N}$ and class labels $\{y_i\}_{i=1}^{N} \in \{\pm 1\}$, where N is the number of entries. The class labels y_i are assumed to be unordered (categorical) and mutually exclusive. This kind of learning consists of two stages:

- Training, in which the model is estimated from a subset N_s of the N pairs of samples in the form:$\{x_i, y_i\}_{i=1}^{N_s}$.
- Validation: stage in which the generalization ability of the model created during the training using samples different from those used in training is evaluated.

After the model estimation, its performance is tested using a set of labeled samples different from the used ones during the training and validation process called test set. The classifier makes an error when assigning a sample to a wrong class.

3.5.1 Linear Discriminant Analysis (LDA)

Given a set of N training samples $\{x_i\}_{i=1}^{N}$ labeled in $(\{c_j\}_{j=1}^{2})$ classes, LDA finds a linear transformation ϕ over x_i, for which the variance within the class c of the projected samples x_i is minimized and the variance between classes is maximized [95]. Such hyperplane is defined as the one that maximizes the ratio S_b and S_c, where

$$S_c = \sum_{c=1}^{C} \Sigma_c = \sum_{c=1}^{C} \sum_{x \in c} (x - \bar{x_c})(x - \bar{x_c})^T \tag{3.37}$$

is the within-class scatter matrix, and

$$S_b = \sum_{c=1}^{C} n_c(\bar{x_c} - \bar{x})(\bar{x_c} - \bar{x})^T \tag{3.38}$$

is the between-class scatter matrix [95]. n_c is the number of samples in the class c ($N = \sum_{c=1}^{C} n_c$), and $\bar{x}$ and $\bar{x_c}$ are the mean vector of X and in class c respectively.

The optimal transformation(ϕ) can be obtained as the eigenvector that corresponds to the smallest eigenvalue resulting from the eigenvalue decomposition of the matrix $S_{cb} = S_c^{-1} S_b$ [95]

$$S_{cb}\phi = \lambda\phi \tag{3.39}$$

ϕ is an $n \times n$ matrix whose columns correspond to the discriminant eigenvectors [61].

3.5.2 Support Vector Machine (SVM)

Support Vector Machines (SVM), are one of the most used machine learning techniques for binary and multi-class classification. Given a set of N training vectors $\{x_i\}_{i=1}^{N}$, $x_i \in \mathbb{R}^p$ that falls in C classes ($\{c_j\}_{j=1}^{2}$) with labels $\{y_i\}_{i=1}^{N} \in \{\pm 1\}$, the SVM attempts to infer a model M that correctly approximates all training vectors x_i to their corresponding target y_i based on a function of the form [37].

$$y(x) = \sum_{i=1}^{N} w_i K(x, x_i) + w_0 \tag{3.40}$$

where w_i are the weights and $K(:,:)$ is a kernel function that defines the inner product between the x_i data vectors in a p-dimensional feature space. Many kernel functions are proposed for SVM implementation. In this research we use the radial basis kernel expressed as,

$$\left\{ K(x, x_i) = \exp\left(\frac{\|x - x_i\|^2}{\sigma^2} \right), \quad \forall i \in \{1 \dots N\} \right.$$

where σ is the proximity parameter (if it is not specified $\sigma = 1$).

There might be an infinite number of functions $y(x)$ that correctly assign x_i vectors to their corresponding label. The main feature of SVM is to find the one with lower generalization error: namely, the function $y(x)$ that minimizes the errors made on the training set while maximizing the separation margin between the two classes [37]. These can be done by computing the hyperplane defined by,

$$h(x) = \begin{cases} 1, & w_s K(x, x_i) + w_0 \geq 0 \\ -1, & otherwise \end{cases}$$

were $1, -1$ are the given labels for the two defined classes. In a p-dimensional space, a hyperplane is defined as a flat affine subspace of dimension $p-1$. It is understood by affine that the subspace must go through the origin [43]. Finding the w_i, w_0 parameters requires to solve the following optimization problem:

$$\min_{w,\xi_i} \frac{1}{2}\|w\|^2 + \gamma \sum_{i=1}^{N} \xi_i \tag{3.41}$$

subject to,

$$\Big\{ y_i(w^T K(x, x_i) + w_0) \geq 1 - \xi_i, \quad \forall i \in \{1, \ldots, N\}$$

where the parameters ξ_i are known as slack variables which penalize data points which violate the margin requirements, and γ is the regularization parameter, a positive constant that controls the influence of the two competing terms [81]. This problem is a quadratic optimization task subject to a set of linear constraints (the Karush-Kuhn-Tucker (KKT) conditions).

3.5.3 K-Nearest Neighbors(KNN)

The K-Nearest Neighbors classifier (KNN) is a non-probabilistic and non-parametric classifier that estimates how likely an unknown sample x_0 is to be a member of a class $\{c_j\}_{j=1}^2$. Given a positive integer K and an unknown sample x_0, the KNN identifies the K vectors in the training set $T = \{x_i, y_i\}_{i=1}^N$, closest to x_0 and represented by N_0.

The basic KNN algorithm applied to a test sample x_0, first calculates the distance between x_0 and T , $\delta = d(x_0, T)$, for $n = 1, \ldots, N$ (where $d(*)$ indicates a distance measure), then determines the indexes $a_1, \ldots, a_k$ of the K smallest

values of δ (such that $\delta_{a_1} \leq \cdots \leq \delta_{a_k}$). According to the KNN rule, x_0 is assigned to the class c_j represented by a majority of its K nearest neighbors in T, represented as k_{c_j}. k_{c_j} given by [75, 91],

$$k_{c_j} = \sum_{i \in N_0} I(y_i = j) \tag{3.42}$$

where $I(y_i = j)$ is the indicator factor ($I(y_i = j)$ if $y_i = j$, $I(y_i = j) = 0$ otherwise), and y_i is the class label of training vector $x_i \in T$.

The only parameters requiring pre-specification are the number of neighbors, K, the distance metric, and the training dataset. When more than one neighbor $K > 1$ is considered, this rule can produce more than a one winning class (it is possible that the number of neighbors closer to x_0 is the same for both classes), leading to the indeterminacy of the class of x_0. To avoid this kind of issue for two classes it is better to restrict K to odd numbers [23]. Without prior knowledge, the KNN classifier usually uses Euclidean distance as distance metric.

3.5.4 Partial Least Squares discriminant analysis (PLS)

Partial least squares discriminant analysis (PLS) is a linear classification technique that generalizes and combines features from partial least squares regression with the discrimination power of a classification technique. Given the feature matrix $X \in \mathbb{R}^{(s \times p)}$ (with s samples and p features) and its corresponding label vector $y \in \{\pm 1\}$ with size s (both mean-centered), partial least-squares (PLS)' main purpose as in multiple linear regression, is to build a linear model in the following way [12],

$$y = X\beta + \epsilon \tag{3.43}$$

where $\beta \in \mathbb{R}^p$ is the regression coefficient vector, and $\epsilon \in \mathbb{R}^s$ is a noise term for the model.

PLS-DA attempts to disclose the main latent variables in X that permit to predict y and describe their joint structure. If y is a vector and X is full rank this can be done by a classical multi-linear regression method. But when s is smaller than p, X is likely to be singular and the regression approach is no longer feasible. Several strategies are proposed to deal with the singularity of X, one of them is to drop some predictors. For choosing which features to delete, PLS uses principal component regression (PCR).

PCR performs the PCA of X and then uses the resulting PC of X as the independent variables of a multiple regression model to predict y. In contrast to PCA, which decomposes X to find the components that best describe X, PLS decomposes X and y simultaneously to find the components in X that best predict y (latent vectors) [1]. The selected latent vectors must maximize the covariance between X and y. The independent variables are then computed as [12],

$$X = PQ^T \tag{3.44}$$

with

$$P^T P = I \tag{3.45}$$

where I is the identity matrix. By analogy with PCA, P is called the score matrix and Q is called the loading matrix. The y is estimated as

$$\hat{y} = P\beta w^T \tag{3.46}$$

w is the 'weight vector'. A more detailed explanation of how to obtain the w vector, as well as the implementation of PLS-DA, can be found in [1, 12].

3.6 Performance Metrics

As mentioned in Section 3.5, once the classifier has been constructed, its performance is tested in new unknown samples. During this test process, the classifier can commit two types of errors.

- Type I: Classify samples from the negative class $\{-1\}$ as member of the positive class $\{+1\}$ (false positive (FP)).
- Type II: Classify samples from the positive class $\{+1\}$ as member of the negative class $\{-1\}$ (false negative (FN)).

An intuitive and simple manner to measure the impact of these errors in the performance of the classifier is to build a confusion matrix. The confusion matrix has two dimensions "Actual class" which refers to the real class label of the sample) and "Predicted class" which refers to the class label estimated by

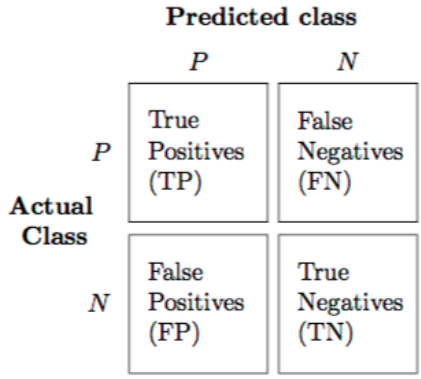


Figure 3.8: General structure to build the confusion matrix from the classifier output.

the classifier for the samples in the test set. Fig. 3.8 shows the general structure to build the confusion matrix.

The confusion matrix is not a performance measure by itself, but there are numerous metrics associated to it. Here, we only introduce those used during the study. The performance of the algorithms proposed in this project is assessed in terms of the sensitivity, specificity, accuracy, false positive rate (FPR) and positive predictive value (PPV). The sensitivity measures the probability that a positive sample is correctly identified as such, and is defined as,

$$Sen = \frac{TP}{TP + FN} * 100\% \tag{3.47}$$

Opposed to this, the specificity measures the probability that a negative sample is correctly classified as such, and is defined as,

$$Spec = \frac{TN}{TN + FP} * 100\% \tag{3.48}$$

where TP is the number of positive samples correctly classified as such, FP is the number of negative samples classified as positives, TN is the number of negative samples correctly classified as such and FN is the number of positive samples classified as negative samples.

The accuracy is the proportion of well classified samples (both negatives and positives) among all samples examined, and is defined as,

$$Acc = \frac{TP + TN}{TP + FP + TN + FN} * 100\% \tag{3.49}$$

The false positive rate denotes the Type I error ratio and is defined as,

$$FPR = \frac{FP}{FP + TN} * 100\% \tag{3.50}$$

And the PPV is the probability that samples classified as positives are truly positive samples, and is termed as,

$$PPV = \frac{TP}{TP + FP} * 100\% \tag{3.51}$$

3.7 Conclusions

This Chapter introduced different techniques that are used throughout this thesis. In Chapter 5 HHT and CWT are used to expand the EEG signal to build tensors from it. The CPD and BTD are applied to analyze the built tensors. The optimal parameters for the CPD and BTD are estimated by means of a combination of the tensors' ml-rank analysis looking at its MLSVD, and MD and DP used to estimate the decomposition quality. To assess the usefulness of the features obtained from the CPD and BTD for the NCES classification the Bhattacharyya distance is measured. An LDA, an SVM, and KNN models are trained to detect NCES. The model's performance is accessed in terms of the specificity, sensitivity, accuracy, and FPR.

Chapter 6 implements an incremental approach of an SVM for NCES detection. In this Chapter, the MRMR is used to select the optimal support vector subset in every training round. The performance of the method is assessed in terms of specificity, sensitivity, and PPV.

Chapter 4

EEG Data

4.1 Introduction

In this Chapter, we describe the database used during this thesis to train and validate the methods here proposed. The recordings come from two different medical settings, and two different hospital facilities. The main part of the data come from the ambulatory clinical Epilepsy Unit of the Cuban Neurological Restoration Center (CIREN). The rest of the recordings were obtained at ICU of the Clinical Surgical Hospital Hermanos Ameijeiras. Both hospitals are located in Havana City. The NCES data was annotated by three expert clinical neurophysiologists. This Chapter provides insights about the database composition, and how this is handled in Chapters 5 and 6.

4.2 EEG Data description

The database built consists of 14 recordings, 9 from ambulatory and 5 from ICU medical settings. The ambulatory EEG data were collected as part of patients' clinical assessment at the Epilepsy Unit of the CIREN. The ICU data were collected at the ICU of the Clinical Surgical Hospital "Hermanos Ameijeiras". Data were anonymized prior to their use in this study. All procedures were reviewed and approved by the Ethical Committees of the CIREN and "Hermanos Ameijeiras" Hospital respectively.

Since the patients come from different hospitals and areas, the acquisition protocol shows some differences; yet the sampling rate for all recordings is 200

Hz. The number of electrodes used for the recordings varies between 8 and 19. In all cases, the electrodes were placed according to the 10-20 montage system. Fig. 4.1 illustrates the different electrode setups. The use of the 10-20 montage during the recording process guarantees that the electrodes' name and position do not change across the records.

All patients in the study are adults (ages ranging between 18 and 57 years old) with various brain disorders that present NCES. The visual inspection and labeling were performed by three neurophysiologists. A total of 139 seizures have been registered, all of them NCES according to the neurophysiologist' diagnosis (55 with coma/stupor). Figures 4.2 and 4.3 show examples of NCES from ambulatory and ICU settings respectively. A more detailed description of the data is given in Table 4.1.

Some EEG protocols for epilepsy monitoring used in this study, include the simultaneous recording of several physiological variables. The principal reasons for recording other physiological variables than the EEG are to obtain data on the patient behavioral aspects and to differentiate artifacts in the EEG that could resemble ictal activity. The recorded physiological signals apart from the EEG in these recordings are the electromyogram (EMG: electrical activity generated in muscles), the electrooculogram (EOG: electrical potentials generated by the vertical and horizontal eye movements), and the electrocardiogram (ECG: electrical activity produced by the heart).

The main objective when including these other physiological signals in the study was to support the diagnosis of the nonconvulsive nature of the recorded seizures. The three signals mentioned above increase their activity during a seizure. Therefore, if the EEG clearly displays a seizure and the fluctuations recorded in the other physiological parameters do not show significant changes, the nonconvulsive nature of the seizure can be confirmed.

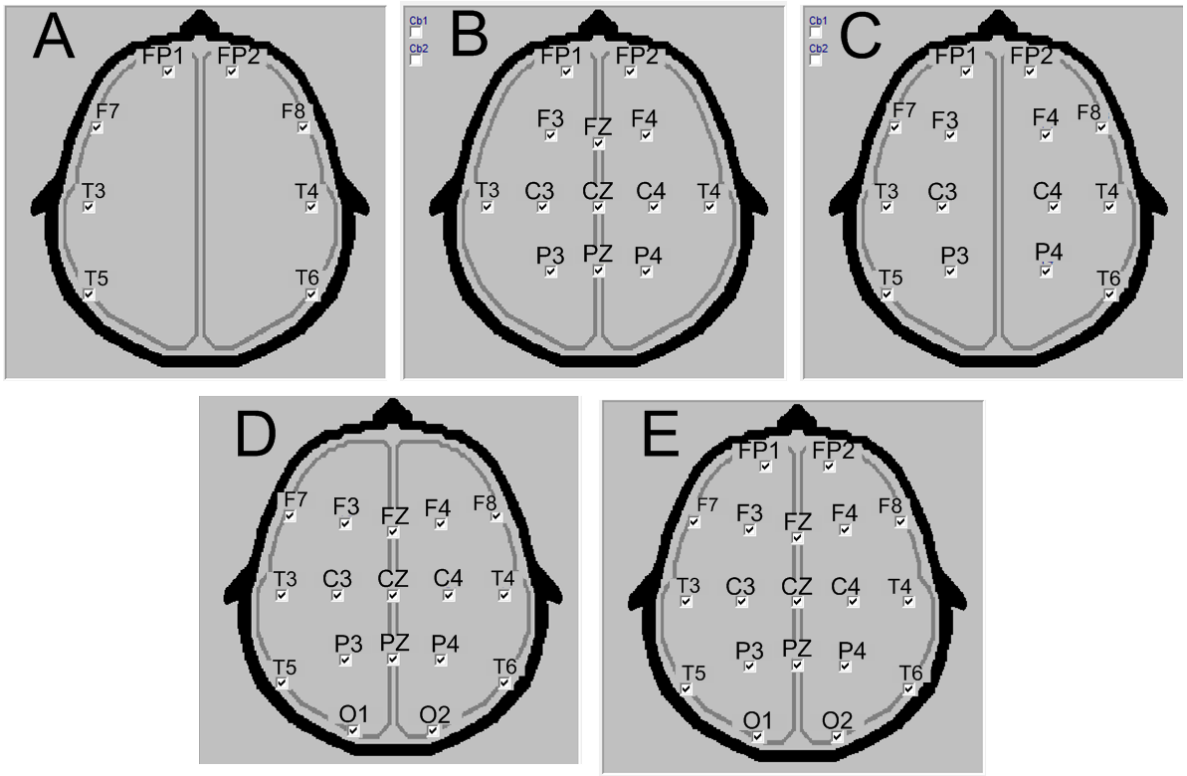


Figure 4.1: Modified 10-20 electrode montage settings used in the EEG recording process. Setups of A)8, B)13, C)14, D)17, and E)19 electrodes.

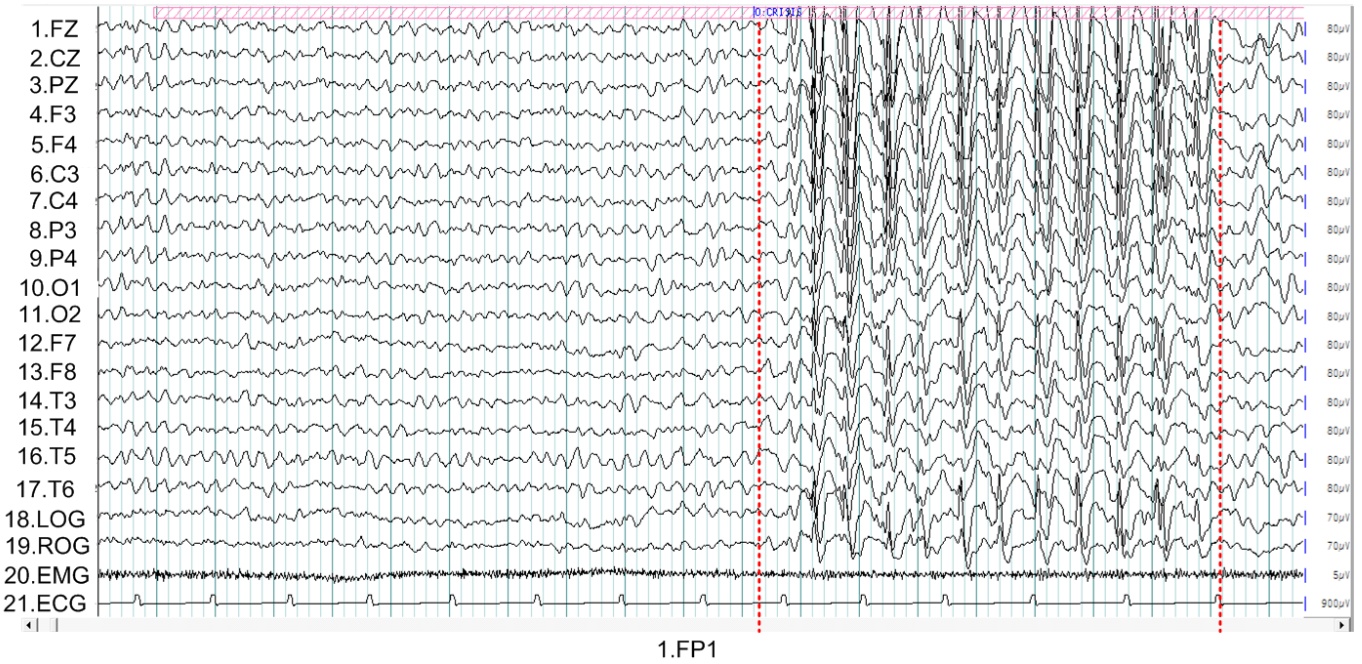


Figure 4.2: Absence seizure recorded from an ambulatory patient (delimited by red lines). In the EEG segment, a clear evolution in the EEG pattern can be observed in frequency and morphology during the seizure. In contrast no changes can be noticed in the EMG signal registered. The absence of changes in the EMG confirms the lack of motor signs during the episode.

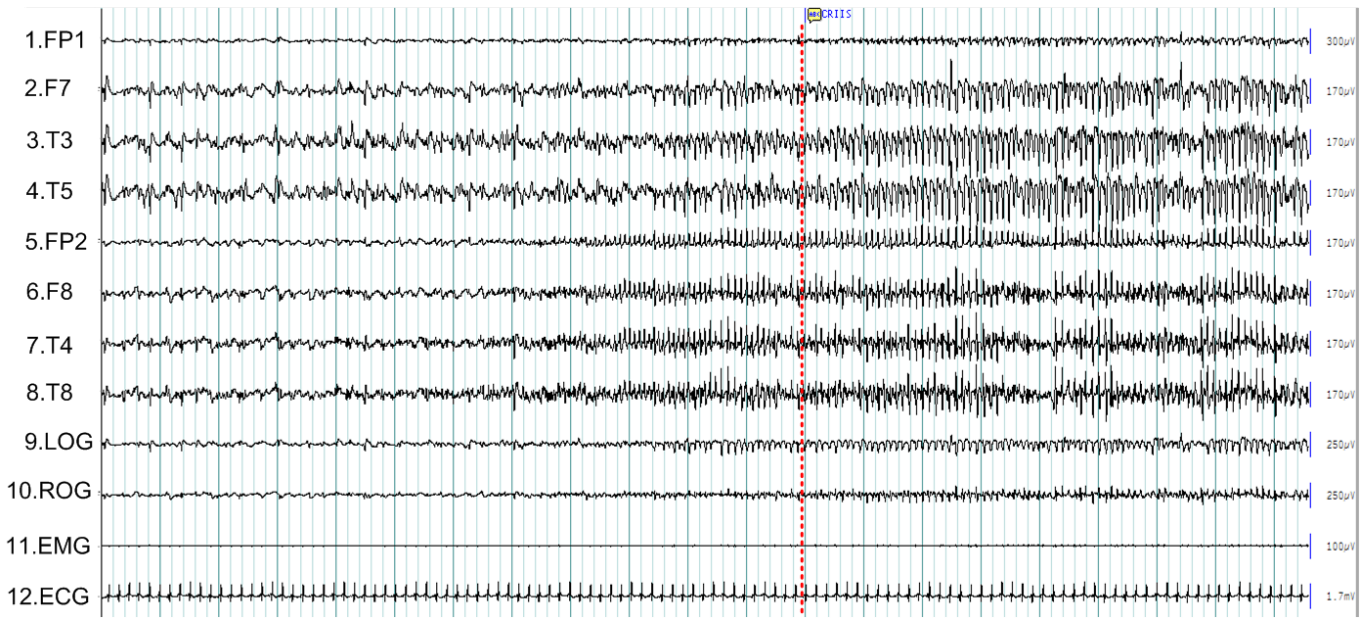


Figure 4.3: NCES from a ICU stuporous patient. As can be noticed the EEG evolution in this patient seizure is not as clear as it is for the absence seizure from Fig. 4.2. The absence of muscular activity can be confirmed by the lack of changes in the EMG during the seizure.

Table 4.1: Description of the database of EEG recordings. The Table describes the protocol and type of medical setting for each recording, as well as the patient etiology. As can be seen, the etiology of the patients and the duration of the recordings is very different. Regarding the recording's duration, a clear difference can be appreciated between the ones from the Ambulatory setting and the ones from ICU. In the table 'x' and 'o' indicate whether the signal was recorded or not for the patient respectively.

Patient data				*Record*		*Protocol*				
Case	*Gender*	*Diagnosis*	*Type*	*Duration(h:min:s)*	*Seizures #*	*Channels*	*ECG*	*EOG*	*EMG*	*Video*
1	*M*	*Temporal lobe epilepsy.*	*Ambulatory*	*00:38:48*	*6*	*19*	*o*	*o*	*o*	*o*
2	*M*	*Focal Epilepsy.*	*Ambulatory*	*00:16:44*	*3*	*19*	*o*	*o*	*o*	*o*
3	*M*	*Temporal lobe epilepsy.*	*Ambulatory*	*00:48:29*	*13*	*19*	*o*	*o*	*o*	*o*
4	*F*	*Partial Temporal Lobe Epilepsy.*	*Ambulatory*	*00:14:05*	*5*	*19*	*o*	*o*	*o*	*o*
5	*F*	*Partial psychic seizures.*	*Ambulatory*	*00:16:05*	*2*	*19*	*o*	*o*	*o*	*o*
6	*F*	*Lenns-Gastaut. Atonic Neck.*	*Ambulatory*	*00:39:30*	*12*	*19*	*o*	*o*	*o*	*o*
7	*F*	*Head trauma.*	*Ambulatory*	*00:28:15*	*2*	*17*	*x*	*x*	*x*	*o*
8	*F*	*Spike wave and Polispike wave.*	*Ambulatory*	*08:05:58*	*6*	*17*	*x*	*x*	*x*	*o*
9	*M*	*Absence seizures.*	*Ambulatory*	*07:42:52*	*34*	*17*	*x*	*x*	*x*	*o*
10	*F*	*Leukosytosis.*	*ICU*	*19:36:38*	*40*	*19*	*x*	*x*	*x*	*x*
11	*M*	*Higly malignant astrocytoma.*	*ICU*	*03:34:50*	*2*	*13*	*x*	*o*	*x*	*x*
12	*F*	*Unknown*	*ICU*	*21:12:46*	*6*	*8*	*x*	*x*	*x*	*x*
13	*F*	*Connective tissue mixed disease.*	*ICU*	*00:47:20*	*5*	*19*	*x*	*x*	*x*	*x*
14	*F*	*Cerebral gliomatosis.*	*ICU*	*01:02:32*	*3*	*14*	*x*	*x*	*o*	*x*

4.3 EEG training and testing data

The recordings are split in training and testing sets to perform the seizure detection. The first seizure and the seizure-free EEG prior to it are used to construct the training set. This means that this set includes the EEG data from the recording onset to the end of the first NCES detected by the doctor. Fig. 4.4 shows an example of EEG data used to built the training set. The rest of the EEG recording, namely all remaining NCES and the seizure-free EEG (n-NCES) is used as test data. An example of the test set EEG data is shown in Fig. 4.5.

The EEG is analyzed in segments of 3s (epochs). Hence, the training sets defined for each case are composed with all the epochs from the first seizure (= positive class) and the same number of n-NCES epochs from the EEG data selected for training as explained before (=negative class). The test set is composed of all epochs from the EEG data selected for testing as explained in the previous paragraph. The training and test sets defined for each case are detailed in Table 4.2. It was impossible to balance the training sets for the cases 2 and 3 since the duration of the EEG prior to the first seizure was too short to meet the same number of epochs from the seizure (the first seizure started after a few seconds of recording).

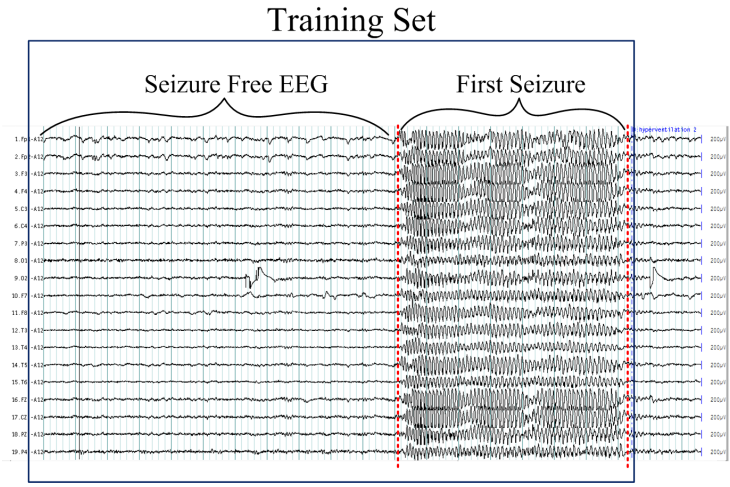


Figure 4.4: Description of an EEG data training set. The figure illustrates the EEG recording from the beginning of the recording process until the end of the first NCES detected. The red bars indicate the beginning and end of the seizure. The blue box indicates an example of the EEG data selected as training set.

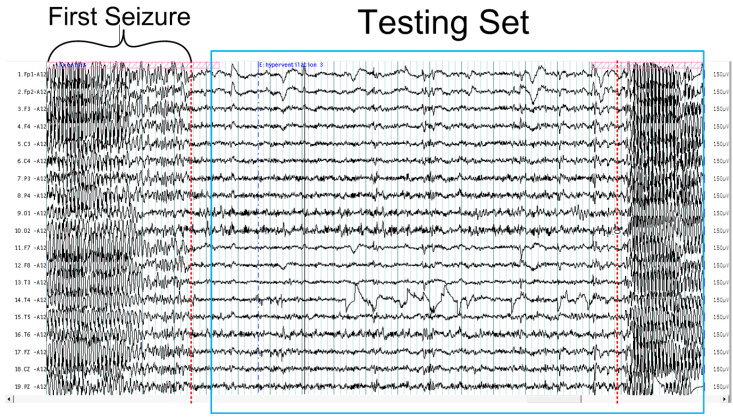


Figure 4.5: Description of an EEG data test set. The figure illustrates the EEG recording from the end of the first NCES detected to the beginning of the second seizure. The two red bars indicate the ending of the first NCES and the beginning of the second NCES respectively. The blue box indicates an example of the EEG data selected as test set.

Table 4.2: Available Data for NCES and seizure free (n-NCES) class and the resulting training sets for each EEG recording given in epochs of 3s duration

	Epochs							
	Available Data		First Seizure and prior EEG				Test Set	
			Modeling Data		Training Set			
Recording	NCES	n-NCES	NCES	n-NCES	NCES	n-NCES	NCES	n-NCES
1	58	345	4	177	4	4	54	168
2	39	127	26	5	26	5	13	122
3	140	424	12	1	12	1	128	423
4	73	112	8	30	8	8	65	82
5	10	18	3	7	3	3	7	11
6	84	269	13	18	13	13	77	251
7	6	237	3	143	3	3	3	94
8	13	4827	3	18	3	3	7	4809
9	163	4558	3	24	3	3	160	4534
10	1074	3346	19	125	19	19	1055	3221
11	16	1841	12	1381	12	12	4	460
12	875	1036	87	152	87	87	788	884
13	90	860	16	33	16	16	74	827
14	44	472	13	59	13	13	31	413
Total	2685	18472	222	2173	222	190	2466	16299

Chapter 5

NCES Detection using scalp EEG

5.1 Introduction

In this Chapter we introduce a new method to detect NCES. The primary aim of this method is to alleviate the shortcomings of the state-of the-art as reviewed in Chapter 2. In particular, this method attempts to identify the NCES based on their similarity to the first NCES detected by the physician on the EEG [64]. The explored features are obtained by means of a multiway analysis of the EEG signal represented as a third order tensor $X \in \mathbb{R}^{(F \times T \times Ch)}$ with modes $frequency \times time \times channels$.

The tensors are computed by expanding EEG segments of 3 s duration using CWT or HHT. The WA is a popular tool for seizure EEG tensor construction [2, 40]. However, we propose the HHT since it resolves $time \times frequency$ events with a finer resolution than CWT and thanks to the adaptive nature of the empirical mode decomposition it yields a more meaningful physical interpretation of the underlying EEG data.

The tensor decomposition performed here are CPD and BTD. It has been shown that decompositions of multiway EEG representations can extract seizure sources and accurately characterize the seizure pattern [2, 24, 40]. The multiway analysis exploits the EEG high-dimensional structure by analyzing its spectral, temporal and spatial properties simultaneously. As such, this approach represents a neat and practical alternative to analyze large, heterogeneous and multidimensional

feature sets. Moreover, it eliminates the need for thresholding since the classifiers will implicitly evaluate the similarity of the signatures. Fig. 5.1 shows a general block diagram of the proposed method.

5.2 Tensor Formulation

The EEG from each recording is divided into segments (epochs) of 3 seconds long. All epochs are expanded in the time-frequency domain using a CWT or HHT. A 3rd order tensor is built from every epoch with '$frequency \times time \times channels'$ axes.

The EEG is expanded into a three-way tensor $X \in \mathbb{R}^{(F \times T \times Ch)}$ with the CWT as described in Section 3.2.1. Since the NCES activity is below $5Hz$ [53], five wavelet scales were computed, which correspond to $1 - 5Hz$ frequencies, with 1 Hz resolution. The tensors obtained with CWT will be referred further on as W_Tensor.

The tensors with the HHT, further on defined as H_Tensor, are built as described in Section 3.2.2.The two tensor sets obtained, W_Tensor and H_Tensor, are decomposed using, CPD and BTD. The EEG epochs that contain seizure activity will be labeled as positives and the rest as negatives.

To differentiate the feature sets, those obtained from a tensor built with HHT are denoted by 'h', 'w" is used for those obtained with CWT. Space, Frequency and Time signatures obtained from the tensor decomposition, are referred as 'spac', 'freq', and 'time'. The training using all assembled signatures together is specified as 'asm'. To tag the decomposition method 'CP' and 'BT' will be used to refer to CPD and BTD respectively. As an example, the Space signature obtained by the CPD decomposition of a tensor built with CWT, will be denoted as: w_spac_CP .

5.3 Tensor Decomposition

The tensors obtained by CWT or HHT are analyzed using CPD and BTD. CPD provides a trilinear description of the data, which is a very compact and interpretable model. This model assumes that the seizure pattern maintains the same frequency and topography within a certain period of time. CPD has been applied successfully in several studies that represent the EEG as a third-order tensor to recognize epileptic seizure activity [2, 40, 24, 6]. This fact makes CPD an attractive method for our study.

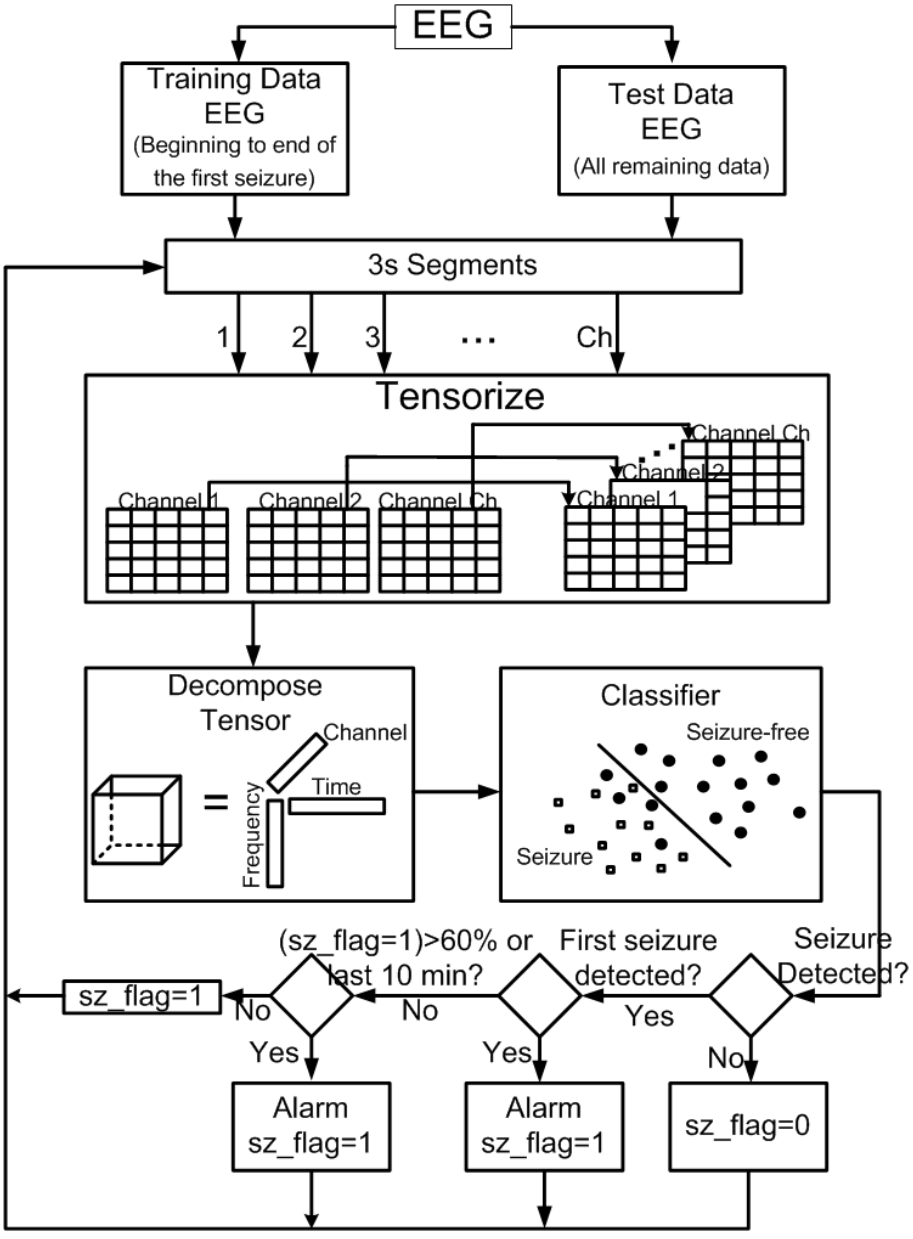


Figure 5.1: General block diagram of the proposed NCES detector using scalp EEG.

On the other hand, BTD is more flexible, and facilitates a more accurate model when the data frequency or topography changes over time. Some studies have demonstrated that BTD outperforms CPD for EEG applications where the pattern of interest shows evolution in frequency, morphology and/or topography [40, 100]. Epileptic seizures have been reported as having such kind of behavior. Therefore, BTD exploration is also interesting for this application.

5.3.1 Tensor Rank Estimation

Before performing CPD and BTD, the tensor rank must be determined. To determine the tensor *ml_rank*, the MLSVD is performed on all tensors obtained from the training sets defined in Chapter 4 for both tensor sets (for more information about the available data see Table 4.2). Then, the obtained singular values (SV) are inspected in order to truncate the MLSVD core such that the SV represent more than 95 % of the data variance. The truncated MLSVD core was used to initialize the LMLRA. The LMLRA method is then performed to find the multilinear rank that best approximates the tensor in the least squares sense [18].

As before, this procedure is performed for all tensors obtained from the 3 s EEG epochs corresponding to the first seizure and all available n-NCES epochs prior to it. Therefore, for every 3 s segment, we obtained a rank value in each of the three modes (i.e. frequency, time and space, (R_f, R_t, R_{ch})). Afterwards, rank histograms are constructed by counting the number of tensors with a certain rank, as shown in Fig. 5.2. Based on these histograms, the most frequently occurring ranks at each mode are selected as upper bound to perform CPD and BTD.

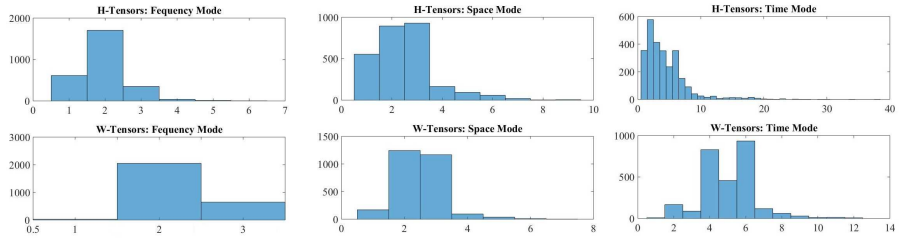
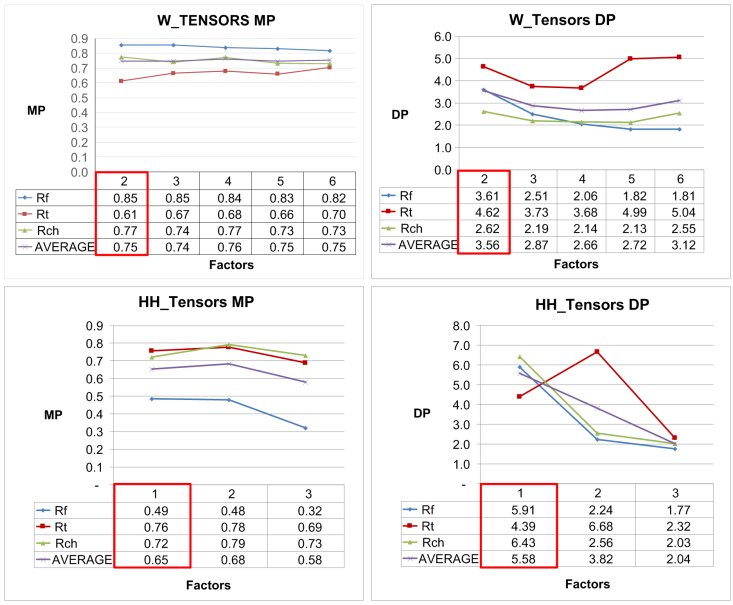


Figure 5.2: LMLRA analysis Frequency, Time and Space mode histograms. Axis X and Y represent the number of components and tensors respectively.



W_TENSORS MP	2	3	4	5	6
Rf	0.85	0.85	0.84	0.83	0.82
Rt	0.61	0.67	0.68	0.66	0.70
Rch	0.77	0.74	0.77	0.73	0.73
AVERAGE	0.75	0.74	0.76	0.75	0.75

W_Tensors DP	2	3	4	5	6
Rf	3.61	2.51	2.06	1.82	1.81
Rt	4.62	3.73	3.68	4.99	5.04
Rch	2.62	2.19	2.14	2.13	2.55
AVERAGE	3.56	2.87	2.66	2.72	3.12

HH_Tensors MP	1	2	3
Rf	0.49	0.48	0.32
Rt	0.76	0.78	0.69
Rch	0.72	0.79	0.73
AVERAGE	0.65	0.68	0.58

HH_Tensors DP	1	2	3
Rf	5.91	2.24	1.77
Rt	4.39	6.68	2.32
Rch	6.43	2.56	2.03
AVERAGE	5.58	3.82	2.04

Figure 5.3: Obtained MP and DP for H_Tensors and W_Tensor CPD signatures computed for several numbers of factors. The displayed MP values are computed for the feature matrices of the first NCES. The MP and DP values displayed on the charts are averaged over the number of features of each mode. The red boxes indicate the selected decomposition.

The best tensor model for the classification task is selected by computing the modeling power (MP) and the discrimination power (DP) of the signatures obtained from the CPD and BTD with a certain rank. Fig. 5.3 shows the computed MP and DP values for the signatures obtained from H_Tensor and W_Tensor after applying CPD. Based on the LMLRA (see Fig. 5.2) the element of the *ml-rank* triplet (R_f, R_t, R_{ch}) most commonly occurring at each mode is inspected. Then, the largest rank among the three modes, denoted as R, is selected as an upper bound to perform CPD. This led to performing CPD with $R = 1, 2, 3$ for H_Tensor and $R = 2, \ldots, 6$ for W_Tensor. BTD is applied to all possible rank combinations of $R_f = 1, 2$, $R_t = 1, 2, 3$ and $R_{ch} = 1, 2, 3$ for H_Tensor and $R_f = 2$, $R_t = 4, 6$ and $R_{ch} = 2, 3$ for W_Tensor (where R_f, R_t, and R_{ch} are the possible rank values of the Frequency, Time and Space modes respectively). Combinations violating the rules $R_f \leq R_t \times R_{ch}$, $R_t \leq R_f \times R_{ch}$ and $R_{ch} \leq R_f \times R_t$ were removed. This resulted in a total of 8 different tensor representations for H_Tensor (excluding the *ml-rank*-$(1, 1, 1)$ since it is equivalent to an $R = 1$ CPD), see Fig. 5.5, and 3 tensor representations for

W_Tensor, see Fig. 5.4.

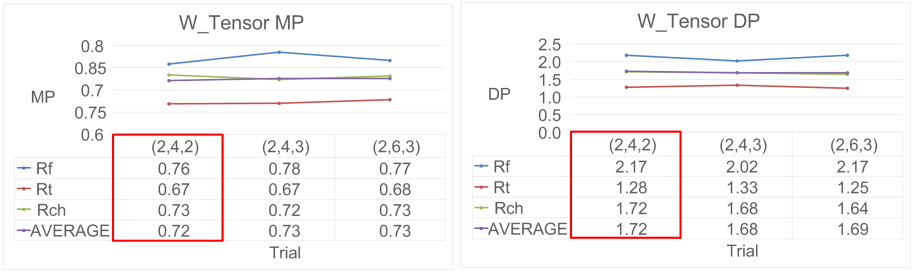
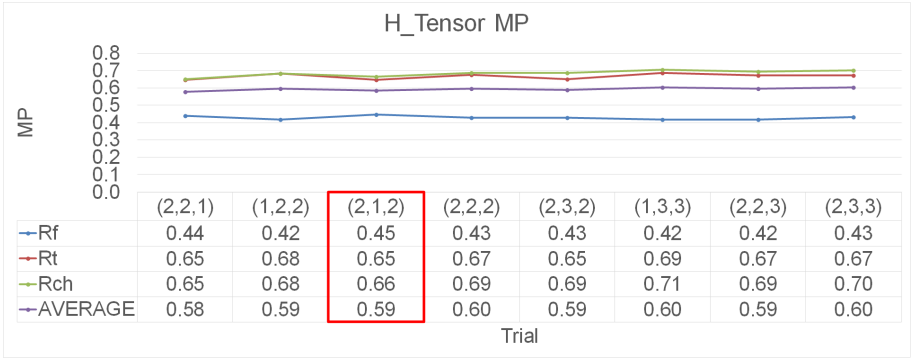


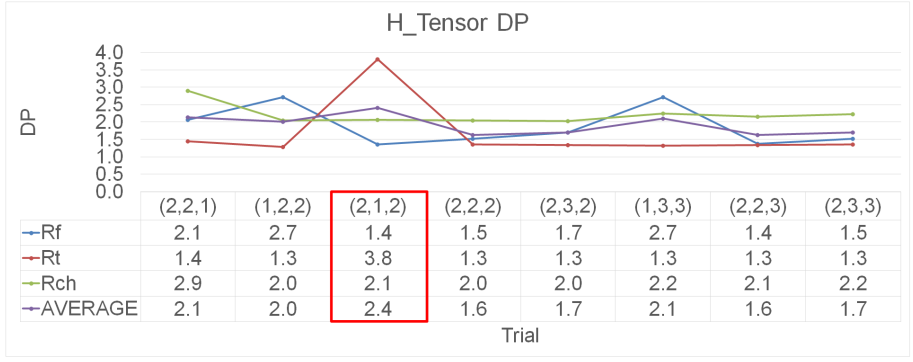
Figure 5.4: Obtained MP and DP values computed for all tensor representations considered for BTD in the W_Tensor *ml-rank* analysis. The displayed MP values are computed for the feature matrices of the first NCES. The MP and DP values displayed on the charts are averaged over the number of features of each mode.The red boxes indicate the selected decomposition.

The average MP value is computed by averaging the values obtained by the three signatures for each tensor representation. The same procedure is performed for the DP. Then, the tensor representation with the average MP closest to 1 and the largest DP (over 1) is chosen. From this analysis, it is decided to decompose the W_Tensor set with *ml-rank* $(2, 4, 2)$ and $R = 1$ for BTD and $R = 2$ for CPD. The H_Tensor BTD analysis is performed with *ml-rank* $(2, 1, 2)$ and $R = 1$ for BTD, and R=1 for CPD. It should be mentioned that a BTD analysis of a tensor with $R = 1$ is equivalent to performing a MLSVD with *ml-rank* (R_f, R_t, R_{ch}).

Note that in the case of H_Tensor, the maximum average MP and DP values of the signatures for CPD are achieved for different tensor representations. The maximum average MP is obtained for $R = 2$ with $MP = 0.68$ and $DP = 3.82$ and the maximum average DP is obtained for $R = 1$ with $MP = 0.65$ and $DP = 5.58$. The same occurs for H_Tensor and W_Tensor signatures for BTD. For H_Tensor BTD the maximum average $MP = 60$ and the maximum average DP is obtained for $(2, 1, 2)$ with $MP = 0.59$ and $DP = 2.4$. For W_Tensor BTD the maximum average MP is obtained for $(2, 4, 3)$ and $(2, 6, 3)$ both with $MP = 0.73$, and $DP = 1.68$ and 1.69 respectively. The maximum average DP for W_Tensor BTD is obtained for $(2, 4, 2)$ with $MP = 0.73$ and $DP = 1.72$. As can be appreciated the difference in MP values for these tensor representations is small compared to the difference in DP values. Since the task to be solved is a classification task, we do not look for an exact approximation of the data but a sufficiently good approximation that allows separating between seizure and



	(2,2,1)	(1,2,2)	(2,1,2)	(2,2,2)	(2,3,2)	(1,3,3)	(2,2,3)	(2,3,3)
Rf	0.44	0.42	0.45	0.43	0.43	0.42	0.42	0.43
Rt	0.65	0.68	0.65	0.67	0.65	0.69	0.67	0.67
Rch	0.65	0.68	0.66	0.69	0.69	0.71	0.69	0.70
AVERAGE	0.58	0.59	0.59	0.60	0.59	0.60	0.59	0.60



	(2,2,1)	(1,2,2)	(2,1,2)	(2,2,2)	(2,3,2)	(1,3,3)	(2,2,3)	(2,3,3)
Rf	2.1	2.7	1.4	1.5	1.7	2.7	1.4	1.5
Rt	1.4	1.3	3.8	1.3	1.3	1.3	1.3	1.3
Rch	2.9	2.0	2.1	2.0	2.0	2.2	2.1	2.2
AVERAGE	2.1	2.0	2.4	1.6	1.7	2.1	1.6	1.7

Figure 5.5: Obtained MP and DP values computed for all tensor representations considered for BTD in the H_Tensors ml-rank analysis. The displayed MP values are computed for the feature matrices of the first NCES. The MP and DP values displayed on the charts are averaged over the number of features of each mode.The red boxes indicate the selected decomposition.

non-seizure data. Therefore, the R and *ml-rank* with higher average DP value is selected in every case.

5.4 Hypothesis Test

The algorithm proposed here, is based on two central assumptions: 1) the distribution of the signatures obtained from the first seizure and the subsequent seizures within a recording is significantly similar, and 2) the distribution of

the signatures obtained from the NCES and n-NCES EEG are significantly different. To provide evidence of this, we perform a hypothesis test.

The test used in both cases is the Wilcoxon signed-rank test. A Wilcoxon signed-rank is a nonparametric test that can be used to determine whether two independent samples are selected from populations with the same distribution. This test is used as the alternative to paired Student's t-test when the normal distribution of the samples cannot be assumed [94]. It performs a two-sided rank sum test of the hypothesis that two independent samples, come from distributions with equal medians, and returns the p-value from the test. P is the probability that the null hypothesis ("medians are equal") for a pair of samples is met. Small values of P refute the validity of the null hypothesis. In all statistical tests, an alpha level of 0.05 is used.

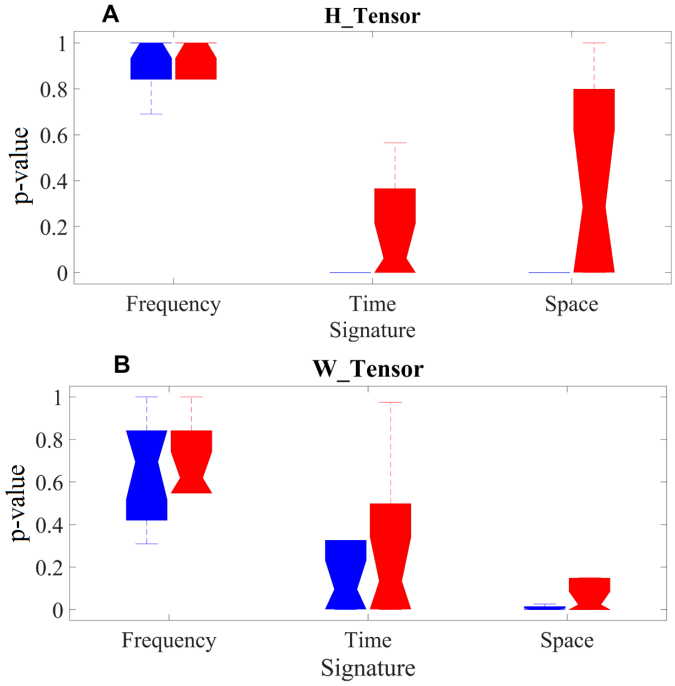


Figure 5.6: p-values resulting from performing Wilcoxon signed-rank test applied to the two performed hypothesis tests. The results in the red boxes correspond to the hypothesis test performed between the first seizure and the subsequent seizures. The blue boxes show the results from the hypothesis test performed between the first NCES and the n-NCES segments. High average p-value (close to 1) means that H0 is accepted.

First, we want to prove the similarity between the first seizure and the rest of the seizures within a recording. The Wilcoxon signed-rank test is performed comparing the signatures obtained with CPD from the first seizure and all subsequent seizures within a recording for both tensor sets. The null hypothesis is defined as H0: the statistical distribution of the signatures from the first seizure and the subsequent seizures within a recording is the same. The alternative hypothesis is assumed as H1: the statistical distribution of the signatures from the first seizure and the subsequent seizures within a recording is significantly different.

Fig. 5.6 displays the results of the performed tests. As can be appreciated in Fig. 5.6 A (red boxes) given the p-values obtained, H0 is accepted for W_Tensor Frequency and Time signatures and must be rejected for Space signatures. Regarding H_Tensor (Fig. 5.6 B (red boxes)), H0 is accepted for all signatures. Frequency signatures showed the highest probability of accepting H0 in both tensor sets, meaning that the feature accurately characterizes seizure samples.

Secondly, the test was applied to prove that 'NCES and n-NCES EEG data distribution are significantly different within a patient recording.' The test was applied to the signatures obtained with CPD of all tensors (both tensor sets) obtained from the training sets specified in Chapter 4. For the Wilcoxon signed-rank test the null hypothesis H0 to be tested is as follows: the statistical distribution of the signatures obtained from the NCES and n-NCES EEG have the same statistical distribution within a patient recording, and the alternative hypothesis H1: the statistical distribution of the signatures obtained from the NCES and n-NCES EEG within a patient recording is significantly distinct.

Fig. 5.6 (blue boxes) shows the results of the second test. According to the p-values obtained, H0 is accepted for the Frequency signatures from both tensor sets. Provided that the Frequency signatures accept H0 for all cases may indicate in our study, that Frequency could not be a good feature for NCES detection. A similar situation is given for the Time signatures from W_Tensor, where H0 is accepted for eleven of the fourteen cases available. H0 is rejected for thirteen of fourteen cases for W_Tensor Space signatures, which means that for these signatures H1 can be accepted. For H_Tensor Space and Time signatures, H0 is rejected in all cases. Hence H1 is accepted for these signatures. From this analysis, it can be concluded that Frequency signatures from both tensor sets and Time signatures from the W_Tensor may not be suitable features to discriminate between seizure and non-seizure samples since these feature values showed the same distribution for both classes. On the other hand, the Space signatures from both tensor sets and the Time signatures from H_Tensor have a good chance of being capable of differentiating NCES and n-NCES samples.

As a conclusion of the performed hypothesis tests, we can say that Frequency

signatures showed the lowest probabilities of having a good performance to separate NCES and n-NCES samples since the distribution of their values is the same for both classes. Time signatures from W_Tensor, and Time and Space signatures from H_Tensor, could have higher probabilities to provide a proper classification of NCES and n-NCES samples. The mentioned signatures showed great potential to accurately characterize the seizure samples (by accepting H0 in the first test) because the statistical distribution of their values differs for each class.

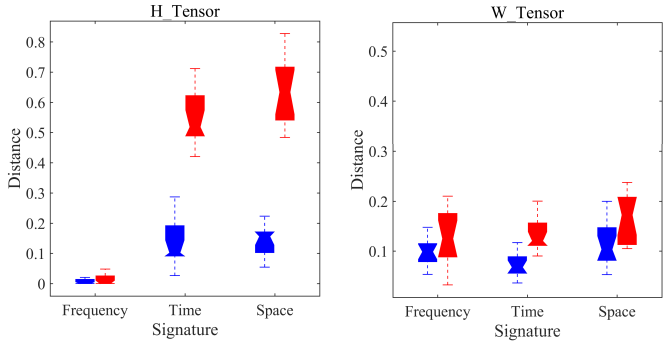


Figure 5.7: Bhattacharyya distance computed for each signature between the first seizure and the subsequent seizures (blue boxes), and between the first NCES and the n-NCES segments (red boxes). These distance values were calculated for all signatures obtained from all tensors in all training sets of all cases specified in Table 4.2 decomposed with CPD.

To confirm the results obtained in the hypothesis tests, we compute the Bhattacharyya distance from the first seizure to the rest of the NCES and to the n-NCES segments for each signature separately. The results obtained (see Fig. 5.7) displayed a short distance between the first seizure and the subsequent seizures and a larger distance between the first NCES and the n-NCES segments. These results confirmed our initial assumptions and gave an idea of what to expect from each parameter during the classification. It is expected to obtain the best performance for those features (signatures) which show the highest Bhattacharyya distances. Fig. 5.8 shows an example of the results obtained. As can be seen in Figure 5.8 the space signatures from both groups are clearly separated for the recordings displayed.

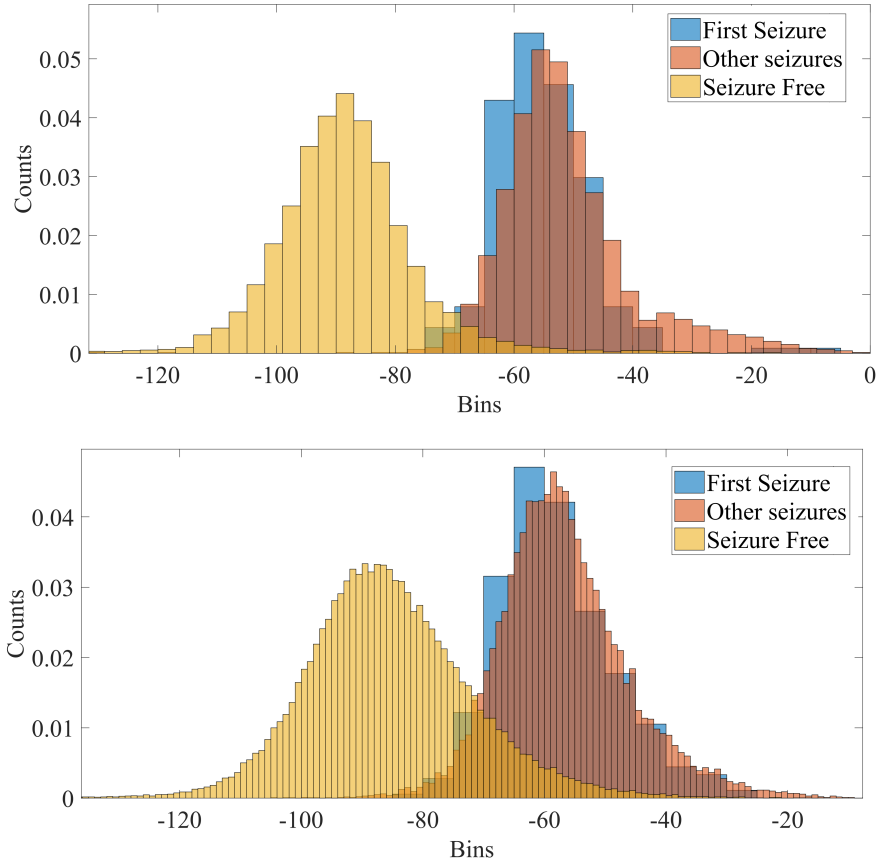


Figure 5.8: Histograms generated by the Space signature distribution of the first seizure, subsequent NCES and n-NCES data for two different cases described in Table 4.1(Case3 (A) and Case10 (B)). The Space signatures were obtained for the H_Tensor decomposed with CPD.

Table 5.1: Bhattacharyya distance computed between the first seizure and the subsequent NCES and n-NCES segments. These distance values were calculated for the Space signature obtained from the H_Tensor decomposed with CPD.

Case	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Other seizures	0.165	0.223	0.222	0.161	0.167	0.171	0.284	0.056	0.103	0.063	0.151	0.054	0.133	0.112
Seizure Free	0.518	0.531	0.541	0.701	0.660	0.484	0.713	0.601	0.717	0.585	0.607	0.827	0.828	0.825

5.5 Classifiers

Three classifiers were trained to perform the NCES detection: a KNN with $K = 3$, a Radial Basis SVM (SVMRB) and a LDA.

Since we are separating two classes, the K value must be 3 or a higher odd number. To determine the value K for the KNN classifier, the values $K = \{3, 5, 7\}$ were evaluated. For $K = 5$ and $K = 7$ the classification results were slightly better (in the order of 10^{-3} and 10^{-2} respectively) than the ones obtained with $K = 3$. On the other hand, the execution time for $K > 3$ increased, in some cases it even doubled. Therefore, the slight improvement in classifier performance achieved for the higher values $K = 5$ and $K = 7$ was considered not significant enough to justify a higher value of K.

The implementation used for the SVMRB estimates automatically the optimal radial basis kernel and the regularization parameter. The kernel optimization is performed by quadratic programming using the QP solver provided by MatLab R2017a. The regularization parameter is estimated by the leave-one-out cross-validation. The implementation of all classifiers is available at the MATLAB toolbox PRTools V.5 [27].

The classification is performed for each patient individually. The classifiers are trained with the training sets defined for each recording in Section 4.3. The classification was performed in two ways: (1) with the signatures from the different modes (Frequency, Time, Space) separately, (2) with all mode signatures assembled. The classifiers' performance is evaluated on the test set in terms of sensitivity, specificity, accuracy and false negative rate.

5.6 Results and Discussion

The classification results are presented for the epoch based analysis since the duration of the epochs is the minimum duration of an epileptic event in the EEG. This means that an alarm should be displayed after each positive detection. This is not a practical implementation in a real life dynamical setting. For the real-life setting, an alarm will be set at the seizure activity onset. Then, subsequent alarms are suppressed and a new alarm is allowed after 10 minutes (lower bound established for NCSE diagnosis [83]) of continuous seizure activity.

The classifiers' performance for the Frequency signatures is shown in Fig. 5.9 A. The best classification results are obtained by a KNN classifier for h_freq_CP signatures with 74.6, 73.7, and 75.6% of specificity, sensitivity and accuracy values respectively, and with a false detection rate of 0.385. According to

the analysis of MP values, Frequency signatures offer the best model of the data when they are obtained from the decomposition of the W_Tensor set either with CPD or BTD. The performance for the frequency signatures for all decompositions is very similar. This analysis revealed that even if the signatures allow an accurate prediction model of the original data, as is the case for W_Tensors signatures, discrimination between classes is not guaranteed. The w_freq_CP, w_freq_BT, and h_freq_BT signatures achieved sensitivity and accuracy values below 80%.

Regarding the Time signatures, the highest DP value is obtained with h_time_CP signatures. Consequently, the best classification results for this feature are achieved by h_time_CP signatures with a KNN classifier (see Fig. 5.9 B). The classifier reached 98.6, 90.9 and 98.0% of specificity, sensitivity and accuracy values respectively. This feature achieved a false positive detection rate of 0.010.

The combination of the signatures obtained with CPD and BTD is unique. Yet, the components computed (signatures) on their own are not. When more than one component is computed, it is expected that some of them fit the pattern of interest (hopefully most of them), and others model the noise in the data or other underlying patterns that may appear. The order in which these components are obtained is not unique and may vary from epoch to epoch. It is not possible to determine which component contains NCES information based on it. Hence, the features computed from different epochs may not correspond to each other, thereby hampering the classification based only on the signatures.

We further discuss the Space signature performance. As low DP values were obtained for W_Tensors CPD and BTD space signatures, the classifier performance for these signatures is lower than that obtained for the H_Tensors signatures, Fig. 5.9 C. The highest specificity, sensitivity and accuracy values for the Space feature are achieved by SVMRB with 99.0%, 98.8% and 99.0% respectively for the h_spac_CP signatures. The KNN classifier obtained similar results for these signatures with a specificity of 98.6%, a sensitivity of 99.8%, and accuracy of 98.5%. The h_spac_CP signatures combined with SVMRB achieved a false positive detection rate value of 0.014.

The training set using the assembled signatures obtained similar results to those achieved by the space signature. The best classifier for the assembled signatures training is SVMRB for H_Tensor CPD signatures with a specificity value of 98.5%, a sensitivity value of 99.0% and accuracy of 98.7%.

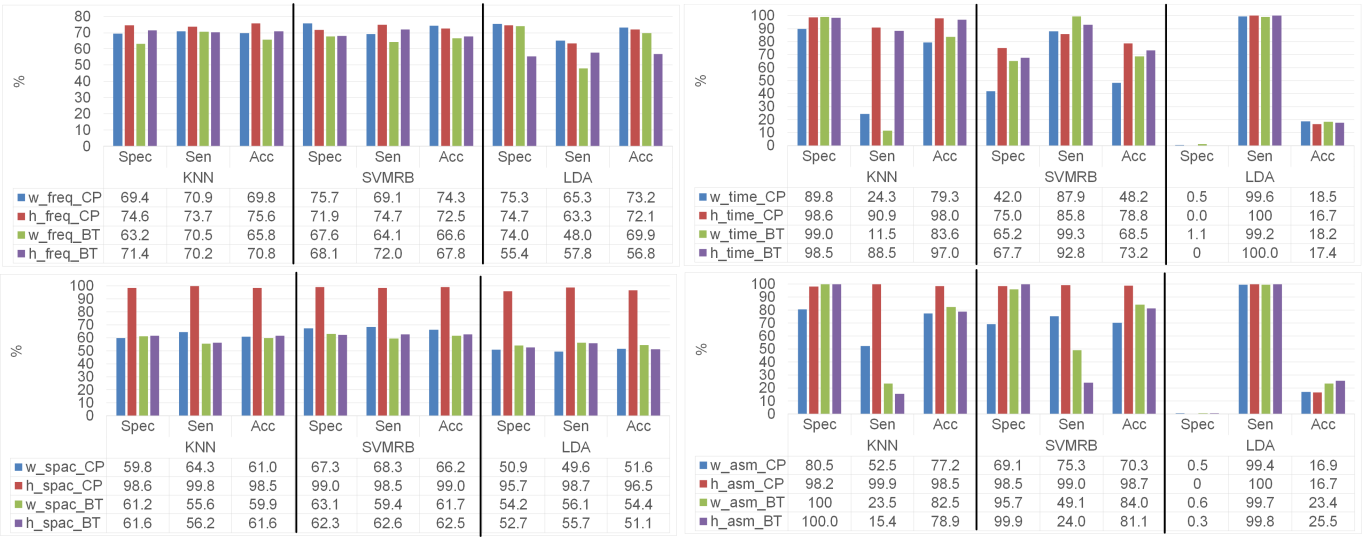


Figure 5.9: Specificity, Sensitivity, and Accuracy scores obtained in the classification of the test set for KNN, SVMRB and LDA. The results for the different signatures are displayed in the following order: A) Frequency, B)Time, C)Space, and D) Assembled signatures. The columns in the graph for each classifier stand for specificity, sensitivity, and accuracy respectively.

The performed experiments showed that the tensor built with HHT separates better the NCES data, thereby outperforming the tensor obtained with the CWT. Note that we also tested datasets created using the energy of the wavelet coefficients and other mother wavelet. None of these data sets obtained better results in the training than those obtained using the wavelet coefficients computed with the Mexican Hat mother wavelet or HHT.

The H_Tensor signatures obtained with CPD outperform those computed with BTD in most of the training sets. This is somewhat unexpected, as NCES is known to evolve along the EEG. A low-rank BTD is a suitable model for analyzing such evolving patterns [40]. The NCES can evolve along the EEG recording in frequency, morphology (time), or location (space). The Salzburg Consensus Criteria for NCSE proposed in [53] by adopting the American Clinical Neurophysiology Society (ACNS) criterion provided more detailed, unambiguous guidelines about the characteristics of this “evolution" in each of these dimensions which could clarify this unexpected result. According to the ACNS, evolution in frequency is defined as at least 2 consecutive changes in the same direction by at least 0.5 Hz, e.g. from 2 to 2.5 to 3 Hz, or from 3 to 2 to 1.5 Hz; Evolution in morphology implies at least 2 consecutive changes to a novel morphology; Evolution in location is defined as sequentially spreading into or sequentially out of at least two different standard *10-20* electrode locations. In order to qualify as “present", a single frequency or location must persist at least 3 s [53]. This implies that the changes usually occur slowly, and a certain pattern tends to persist 3 s or longer. The CPD model assumes that the source data preserve the same frequency and location within the observed epoch [40]. Due to the analyzed epochs duration (3 s), it is therefore unlikely to observe a lot of evolution in frequency, time or space occurring within the same epoch. Hence, the NCES data analyzed in these small epochs fulfill the trilinear criteria justifying the use of CPD and explaining why BTD overfits the model.

Although HHT and CWT are used for this proposal, different tensorizations would also be possible, e.g. empirical wavelet transform and tunable-Q wavelet transform [8, 9], which were successfully used to detect epileptic seizures. Such methods could be tried on NCES data in order to demonstrate whether these algorithms are capable of capturing this seizure activity as well.

Both classifiers, KNN and SVMRB, showed similar performance in the classification for different training sets. The classifier's performance depended on the signature used to build the model (e.g. h_spac_CP combined with SVMRB outperforms the combination of h_spac_CP and KNN, h_time_CP combined with KNN outperforms h_time_CP combined with SVMRB). Therefore, is not possible to establish a clear superiority of one over the other given the obtained results.

The LDA classifier showed to be able to classify NCES data with the Frequency and Space signatures. This classifier obtained values of specificity, sensitivity, and accuracy over 95% for the h_spac_CP. However, for high-dimensional features such as Time and Assembled signatures LDA showed to be unable to identify the NCES from n-NCES data. For the Time and Assembled signatures, LDA displayed high sensitivity but low values of specificity and accuracy.

As the achieved false positive rate values suggest, Fig. 5.10, the number of non-seizure events classified as seizures is low. The majority of these misclassifications are due to the appearance of long duration paroxysms (Fig. 5.11) or a large number of successive spikes in the EEG. These events, are long enough to resemble a seizure but are too brief to be classified as a seizure according to the physician's criteria. False positive detections also occur at epochs with pre-ictal or post-ictal EEG activity.

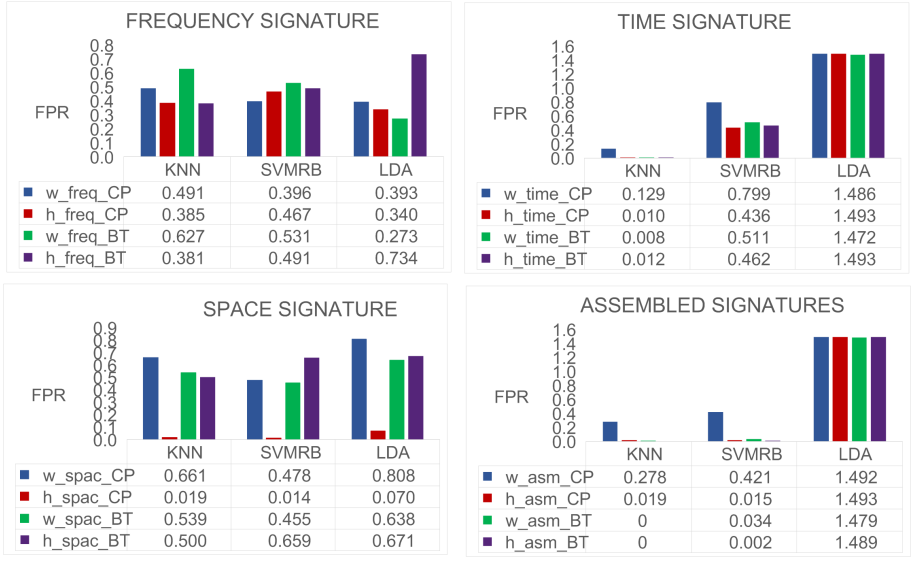


Figure 5.10: FPR values displayed by the KNN, SVMRB and LDA classifiers during the different trainings.

False negative detections are found in recordings longer than 10 hours (Fig. 5.12). What was observed in our experiments is that the seizure pattern might evolve after some hours. This evolution, in our case, was observed in patients with long-term EEG recordings (patients 10 and 12) after 10h of recording. It is not possible to predict when these changes will appear. According to the medical

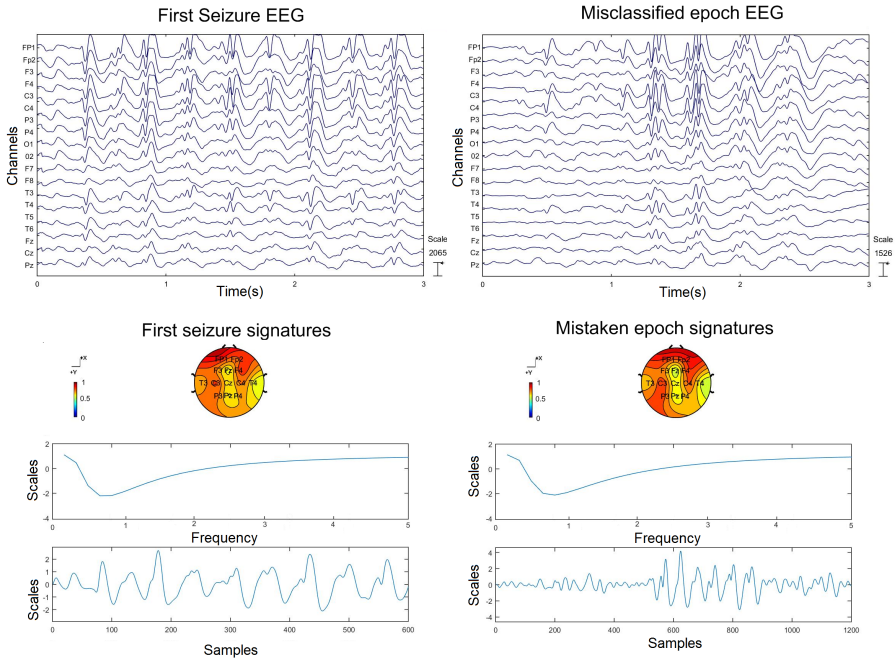


Figure 5.11: H_Tensors CPD: false positive classification example. Misclassified n-NCES epoch shows a short paroxysm of approximately 1 s of seizure-like activity in the EEG. This activity is captured by highly similar signatures as the first seizure's time and frequency signatures. This leads to the detection of the paroxysm as a seizure. Signatures are represented in the following order: Space, Frequency and Time.

literature [71] it is less probable that these changes occur within one hour than in longer time windows. The evolution of the seizure will depend on the patient etiology and response to the treatment. The proposed time window (10 min) is based on our data, this parameter can be adjusted by the doctors taking into account patient information that we ignore such as the seizure etiology, cerebral state, and recording time. To avoid discrepancies between the patterns in the set used for training and the subsequent seizures we propose to use an incremental learning approach (see next Chapter). Including new EEG patterns from subsequent seizures to the initial training set could help to solve this. The proposal is not to reset the signatures on the training at every new training, but add new patterns to the training set, and retrain the classifier. We consider this

will make our method robust to changes in the seizure pattern. There is no need to stop the classification to retrain the classifier, this can be done in parallel. This is a strength when comparing our proposal to threshold-based methods. The threshold based methods are unable to respond to seizure and background EEG pattern changes. This lack of adaptability makes the threshold based methods less accurate for long-term EEG recordings.

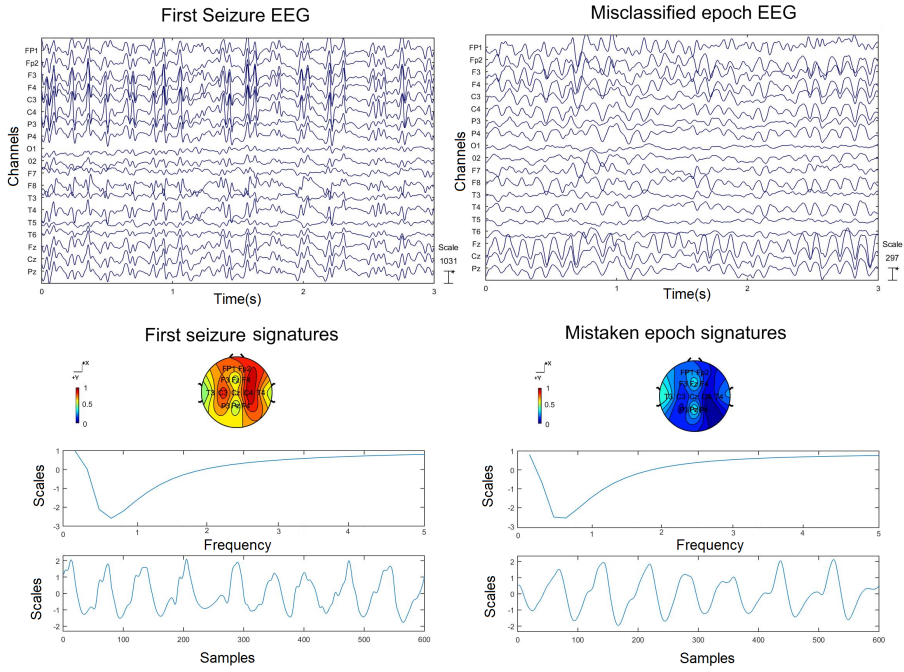


Figure 5.12: H_Tensors CPD: false negative classification example. Seizure epoch classified as a non-seizure. The EEG seizure patterns from the misclassified epoch are dissimilar to the patterns from the first seizure. The space signature from the misclassified epoch shows an inverted activation pattern compared to the one shown by the first seizure. This inverted pattern could be interpreted as a lighter seizure or a non-seizure EEG segment. The frequency component from the misclassified epoch is steeper than the first seizure frequency component. The time component from the misclassified epoch is smoother than the one from the first seizure. However, the time component from both epochs are quite similar. The dissimilarities found between the computed components could explain the classifier failure. Signatures are represented in the following order: Space, Frequency and Time.

Some misclassifications are caused by high-frequency artifacts probably related to poor electrode contact. These are only found at the epochs where the artifact affected more than one electrode at the same time instant. All signatures proved to be robust to artifacts affecting isolated electrodes.

Based on the performed analysis, the best combination for NCES detection is provided by the Space signatures as features and the SVMRB as classifier. The Space signature must be obtained with HHT and CPD as tensorization and tensor decomposition method respectively.

Table 5.2: Rough comparison of the proposed method performance with other existing methods in terms of accuracy, sensitivity and specificity. N/R stands for "Not Reported".

Method	Specificity	Sensitivity	Accuracy
Proposed Method	99.0%	98.8%	99.0%
Liang et. al. [54]	N/R	N/R	97.5%
Xanthopoulos et. al. [96]	N/R	N/R	95.0%
Kollialil et. al. [50]	N/R	N/R	99.6%
Fatma et. al. [28]	88.02%	100.0%	94.01%
Jacquin et. al. [42]	96.0%	83.0%	89.5%
Minasyan et. al. [57]	99.0%	71.0%	85%
Khan et. al. [48]	96.9%	86.8%	91.85%
Sharma et. al. [72]	93.3%	100.0%	N/R

Table 5.2 presents a rough comparison of our results with the results achieved by other methods for NCES detection, in terms of sensitivity and specificity. As can be noticed, the methods from Sharma et. al. [72] and Fatma et. al. [28] obtained superior results in terms of sensitivity than those achieved by the method proposed here. The database used in these publications consists of 13 EEG recordings, 6 of which have only one seizure. The parameters extracted from epochs of the same seizure tend to display a high similarity. Hence, any classification task made with parameters from epochs of the same seizure will reach high sensitivity values.

In terms of accuracy, Table 5.2 shows that our method is only outperformed by Kollialil et. al. [50]. The method proposed by Kollialil uses an SVM to classify between normal, interictal and seizure EEG. In this publication, three different labeled datasets are used. The classification was not performed in a continuous recording. It should be noted that with such test data, the classifier does not face issues as artifacts, EEG pattern changes or paroxysms as done in our study.

Regarding the execution time, which is a major concern for detection algorithms,

our method displays a runtime from 6.04 s to 3.3 min for the training step. All tests are performed on a computer with an Intel Core-i3 processor at 1.70 GHz with 8 GB of RAM. The training results are obtained for the EEG recordings with the shortest (9 s) and the largest (4.35 min) first seizure respectively. It should be noted, that the training will only be performed once at the beginning of the monitoring. The training step includes the tensorization and analysis of all first seizure epochs and the classifier training. Hence, the execution time will depend on the first seizure duration. It could be decided to use only part of the seizure for training if this is very long. Yet, it should be kept in mind that doing so fewer training patterns will be included. In this sense, there is a compromise between the number of positive epochs to include in the training set and the quality of the classifier.

The classification runtime ranged from 0.37 to 3 s which means that our method can run in real-time. These results are obtained by classifying all epochs from the test set. The duration of the classification step will depend on the pattern complexity. As the pattern becomes more complex, more time is required to perform the EMD. The average detection delay of the algorithm is 5.4 s.

Regarding the alarm system, as defined in the beginning of this Section (not setting an alarm at every positive detection), we are referring to the sound alarm. The algorithm will set a mark in every EEG window if this is recognized as a seizure. Our purpose is to alert the doctors, when the seizure activity begins, and also alert them for the risk of this seizure activity to progress into SE (after 10 minutes of persistent seizure activity). To diagnose SE for ICU patients 30 min of continuous seizure activity is required.

In practice, it means that sensitivity is higher than the one reported: even if some epochs of a seizure are not detected (sensitivity for that particular seizure is less than 100%), if at least one epoch of a seizure is detected, there will be an alarm. It also means that in practice there are less false negatives (if there are 2 false negatives within 10 minutes from each other, there is no separate alarm for the second one).

5.7 Conclusion

The presented Chapter introduced a NCES detection method that exploits techniques such as multiway data analysis and HHT. It is the first algorithm to apply HHT to model seizure data with tensors. This is the first algorithm that uses this combination for NCES detection. The HHT provides an accurate definition of particular events in the $time \times frequency$ space capturing the NCES data in a tensor of rank close to $R = 1$. This turns all signatures

obtained from H_Tensor decomposition in relevant features to perform the NCES detection.

The best results at the classification were achieved for H_Tensor Space and Assembled features computed with CPD combined with the SVMRB classifier, closely followed by H_Tensor Space features computed with CPD combined with KNN.

According to the obtained results, the proposed method proved to be one of the best NCES methods so far. The method was validated on continuous clinical data in a realistic training setting proving that the method can run in real-time.

The number of false positives for the training with H_Tensor Space and Assembled signatures obtained with CPD was low. Yet, the appearance of EEG activity resembling the seizure affected the performance of these signatures. The false negatives found were related to EEG pattern changes at the seizures distant in time from the first recorded seizure. Due to the seizure EEG pattern evolution caused by the underlying disease progression, adaptive training strategies must be implemented to improve this as shown in the next Chapter.

Chapter 6

Nonconvulsive epileptic seizure monitoring with incremental learning

6.1 Introduction

In the previous Chapter, was introduced a method which accurately detect NCES. However, this approach showed to be less accurate when the EEG morphology evolved over time and showed different characteristics from the morphology at the beginning of the recording. The changes in the EEG morphology in both NCES and n-NCES activity may be caused either by a reaction to sedative medication or by the patient's disease evolution, e.g., potentially reversible complications such as elevated intracranial pressure [89, 19]. It is not the first time that this kind of pattern drift is reported to be an issue for seizure detection methods. Kharbouch et al. [49] found that the evolution of the earliest seizure activity could differ within an individual, especially when seizures are generated in different brain regions. Some authors address this problem using a multi-task learning approach [87] thereby improving the differentiation between seizure and non-seizure EEG.

Specialists are trained to identify epileptic seizure patterns in the EEG and are able to identify them in more complex scenarios, e.g., when the background EEG resembles ictal activity. Similarly, automatic seizure detection methods need appropriate training to be able to solve a complex signal processing and

pattern recognition task given the complex nature of the EEG signals. An approach to solve this task is to model it as a supervised learning problem. From the supervised learning point of view, to discriminate between the two classes: NCES and n-NCES, a dataset of examples labeled by the clinician is used to train a machine learning or pattern recognition algorithm. Given a set of S training vectors $\{x_s\}_{s=1}^{S}$, $x_s \in \mathbb{R}^p$ with p as the number of features with labels $\{y_s\}_{s=1}^{S} \in \{\pm 1\}$, the task is to infer a model M that correctly *assigns* all training vectors x_s to their corresponding target y_s from this data.

The most commonly used training strategy for a machine learning algorithm is the batch method. In the batch method, the algorithm has a fixed number of samples and uses them to train a classifier. This classifier is used for classifying new samples without further modification. This approach has been applied successfully for the NCES detection in several methods proposed in the literature [50, 72, 28, 65, 66].

However, in a real-life EEG monitoring setting, at the beginning of the monitoring, there is no data, but it becomes available over time. Most of the existing algorithms for NCES detection are non-patient-specific, i.e. using training data from different patients. The seizure morphology can vary a lot from one patient to another. Therefore, a patient-specific solution would be preferred. In such a solution the patient's EEG has to be recorded for some time before starting the classification. Afterwards, the recorded data is used to train a classifier. Finally, the trained classifier is used to make decisions about the upcoming EEG. As new information is continuously becoming available, and the patients' seizures may evolve in morphology, the classifier must be updated online using the latest EEG data.

Some approaches of online learning models estimate a new model M_{t+1} using each new sample $\{x_s, y_s\}$ and the previous model M_t. We call such an approach an exhaustively updated online learning method. However, if the classifier is updated after the arrival of each new sample, the model grows proportionally with the number of analyzed samples. It is not possible to determine the number of samples *to be analyzed* in advance. Thus, it is not possible to estimate how large the model will become. The computational resources in any real system are limited and cannot store an infinite number of data samples.

Incremental Learning refers to online learning techniques that have limited memory resources [33]. With an incremental learning approach, the model updating must follow specific rules. First, the representation of the already analyzed data (classifier knowledge) must be compact (e.g., efficient statistics of the data, an alternative compact memory model) and free from redundancy to fit into the limited memory and maintain a fixed computational complexity. This property is called controlled memory growth [63]. Second, in the incremental

learning scenario, a model cannot grow indefinitely even if new samples are continually arriving. Hence, the algorithm must discard information during the updating process. This property is called forgetting capability. [63].

Among the existing classifiers, the SVM techniques are powerful tools to build learning models with excellent generalization capabilities even for sparse and high dimensional data. Several methods proposed in the literature that use SVM for NCES detection and classification [54, 62, 50], obtained good results, proving their usefulness for this task. Also, several incremental learning algorithms are proposed to extend the SVMs [14, 98, 85, 35]. Among these methods, the approximate techniques [92, 52, 99] have been demonstrated to be better suited for online epileptic seizure detection. These techniques discard the less relevant training vectors at each incremental step, thus reducing the memory requirements.

This Chapter presents an SVM-based incremental learning method which provides a good performance, comparable to that of the batch algorithm, while reducing the memory requirements. The incremental learning algorithm here presented is an extension of the method introduced in Chapter 5, from now referred as 'Batch method'. A hard threshold is used to reduce the number of support vectors selected to build the classifier decision function. This threshold provides a trade-off between performance and memory reduction. To introduce the incremental learning in the Batch method, three methodologies are tested to reduce the number of support vectors needed to build the decision function of the classifier: 1)Discard a fixed number of support vectors after the classifier retraining, based on the threshold defined. From now referred as 'Hard_IL', 2)Select an optimal support vector subset using cross-validation. From now referred to as 'Cross_IL', and 3). An approximate technique for incremental SVM proposed by [98] based on MRMR feature selection [25], referred in this text as 'MRMR_IL'. To evaluate the suitability of these incremental-learning approaches for NCES detection, first, a comparison between them is established. Secondly, the incremental learning approach with the best performance is compared with the Batch method and three batch methods from the literature [50, 72, 28] discussed in Chapter 5 (see Table 5.2), that reported better performance than the Batch method in some sense.

6.2 Incremental learning

In the literature, there are several ways to define incremental learning, and the term is not always used in the same context. This is the case with the algorithms denoted as online learning in this work. These algorithms use the training

samples one by one, without knowing the amount in advance, to optimize the classifier cost function. Different learning algorithms can be adopted, to perform the online training: 1) a fully online approach that adapts the model immediately after analyzing a single sample, 2) accumulate a small number of samples (mini-batch techniques) discarding them later, 3) or an entire batch learning approach, which stores all samples internally. The last two approaches work in batch mode for the inference of M_t by accumulating all samples in memory up to time step t. Incremental learning on the other hand, due to its limited memory resources counts on a compact representation of the already observed samples, and at the same time, it has to deliver accurate results for all relevant events that may arise from the data.

Despite the essential differences between them, incremental and online learning deal with similar challenges, in addition to the memory limitations. The fundamental dilemmas for these two learning models can be summarized as [33]:

- Concept drift: A concept drift occurs when the data distribution changes over time. Different types of concept drift may be identified depending on the location where the changes take place: if it is in the input data distribution, it is called virtual concept drift. If the changes occur in the underlying functionality itself, it is referred as real concept drift.
- The stability-plasticity: This dilemma is related to noisy environments and concept drift situations. In such a challenging setting, the question is when and how to update the model. A fast model updating enables high adjustment to new data but will neglect the past patterns equally fast. On the other hand, if it is performed slowly, retaining the old data for a longer time, the system adaption to the data changes is decreased. Many online learning approaches are not able to solve this dilemma since they exhibit the so-called catastrophic forgetting behavior.
- Adaptive model complexity: Since the data is unknown for incremental and online learning, it is impossible to estimate the model complexity in advance. Depending on how much the data change over time, a more complex model becomes necessary. However, the general model complexity is limited by the available resources. This situation requires a redistribution of the resources to ensure the proper functioning of the system.
- Efficient memory model. Given the limited resources availability of the incremental learning models, the observed data must be stored in a compact form at risk of being discarded. For both learning settings, the

memory adaptation must be carefully designed since it directly influences the stability-plasticity dilemma.

In the same way, as described in Chapter 5 the incremental learning approach introduced here, proposes to start the NCES monitoring after the clinicians identify the first epileptic seizure. The EEG recorded from the recording onset until the end of the first seizure is divided into nonoverlapping segments of 3s. Then, an SVM model is trained with the space signatures obtained from the CPD of all tensors built with HHT from the training sets defined in Chapter 4. The trained SVM becomes the first model M_t for the incremental learning algorithm. The purpose of the incremental learning, in this case, is to update M_t to provide adaptability to different EEG patterns within the same patient.

As stated before, the EEG signal is prone to morphology changes which can lead to a virtual or a real concept drift. The kind of concept drift that would appear depends on whether the changes prevail over time, or the previous state is regained. In such an environment, it is difficult to establish when and how to update the model.

Regarding to when to update the model, we propose to collect the EEG data during a time window of duration t and train a new classifier with this data. The number of times that this step is done will depend on the EEG recording duration and the chosen length for the time window. The value of t is selected arbitrarily as $t = 10\ min$ for short EEG recordings ($< 2\ h$) and $t = 2\ h$ for longer recordings ($> 2\ h$). These values are chosen evaluating the real chances of the appearance of morphology changes in the EEG. The clinicians establish the recording time for a patient considering how long it would take to register an epileptic event given the patient etiology. Hence, it is assumed that the changes emerge sooner in short EEG recordings and takes longer to show up in longer EEG recordings. The parameter t can be specified by the specialist at the start of the EEG monitoring. The data collected during this time is defined as a training round.

How to update the model is another nontrivial issue to solve in incremental learning problems. The new data samples $x_{n_{new}}$ become available over time, but no labels $y_{n_{new}}$ are provided for these data points. From the supervised training point of view, it is impossible to estimate M_{t+1} without known labels for $x_{n_{new}}$. The $x_{n_{new}}$ available for the estimation of M_{t+1} are samples previously classified by M_t. Hence, the labels estimated by M_t for $x_{n_{new}}$ are used as the labels $y_{n_{new}}$ for estimating the new model M_{t+1}. However, the labels estimated by M_t can be erroneous. If this is the case, then the error can be propagated into the classifier during the updating process, leading to a concept drift. Concept drifts are problematic since they lead to conflicts in the classification. The classifier performance will decrease until the model can be updated accordingly.

To reduce the chances of generating a concept drift with the labels provided by the model M_t,a double set of labels are predicted, one using the batch method, and additionally, another set using the PLS method [67]. The PLS prediction model, is trained with the x_n samples used to estimate the model M_t as feature matrix and their corresponding y_n labels as target. Then, the trained PLS model is used to predict the $y_{n_{new}}$ labels of the $x_{n_{new}}$ samples. The PLS model is implemented with online learning with a mini-batch approach, meaning the model is updated with the data collected in a time window of duration t. In order to verify the suitability of the PLS method within this context, we first tested its performance using a batch approach, prior to applying it for the double labeling in the incremental learning setting. The PLS model was tested on the test set of the 14 EEG recordings, it showed a specificity of 98.4% and a sensitivity of 97.9% predicting the labels of the EEG segments. Due to the double labeling, two sets of labels are available for $x_{n_{new}}$; $y_{n_{new1}}$ provided by the model M_t and $y_{n_{new2}}$ provided by PLS. The samples $x_{n_{new}}$ for which both methods estimate the same label ($y_{n_{new1}} = y_{n_{new2}}$) are selected for estimating and testing M_{t+1}.

The updating is performed in two steps. First, the model M_{t+1} is estimated only with the new samples $x_{n_{new}}$. In this case, the learning is not incrementing, but it is adapting to the new EEG patterns. As incremental learning algorithms do not store batches of previous data due to limited computational resources of the system, and because we are looking to add new EEG patterns while maintaining the old ones, a compact representation for old data must be found. Since the model used is an SVM, the support vectors from the model M_t are a condensed representation of the old data. Hence, finally the classification of the new samples can be done by using the SV from M_t and M_{t+1}. Then, after every training round, a new model is estimated, and new support vectors will be added to the solution.

After every update, the estimated model grows linearly with the number of new analyzed samples. In our specific case, the number of support vectors of the SVM will grow proportionally with the number of new training samples after every training round [14, 59, 63]. It is not possible to estimate how large the model will become if, as stated before, the number of times the model is updated depends on the recording duration. Therefore, it is necessary to take some actions to limit the number of support vectors in the solution. A regularization factor ϕ is introduced, to limit the growth of the number of support vectors. The factor ϕ is a positive integer between 0 and 1, defined as:

$$\phi = N_{sv_{M_t}}/N_{sv} \tag{6.1}$$

where $N_{sv_{M_t}}$ is the number of support vectors from the model M_t, and N_{sv} is

the sum of the number of support vectors from M_t and M_{t+1} . A $\phi = 0$ indicates that no new support vectors are selected for the solution (this is equivalent to the batch method). A $\phi = 1$ indicates that all available support vectors are included in the solution, implying that the number of support vectors in the model grow indefinitely. ϕ value will depend on the number of support vectors that become available after every training round. Applying ϕ factor means that the length of the selected subset of support vectors will be restricted in the following way,

$$\hat{N}_{sv} = \phi N_{sv} \tag{6.2}$$

where $\hat{N}_{sv}$ is the size of the final subset of support vectors. The factor ϕ controls the model growth, reducing the growth rate on the number of support vectors, but does not guarantee the optimality of the resulting subset. The three methodologies defined in the introduction as Hard_IL, Cross_IL, and MRMR_IL are tested, to select an optimal subset of support vectors. Fig. 6.2 describes all three incremental learning approaches implemented.

The new samples are divided into training and test set. A new model M_t is trained with the training set. The support vectors from the model the M_{t-1} and M_t will become the support vectors for the model M_{t+1}. To control the model growth, Hard_IL accumulates the support vectors of a number of training rounds, tr. Then, after tr training rounds, at every new training round, the oldest $N_{sv} - \hat{N}_{sv}$ support vectors are discarded to obtain the support vector subset for the model M_{t+1}. tr is a user-specified parameter for the number of training rounds after which the oldest support vectors should be removed. $tr = 1$, means that after every training round the oldest $N_{sv} - \hat{N}_{sv}$ support vectors are discarded. This could lead to a catastrophic forgetting behavior after some training rounds. The value of tr must establish a trade-off between the number and the redundancy of available support vectors in the set after each training round determined by the size of the training window t. The larger the value of t, the more support vectors become available increasing the model complexity. For small values of t the model growth is decreased but the redundancy of the support vector set may be larger. If the classifier is updated in short intervals, there are higher probabilities of modeling the same EEG patterns, adding redundant support vectors to the classifier. Hence, it is convenient to set small values for tr. For this implementation, $tr = 2$ was selected. As can be appreciated in Fig. 6.1 for $tr > 2$ the execution time increases without significant improvements in the algorithm accuracy. Fig. 6.2 B describes the Hard_IL algorithm. Finally, the models M_{t-1} and M_{t+1} are tested in the test set, and the model with the best performance will be used for the classification until the next update of the classifier.

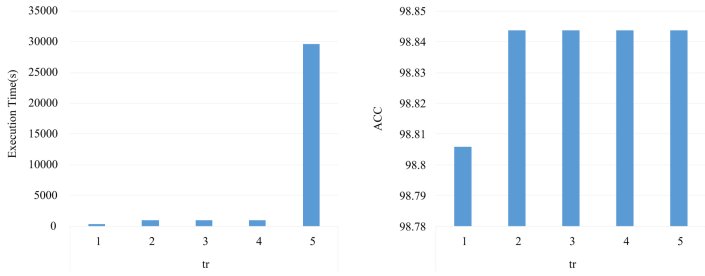


Figure 6.1: Hard_IL execution time (left) and accuracy (right) values obtained for different values of tr. The test was conducted for the recording number 10 which is the one that completes more training rounds.

For Cross_IL the labeled new samples $x_{n_{new}}$ are split into a training and test set. A model M_t is trained with the training set. Given the support vectors from the models M_{t-1} and M_t a 5-fold (of length $\hat{N}_{sv}$) cross-validation is applied to find the optimal subset of support vectors. This means that the set of all available support vectors is shuffled into five different subsets. This is implemented by randomly splitting the support vectors in five sets so that the size of these five sets is $\hat{N}_{sv}$. Hence, performance of the five subsets and the model M_{t-1} is evaluated on the test set. The model with the best performance is used to update the current classifier. Fig. 6.2 C illustrate the diagram of the Cross_IL algorithm.

In the MRMR_IL approach, as implemented for Cross_IL the labeled new samples $x_{n_{new}}$ are split in a training and test sets, and a model M_t is estimated from the training set. Then the most relevant support vectors from the model M_t and M_{t-1} are selected as support vectors of the model M_{t+1}. The MRMR_IL approach implements the method proposed in [98] to estimate the optimal subset of support vectors. In [98], Yang et al. proposed an incremental SVM algorithm where the selection of an optimal subset of support vectors for the SVM is formulated as equivalent to a feature selection problem. This is indeed the case, if the SVM model is written in matrix form as

$$y = w^T \tilde{K} + w_0 \tag{6.3}$$

where w^T is the weight vector, $\tilde{K} = [k(x, x_1), \ldots, k(x, x_{N_{sv}}]^T$ is the kernel matrix with N_{sv} candidate support vectors.

Equivalently, in the feature selection domain for the linear case we want to find the feature subset that maximizes the function

$$f(s) = \beta^T x + b_0 \tag{6.4}$$

where $x = [x_1, \ldots, x_{\hat{N}_{sv}}]$ with x_i as selected feature value, β^T the corresponding weight vector and b_0 is the bias term.

By exploiting the similarities between (6.3) and (6.4) the kernel matrix $K \in \mathbb{R}^{N_{sv} \times N_{sv}}$ can be interpreted as a data matrix in the feature selection context where each row of K corresponds to an observation and each column of K corresponds to a feature (support vectors). Hence, the number of support vectors can be reduced by dropping columns in the kernel matrix K, while keeping the number of rows unchanged [98]. To select the best subset of support vectors, Yang proposed the feature selection MRMR introduced in [25].

Following the MRMR scheme to find the optimal subset of support vectors, the first step is to select the column of K with the highest F_i computed with (3.32) (Section 3.4.2). This column is the first element of the final subset. At each iteration, the rest of the unselected support vectors are evaluated with (3.32), and the one with the largest F_i is added to the subset. The subset is evaluated according to (3.31) and (3.34) (Section 3.4.2), and the one that maximizes the relation in (3.35) (Section 3.4.2) is selected. Finally, the models M_{t-1} andM_{t+1} are tested in the test set and the model with the best performance is selected to update the classifier.

As mentioned before, it is crucial to balance how much to learn and how much to forget. About the forgetting rate of this system, it should be noticed that the model M_{t+1} keeps all relevant support vectors from the previous models M_t and $M_t - 1$. This characteristic of the model updating allows that even if there is a fast model adaptation to new patterns, the old patterns will be present in the model if they are relevant for the classification. Fig. 6.2 D illustrate the algorithm implementation.

6.2.1 Alternative batch approaches

The performance of Hard_IL, Cross_IL, and MRMR_IL is first mutually compared, to establish the best method to select the subset of support vectors. The method with the best performance is then compared with the Batch method and the batch methods proposed in [50],[72] and [28] (from now referred in the text as 'Koliallil', 'Sharma' and 'Fatma' respectively). Since the Batch method is the basis for all incremental learning methods proposed here, it is logical to compare their performances to assess the achieved improvement, if there is any.

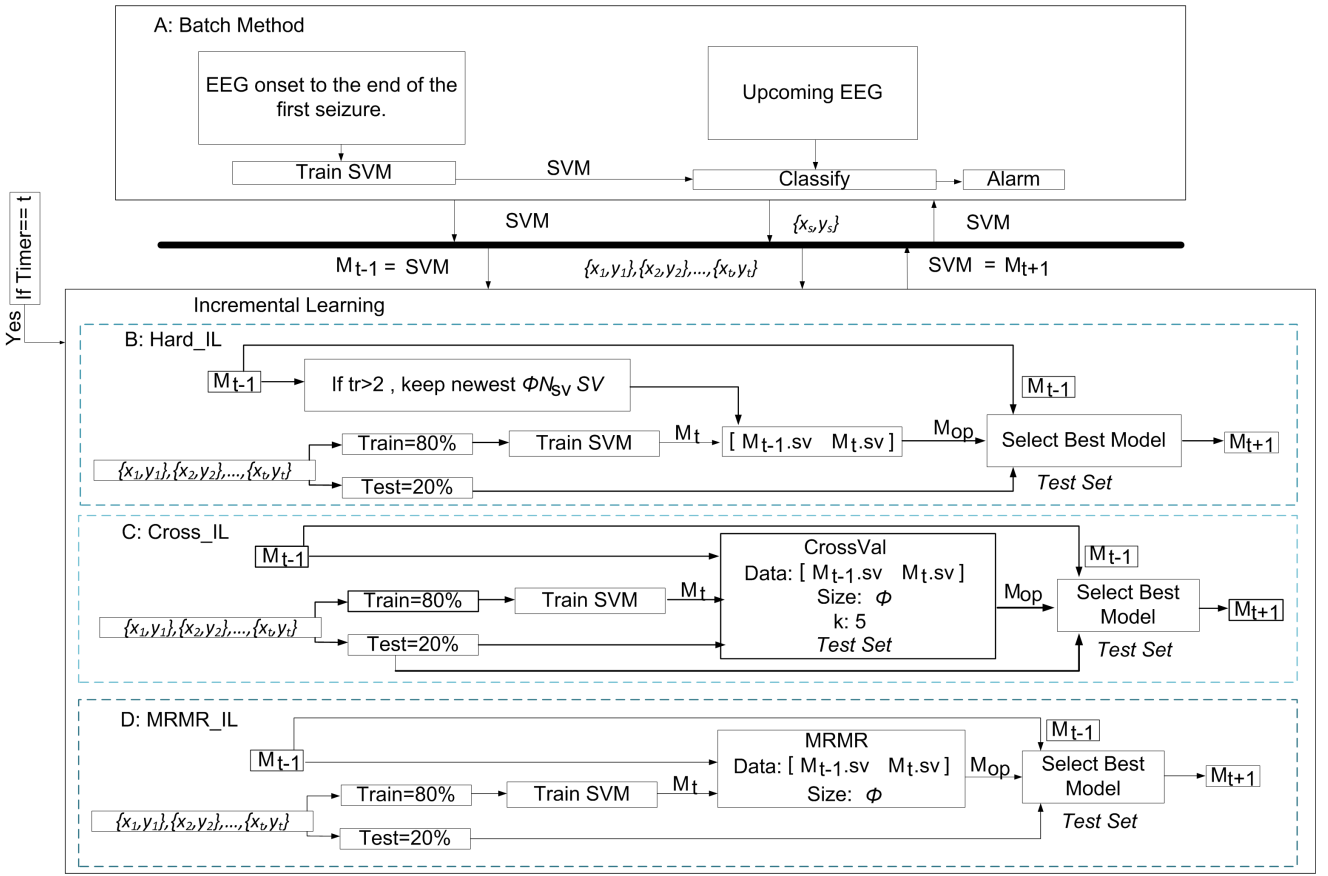


Figure 6.2: Flow chart of the NCES detection method with incremental learning. Section (A) describes the Batch method (see Fig. 5.1 for details). The three implementations of incremental learning approaches tested: HardIL, Cross_IL, and MRMR_IL are described in sections (B),(C), and (D) of the diagram. For the graph simplicity, $\{x_1, y_1\}, \{x_2, y_2\}, \ldots, \{x_t, y_t\}$ refers to the samples resulting from the double labeling. SV stands for support vector.

The three methods, [50],[72] and [28], are implemented and tested on our data. For each implementation, we tune the parameters to maximize the classification performance. We include some minimal variations in the training to tune the methods to the available dataset. First, the training set is composed of the first NCES and the same number of epochs of n-NCES EEG prior to it, as described in Section 4.3. For Sharma and Fatma, the EEG was segmented using the epoch length defined by the authors ($1s$). In the case of the Kollialil method, two ways of training are implemented. First, the classifier is trained with all data from 7 of the 14 cases available, and the unseen cases are used to test the classifier. The cases are selected in such a way that each set had ambulatory and ICU cases. Second, given that this method is not patient-specific, the training set is composed of all training sets defined in Table 4.2 assembled. These variations are included to ensure equality of conditions for all methods using the same number of patterns for the training stage. The performance of all methods is assessed in terms of the sensitivity, specificity, PPV and false positives per hour.

6.3 Results and Discussion

6.3.1 Incremental Learning Approaches Comparison

All incremental learning approaches, as can be appreciated in Table 6.1 outperformed the Batch method in at least one of the metrics assessed.

The Hard_IL approach showed an increase in specificity, PPV, and a lower number of false detections per hour compared to the Batch method. However, the method also experiments a drop in sensitivity. The arbitrary removal of support vectors seems to cause the forgetting of relevant information for the NCES classification. It is possible that for more extended recordings, the catastrophic forgetting phenomenon could appear if the EEG morphology changes too fast.

Cross_IL sensitivity values are also lower than the ones obtained by the Batch method, yet displayed an increase in specificity and PPV. This method displayed the lowest number of false detections with only 1.6 false positives per hour. The results obtained with Cross_IL demonstrate to be unstable since they depend on the support vector subset resulting after the cross-validation process (not always the same, and not always better). The performed tests show that the winning subset does not always show the best performance for new samples.

The MRMR-IL displayed values of specificity, sensitivity above 99% and a PPV of 90.1 %. The low average PPV value, despite the high sensitivity and specificity values, may be because the test sets are extremely unbalanced in

some cases, and the PPV is very sensitive to false detections. For this reason, even when sensitivity and specificity values are close to 100%, the PPV is not as good as expected. For example, the test set of case 8, has only seven positive samples and 4809 negatives. If the classifier has 99.0% specificity, this represents already 48 false positive detections, this results in a low PPV value, even if the specificity value is as good as 99%. In general, the use of unbalanced test sets results in low PPV values. The MRMR_IL generates an average of 2.4 false detections per hour, 3.1 less than the Batch method.

Regarding the errors made by the algorithm during the classification, it was found that the epochs misclassified as false positives corresponded to preictal activity, classified as seizure activity (in the case of the false positives). The preictal activity that appears long before seizure onset (say $4 - 6$ s), is not marked as seizure activity. On the other hand, the end of the seizure is marked immediately after the postictal activity. As can be seen in Fig. 6.3, the patterns of the preictal and postictal activity are very similar. Postictal activity is always labeled as seizure by the doctors. Hence, if the patterns of the postictal activity (similar to preictal) are considered as relevant by the classifier and are added to the training set, later on, the preictal activity could also be recognized as seizure activity.

Some tests were also performed to investigate whether the double labeling was improving the classification. The difference in the classification efficiency can be appreciated in Table 6.2 which displays the false positive and negative detections with and without double labeling. As can be seen in 6.2the double labeling represented a major improvement for some cases. Based on the results obtained, it is our believe that the double labeling in longer recordings will represent a strength for the incremental learning approaches preventing the classifier updating from generating a concept drift.

False negative detections were found for records 1 and 9. The false negatives identified are associated with the errors committed by the classifier before the updating. Every error committed by the classifier before the defined updating time will affect the MRMR_IL performance.

The execution time for the MRMR-IL approach is expected to be the same as the Batch method since the model updating is performed in parallel to the classification. Therefore, the execution time of this algorithm is assumed to be in the same time range as the Batch method (0.37 to 3 s for the classification task). The classifier retraining execution time goes from 0.03 to 14.36 s for $t = 10$ min and from 0.94 s to 1.79 min for $t = 2$ h. This means that the highest delay in the classifier updating is approximately 14.36 s and 1.79 min for a time window of 10 min and 2 h respectively.

Table 6.1: Average Specificity, Sensitivity, PPV, and FP per hour values obtained for the Batch Method and the three implemented incremental learning algorithms. In the table Spec, and Sen stand for Specificity and Sensitivity metrics respectively.

	Batch Method				Hard_IL				Cross_IL				MRMR_IL			
Case	Spec	Sen	PPV	FP/h	Spec	Sen	PPV	FP/h	Spec	Sen	PPV	FP/h	Spec	Sen	PPV	FP/h
1	97.4	96.2	85.0	13.9	98.5	71.7	88.4	7.7	99.1	56.6	90.9	4.6	98.2	98.1	89.7	9.3
2	100	100	100	0	100	100	100	0	100	100	100	0	100	100	100	0
3	99.3	97.6	97.4	3.7	99.3	98.3	97.5	3.7	100	100	100	0	100	100	100	0
4	99.1	100	98.4	4.3	99.1	100	98.4	4.3	100	100	100	0	100	100	100	0
5	100	100	100	0	100	100	100	0	100	100	100	0	100	100	100	0
6	94.8	100	83.8	19.7	96.2	100	87.0	15.2	98.5	100	94	6.1	96.5	100	88.2	18.2
7	100	100	100	0	100	100	100	0	100	100	100	0	100	100	100	0
8	99.4	100	26.3	3.5	99.7	100	40.0	2.0	99.9	100	58.8	0.9	99.9	100	58.8	0.9
9	98.7	89.9	70.4	7.8	99.3	94.3	82.9	4.0	99.8	96.9	95.7	1.2	99.8	94.3	94.3	1.2
10	99.1	99.9	86.5	8.3	97.9	100	93.7	3.6	99.1	100	97.3	0.8	99.07	100	97.1	1.6
11	99.5	100	11.4	7.5	99.1	100	20.0	4.5	99.6	100	33.3	2.2	99.6	100	33.3	2.2
12	100	100	99.8	0.0	100	100	100	0	100	100	100	0	100	100	100	0.0
13	98.3	100	90.8	8.9	98.3	100	90.8	8.9	99.5	100	97.2	6	100	100	100	0.0
14	100	100	100	0	100	100	100	0	100	100	100	0	100	100	100	0
Average	99.0	98.8	82.1	5.5	99.1	97.4	85.6	3.8	99.7	96.7	90.5	1.6	99.5	99.5	90.1	2.4

Table 6.2: Number of false positive and negative detections of the MRMR_IL method with and without double labeling, compared to the erroneous detections made by the Batch Method.

	Batch Method		MRMR_IL			
			Single Labeling		Double Labeling	
Case	FN	FP	FN	FP	FN	FP
1	2	9	2	9	1	6
2	0	0	0	0	0	0
3	3	3	3	2	0	0
4	0	1	0	0	0	0
5	0	0	0	0	0	0
6	0	13	0	13	0	12
7	0	0	0	0	0	0
8	0	28	0	22	0	7
9	16	60	12	53	3	9
10	1	163	1	144	0	31
11	0	27	0	16	0	8
12	1	0	2	0	0	0
13	0	7	0	0	0	0
14	0	0	0	1	0	0

6.3.2 MRMR_IL and Batch methods Comparison

Table 6.3 lists the classification results of the Batch, MRMR_IL, Kollialil, Sharma and Fatma methods for the 14 cases available in the database. The MRMR-IL clearly outperformed the Batch method in all metrics. It should be noted that the performance of the Batch method for this database was already high before including incremental learning, which does not leave much room for improvement. The MRMR_IL show the same performance as the Batch method for the 5 cases for which the classification outcome was perfect, and improved the classification results for the other nine. However, the short duration of the recordings difficult the assessment of the real capabilities of the MRMR_IL approach for NCES detection. MRMR_IL should be tested in a more extensive database with longer EEG recordings to assess the statistical significance of the improvement achieved with the addition of the incremental learning to the Batch method.

The MRMR_IL yields the best results among the compared methods, showing the highest average performances for all metrics across the patients. According to the paired t-test, the MRMR_IL improvement was statistically significant compared to Kollialil, Sharma, and Fatma methods. From the two training processes performed for Kollialil, the first one (using 7 cases for training and the rest for testing) (Kollialil_1), showed results below 70% for all metrics. The second training (assembling the predefined training sets in Table 4.2)(Kollialil_2), increased the performance significantly regarding the specificity and PPV. The sensitivity remained the same for both trainings (67.0%). The results obtained, can be explained by two possible causes. First, the training set was not big enough so that the classifier could learn and perform the classification correctly. Second, the classifier is not able to recognize patterns that are not used during training, when using the Inter-quartile Range (IQR) as feature.

The MRMR-IL outperformed the Sharma method in all metrics assessed except for the sensitivity, where Sharma displayed a 100% average value. The performance of the Sharma method decays in patients for which the training set is extremely unbalanced (case 3) or too small (cases 7, 8 and 9). The method obtained the lowest specificity values and PPV for these cases.

The Fatma method was inferior to MRMR-IL for all the evaluated metrics. This approach achieved an average specificity of 93.4% with a PPV of 41.3%. The algorithm displayed the lowest average sensitivity over all tested methods (around 62%). Since this method does not use classifiers, the only plausible reason for the low sensitivity outcome is that the threshold for the seizure detection must be adapted for each of the datasets.

Concerning the execution time, the Kollialil method outperformed MRMR-IL. The algorithm from Kollialil is executed in 0.28 s to 0.19 s for the classification process of the 3 s epochs. An outlier of 25.25 s was observed for one of the epochs of Case 10. This outlier was disregarded in the time performance analysis. The superiority of Kollialil is given by the fact that the tensorization and the tensor decomposition processes are more time consuming than the IQR range computation. Nevertheless, based on our results, we believe that both, MRMR-IL and its batch method will perform properly in a real-time monitoring setting.

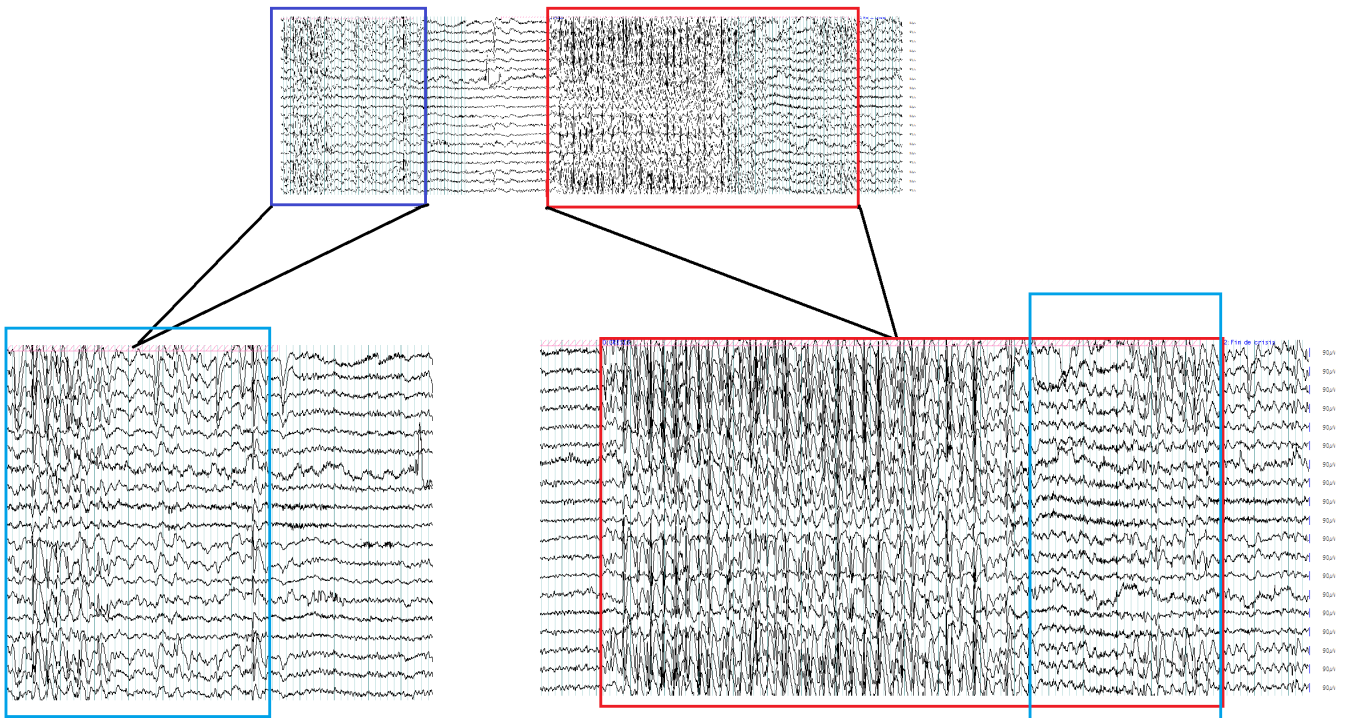


Figure 6.3: Misclassified epoch by The Incremental Learning approach from Case 6. The red box indicates seizure activity. The blue boxes mark the preictal and postictal EEG patterns. As can be observed the patterns from both preictal and postictal activities are quite similar. Since the classifier considers the postictal activity as seizure, this similarity could cause the classifier failure.

Table 6.3: Specificity, Sensitivity, and PPV, scores obtained in the classification for MRMR-IL, Kollialil(1 and 2), Sharma ,and Fatma. In the table Spec, and Sen stand for Specificity and Sensitivity metrics respectively.

	Batch Method			MRMR_IL			Kollialil _1			Kollialil_2			Sharma			Fatma		
Case	Spec	Sen	PPV	Spec	Sen	PPV	Spec	Sen	PPV	Spec	Sen	PPV	Spec	Sen	PPV	Spec	Sen	PPV
1	97.4	96.2	85.0	98.2	98.1	89.7	22.0	58.2	7.1	94.9	29.1	39.1	93.5	100	53.3	96.9	65.5	61.7
2	100	100	100	100	100	100	55.0	57.9	9.3	54.4	94.7	86.5	96.4	100	75.9	84.2	97.4	38.5
3	99.3	97.6	97.4	100	100	100	35.7	63.9	13.7	97.2	5.3	25.9	2.8	100	13.5	85.7	52.8	37.0
4	99.1	100	98.4	100	100	100	68.6	56.5	23.9	56.4	98.6	26.3	81.7	100	60.3	80.9	93.7	60.6
5	100	100	100	100	100	100	26.6	62.5	2.5	96.3	100	7.5	34.4	100	3.9	97.1	81.5	44.0
6	94.8	100	83.8	96.5	100	88.2	22.4	68.6	10.6	62.9	100	15.2	88.5	100	47.7	87.8	94.5	48.0
7	100	100	100	100	100	100	67.8	60.0	0.9	79.7	80.0	1.9	54.7	100	1.3	98.3	31.3	14.7
8	99.4	100	26.3	99.9	100	58.8	54.3	90.0	0.1	62.2	90.0	0.2	25.8	100	0.1	99.4	65.4	9.5
9	98.7	89.9	70.4	99.8	94.3	94.3	36.5	50.7	1.4	94.7	52.1	7.1	20.4	100	2.0	96.6	76.9	26.6
10	99.1	99.9	86.5	99.1	100	97.1	77.3	57.2	56.6	99.5	16.8	53.1	97.1	100	65.3	89.3	86.9	30.7
11	99.5	100	11.4	99.6	100	33.3	37.8	66.7	3.3	99.6	100	9.4	98.4	100	66.7	99.7	88.9	88.9
12	100	99.9	99.8	100	100	100	93.8	84.0	5.6	99.6	95.4	53.3	89.8	100	23.8	100	24.8	99.1
13	98.3	100	90.8	100	100	100	51.7	73.6	9.4	75.5	59.8	11.1	86.8	100	37.4	95.4	3.8	7.6
14	100	100	100	100	100	100	68.9	88.4	3.9	99.8	18.6	7.2	95.8	100	46.4	96.6	12.3	11.4
Average	99.0	98.8	82.1	99.5	99.5	90.1	51.3	67.0	10.6	83.8	67.2	24.5	69.0	100	35.6	93.4	62.6	41.3

6.4 Conclusion

This Chapter proposed a method that uses incremental learning to improve the NCES monitoring. The proposed algorithm, namely MRMR-IL is based on the solution introduced in Chapter 5. The MRMR-IL retrained the classifier periodically to improve the seizure recognition in case of changes in the EEG morphology over time. The obtained results show that the MRMR-IL outperformed the Batch method.

A double labeling scheme is proposed to prevent the generation of concept drift. The double labeling is shown to improve the MRMR_IL performance.

Three detection methods proposed by 3 other authors Kollialil, Sharma , and Fatma were evaluated in the available database and compared to MRMR_IL. MRMR_IL was shown to outperform the three methods in all measured metrics.

The MRMR_IL method proved to be an effective tool for NCES monitoring in a real-time setting which might represent an enhancement in the detection of NCES and NSE diagnosis. For further application, it is necessary to test the method with new and more extensive data, with longer duration.

Chapter 7

Conclusions

This thesis introduced novel EEG-based techniques for real-time detection of NCES and early alert of the NCSE. Such methods can be useful in various medical settings, providing support to the clinicians for the diagnosis of this medical emergency. In this chapter we review the thesis contributions. The limitations of the research and methods are discussed, and some directions for future work are proposed to address these limitations.

7.1 Summary of obtained results

This research introduced a novel EEG-based NCES detection algorithm. The analysis was done by expanding EEG epochs of 3 s duration using the HHT and building third-order tensors, with modes "$frequency \times time \times channels$". The signatures from each mode are computed using CPD. It was shown that the EEG patterns analyzed in epochs of $3s$ duration satisfy the trilinear model property required for CPD. This partially explains why a BTD approach could not improve the results. A SVM model was trained with the extracted signatures to detect the NCES activity.

The above explained combined methodology is the basis of the seizure detection method proposed in this thesis. The method was evaluated using a test dataset consisting of $\approx$ 57 h of scalp EEG recordings from 14 patients from diverse medical settings diagnosed for epilepsy with a total of 139 seizures, resulting in an overall sensitivity, specificity and accuracy of $\approx 99\%$ respectively, along with a false detection rate of $\approx 0.014/h$. Apart from the selection of the first seizure

that appears in the recording, the proposed method is fully automated.

According to the assessment done, the false detections of the proposed method resulted mostly from appearing paroxysms longer than 0.5 s or a large number of successive spikes in the EEG. These events are very similar to a crisis, but doctors do not consider them as such, due to their short duration, although their appearance is associated with epileptic disorders in the patient. False positive detections also occur at epochs with preictal or postictal EEG activity. In addition, the results show that most EEG epochs of the epileptic seizures missed by this algorithm exhibit a significantly different pattern from the first seizure provided as the training patterns for the classifier training.

The MRMR_IL algorithm proposed in Chapter 6 is an extension of the NCES detection approach described in Chapter 5 in order to deal with false negatives due to morphological changes in the EEG patterns, and shown to improve the overall performance. It uses an SVM incremental learning method combined with an MRMR approach to reduce the number of support vectors selected to update the classifier. This method was shown to outperform other methods proposed in the literature for NCES detection and the method from Chapter 5. More extensive testing must be carried out to assess the real statistical significance of this improvement. The data available are mostly from short-term recordings, which hinders the demonstration of the long-term real benefits of the incremental learning.

A significant issue of the MRMR_IL is estimating the size of the retraining window (time after which the classifier must be retrained). This parameter is shown to influence the method performance. Here we proposed a window of $t = 10\ min$ for recordings lasting less than 2 h and $t = 2\ h$ for longer recordings, based on the analyzed data. The specification of this parameter will depend on the objectives of the doctor with the monitoring. The physician could determine the time of retraining according to the estimated evolution of the patient. Longer windows must be set for patients with slow clinical evolution (e.g. comatose patients) and shorter ones for patients with fast clinical evolution (e.g. ambulatory patients with absence seizure disorders).

The strength of our proposed methods is proven by comparing it with other published NCES detection methods tested on scalp EEG recordings. The methods proposed in the literature were implemented and tested on our data to establish a fair comparison and avoid making conclusions based on different databases. The comparison confirmed that the methods proposed here demonstrates a highly acceptable performance as it detects seizures with high sensitivity and specificity, while the computational cost is low enough for the real-time implementation.

The seizure detection methods proposed in this thesis provided satisfactory results even when using a relatively small training dataset (the training data were selected using the EEG recorded from the beginning to the end of the first seizure). The execution time needed by the algorithm to perform the classification was lower or equal to the duration of the EEG epoch analyzed. Therefore, from a practical point of view, the proposed method has the potential to perform properly in a routine clinical setting. It is also worth noting that although generic seizure detection methods do not need training when applied to a particular patient [50], they must be developed/trained using a very large dataset including various seizure types and non-seizure patterns to be sufficiently general for clinical applications. These large databases are not available for NCES.

7.2 Future work and challenges

Although the proposed methods show potential to be used in routine clinical environments for automated seizure detection, more studies are required to improve the method and address its limitations. The majority of the scalp EEG recordings used in this work comes from an ambulatory setting; this means that most of the recordings are short-term. None of the recordings used has NCSE. Therefore the alarm system for NCSE remains untested. Hence, a necessary future step is to evaluate the proposed methods extensively using EEG data of longer duration to assess their real potential. The future datasets should include more recordings from ICU patients to better investigate the performance of the methods on NCES from brain-injured patients, and the alarm system on recordings with NCSE.

NCES is very common in critically ill children and is associated with unfavorable outcome (especially in patients with high seizure burden) despite the clinical effort to manage them. In a study conducted by Jette et al. in [45] in a population of 117 critically ill children, 44% of them were diagnosed as having NCES (detected on cEEG). The methods proposed and discussed in this thesis are developed and tested on EEG obtained from adult patients. However, their usefulness for critically ill children must be considered. The extension and validation of the proposed methods to the Pediatric Intensive Care Unit (PICU) and Neonatal Intensive Care Unit (NICU) could enhance the management of the NCES improving the outcome of the children hospitalized in this units.

Predicting and preventing epileptic seizures is the ultimate goal to achieve with a real-time monitoring system. The occurrence of NCES is associated with an increased risk for developing subsequent epilepsy. To determine whether the

early identification and management of non-convulsive epileptogenesis reduce the risk of the appearance of NCES and NCSE, early markers (EEG patterns or other biomarkers) of epileptic activity need to be identified. Hence further research is needed, even if it is not demonstrated that treating NCES symptoms reduces the chances of the later appearance of the seizure. We believe that a valuable addition to the methods proposed in this thesis would be a seizure prediction algorithm. Such an algorithm would allow conducting studies about the impact of treating the symptoms of NCES in the patient's prognosis.

The diagnosis of NCES and NCSE is often hampered by the lack of clarity of the EEG patterns, especially those from the ICU. Therefore, a seizure detection method based only on the interpretation of these EEG patterns is expected to show low performance in recordings where these patterns are unclear. To better understand these EEG patterns and their meaning of both global and patient-specific levels, a combination of the continuous EEG with other physiological parameters is needed. It is our firm belief that the performance of our proposed methods can be further improved by multimodal monitoring, e.g. by including other parameters not derived from the EEG (e.g., ECG).

The methods proposed here, will only be beneficial for patient care when they are incorporated into monitoring devices and are used extensively in medical services. The further integration and implementation of our proposed methods in a monitoring device must be worked out in the near future to make them available to physicians.

Bibliography

[1] Abdi, H. Partial least squares regression and projection on latent structure regression (pls regression). *Wiley interdisciplinary reviews: computational statistics 2*, 1 (2010), 97–106.

[2] Acar, E., Aykut-Bingol, C., Bingol, H., Bro, R., and Yener, B. Multiway analysis of epilepsy tensors. *Bioinformatics 23*, 13 (2007), 10–18.

[3] Adeli, H., Zhou, Z., and Dadmehr, N. Analysis of eeg records in an epileptic patient using wavelet transform. *Journal of neuroscience methods 123*, 1 (2003), 69–87.

[4] Aherne, F. J., Thacker, N. A., and Rockett, P. I. The bhattacharyya metric as an absolute similarity measure for frequency coded data. *Kybernetika 34*, 4 (1998), 363–368.

[5] Ansari, A. Q., and Sharma, P. A review on automated detection of non-convulsive seizures using eeg. In *Computational Intelligence & Communication Technology (CICT), 2016 Second International Conference on* (2016), IEEE, pp. 283–286.

[6] Atamon, E. A. *Understanding epilepsy seizure structure using tensor analysis*. Rensselaer Polytechnic Institute, 2008.

[7] Beniczky, S., Hirsch, L. J., Kaplan, P. W., Pressler, R., Bauer, G., Aurlien, H., Brøgger, J. C., and Trinka, E. Unified eeg terminology and criteria for nonconvulsive status epilepticus. *Epilepsia 54*, s6 (2013), 28–29.

[8] Bhattacharyya, A., and Pachori, R. B. A multivariate approach for patient-specific eeg seizure detection using empirical wavelet transform. *IEEE Transactions on Biomedical Engineering 64*, 9 (2017), 2003–2015.

[9] Bhattacharyya, A., Pachori, R. B., Upadhyay, A., and Acharya, U. R. Tunable-q wavelet transform based multiscale entropy measure for automated classification of epileptic eeg signals. *Applied Sciences 7*, 4 (2017), 385–403.

[10] Bicciato, S., Luchini, A., and Di Bello, C. Pca disjoint models for multiclass cancer analysis using gene expression data. *Analytica chimica acta 595*, 1 (2007), 43–50.

[11] Blum, A. S., and Rutkove, S. B. *The clinical neurophysiology primer*, vol. 388. Springer, 2007.

[12] Brereton, R. G., and Lloyd, G. R. Partial least squares discriminant analysis: taking the magic away. *Journal of Chemometrics 28*, 4 (2014), 213–225.

[13] Bustamante, I. T., Mata, F. S., Gonz´alez, N. H., Gazapo, R. G., Palau, J., and Ferreira, M. M. C. Application of chemometric tools for automatic classification and profile extraction of dna samples in forensic tasks. *Bioinformatics 19*, 5 (2003), 571–578.

[14] Cauwenberghs, G., and Poggio, T. Incremental and decremental support vector machine learning. In *Advances in neural information processing systems* (2001), pp. 409–415.

[15] Chandrashekar, G., and Sahin, F. A survey on feature selection methods. *Computers & Electrical Engineering 40*, 1 (2014), 16–28.

[16] Chang, B. S. *Atlas of ambulatory EEG*. Elsevier, 2005.

[17] Chang, N.-F., Chen, T.-C., Chiang, C.-Y., and Chen, L.-G. On-line empirical mode decomposition biomedical microprocessor for hilbert huang transform. In *Biomedical Circuits and Systems Conference (BioCAS), 2011 IEEE* (2011), IEEE, pp. 420–423.

[18] Cichocki, A., Mandic, D., De Lathauwer, L., Zhou, G., Zhao, Q., Caiafa, C., and Phan, H. A. Tensor decompositions for signal processing applications: From two-way to multiway component analysis. *IEEE Signal Processing Magazine 32*, 2 (2015), 145–163.

[19] Claassen, J., Perotte, A., Albers, D., Kleinberg, S., Schmidt, J. M., Tu, B., Badjatia, N., Lantigua, H., Hirsch, L. J., Mayer, S. A., et al. Nonconvulsive seizures after subarachnoid hemorrhage: multimodal detection and outcomes. *Annals of neurology 74*, 1 (2013), 53–64.

[20] Dabbaghchian, S., Ghaemmaghami, M. P., and Aghagolzadeh, A. Feature extraction using discrete cosine transform and discrimination power analysis with a face recognition technology. *Pattern Recognition 43*, 14 (2010), 1481–1440.

[21] De Lathauwer, L. Decompositions of a higher-order tensor in block terms—part ii: Definitions and uniqueness. *SIAM Journal on Matrix Analysis and Applications 30*, 3 (2008), 1033–1066.

[22] De Lathauwer, L., and Nion, D. Decompositions of a higher-order tensor in block terms—part iii: Alternating least squares algorithms. *SIAM journal on Matrix Analysis and Applications 30*, 3 (2008), 1067–1083.

[23] De Sa, J. M. *Pattern recognition: concepts, methods and applications*. Springer Science & Business Media, 2012.

[24] Deburchgraeve, W., Cherian, P. J., De Vos, M., Swarte, R. M., Blok, J. H., Visser, G. H., Govaert, P., and Van Huffel, S. Neonatal seizure localization using parafac decomposition. *Clinical Neurophysiology, 120*, 10 (2009), 1787–1796.

[25] Ding, C., and Peng, H. Minimum redundancy feature selection from microarray gene expression data. *Journal of bioinformatics and computational biology 3*, 02 (2005), 185–205.

[26] Dubuisson, S. The computation of the bhattacharyya distance between histograms without histograms. In *Image Processing Theory Tools and Applications (IPTA), 2010 2nd International Conference on* (2010), IEEE, pp. 373–378.

[27] Duin, R., Juszczak, P., Paclik, P., Pekalska, E., De Ridder, D., Tax, D., and Verzakov, S. Prtools 4.1. *A Matlab Toolbox for Pattern Recognition, Software and Documentation downloaded May* (2010).

[28] Fatma, T., Farooq, O., Khan, Y. U. Tripathi, M., and Sharma, P. Automatic detection of non-convulsive seizures: A reduced complexity approach. *Journal of King Saud University-Computer and Information Sciences 28*, 4 (2016), 407–415.

[29] Fiest, K. M., Sauro, K. M., Wiebe, S., Patten, S. B., Kwon, C.-S., Dykeman, J., Pringsheim, T., Lorenzetti, D. L., and Jetté, N. Prevalence and incidence of epilepsy a systematic review and meta-analysis of international studies. *Neurology 88*, 3 (2017), 296–303.

[30] Fisher, R. S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J. H., Elger, C. E., Engel, J., Forsgren, L., French,

J. A., Glynn, M., et al. Ilae official report: a practical clinical definition of epilepsy. *Epilepsia 55*, 4 (2014), 475–482.

[31] Fisher, R. S., Boas, W. v. E., Blume, W., Elger, C., Genton, P., Lee, P., and Engel, J. Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ilae) and the international bureau for epilepsy (ibe). *Epilepsia 46*, 4 (2005), 470–472.

[32] Fisher, R. S., Cross, J. H., French, J. A., Higurashi, N., Hirsch, E., Jansen, F. E., Lagae, L., Moshé, S. L., Peltola, J., Roulet Perez, E., et al. Operational classification of seizure types by the international league against epilepsy: Position paper of the ilae commission for classification and terminology. *Epilepsia 58*, 4 (2017), 522–530.

[33] Gepperth, A., and Hammer, B. Incremental learning algorithms and applications. In *European Symposium on Artificial Neural Networks (ESANN)* (2016), ESANN, 2016, Bruges, Belgium, pp. 1–13.

[34] Glauser, T., Shinnar, S., Gloss, D., Alldredge, B., Arya, R., Bainbridge, J., Bare, M., Bleck, T., Dodson, W. E., Garrity, L., et al. Evidence-based guideline: treatment of convulsive status epilepticus in children and adults: report of the guideline committee of the american epilepsy society. *Epilepsy currents 16*, 1 (2016), 48–61.

[35] Gu, B., Sheng, V. S., Tay, K. Y., Romano, W., and Li, S. Incremental support vector learning for ordinal regression. *IEEE Transactions on Neural networks and learning systems 26*, 7 (2015), 1403–1416.

[36] Guerrero-Mosquera, C., Verleysen, M., and Vazquez, A. N. Eeg feature selection using mutual information and support vector machine: A comparative analysis. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE* (2010), IEEE, pp. 4946–4949.

[37] Hastie, T., Friedman, J., and Tibshirani, R. Boosting and additive trees. In *The Elements of Statistical Learning*. Springer, 2001, pp. 299–345.

[38] Hirsch, L., and Brenner, R. *Atlas of EEG in critical care*. John Wiley & Sons, 2011.

[39] Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the Royal Society*

of London A: Mathematical, Physical and Engineering Sciences (1998), vol. 454, The Royal Society, pp. 903–995.

[40] Hunyadi, B., Camps, D., Sorber, L., Van Paesschen, W., De Vos, M., Van Huffel, S., and De Lathauwer, L. Block term decomposition for modelling epileptic seizures. *EURASIP Journal on Advances in Signal Processing 2014*, 1 (2014), 139–158.

[41] ILAE. Guidelines for epidemiologic studies on epilepsy. *Epilepsia 34*, 4 (1993), 592–596.

[42] Jacquin, A., Causevic, E., and John, E. R. Automatic identification of spike-wave events and non-convulsive seizures with a reduced set of electrodes. In *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE* (2007), IEEE, pp. 1928–1931.

[43] James, G., Witten, D., Hastie, T., and Tibshirani, R. *An Introduction to Statistical Learning: with Applications in R (Springer Texts in Statistics)*. Springer, 2013.

[44] Jasper, H. Report of the committee on methods of clinical examination in electroencephalography. *Electroencephalogr Clin Neurophysiol 10* (1958), 370–375.

[45] Jette, N., Claassen, J., Emerson, R. G., and Hirsch, L. J. Frequency and predictors of nonconvulsive seizures during continuous electroencephalographic monitoring in critically ill children. *Archives of Neurology 63*, 12 (2006), 1750–1755.

[46] Kanik, E., Temel, G., Erdoğan, S., and Kaya, I. Affected states soft independent modeling by class analogy from the relation between independent variables, number of independent variables and sample size. *Balkan medical journal 30*, 1 (2013), 28–32.

[47] Kaplan, P. W. Eeg criteria for nonconvulsive status epilepticus. *Epilepsia 48*, s8 (2007), 39–41.

[48] Khan, Y. U., Farooq, O., Tripathi, M., Sharma, P., and Alam, P. Automatic detection of non-convulsive seizures using ar modeling. In *Power, Control and Embedded Systems (ICPCES), 2012 2nd International Conference on* (2012), IEEE, pp. 1–4.

[49] Kharbouch, A. A. *Automatic Detection of Epileptic Seizure Onset and Termination using Intracranial EEG*. PhD thesis, Massachusetts Institute of Technology, 2012.

[50] Kollialil, E. S., Gopan, G. K., Harsha, A., and Joseph, L. A. Single feature-based non-convulsive epileptic seizure detection using multi-class svm. In *Emerging Trends in Communication, Control, Signal Processing & Computing Applications (C2SPCA), 2013 International Conference on* (2013), IEEE, pp. 1–6.

[51] Koutroumanidis, M., Arzimanoglou, A., Caraballo, R., Goyal, S., Kaminska, A., Laoprasert, P., Oguni, H., Rubboli, G., Tatum, W., Thomas, P., et al. The role of eeg in the diagnosis and classification of the epilepsy syndromes: a tool for clinical practice by the ilae neurophysiology task force (part 1). *Epileptic Disorders 19*, 3 (2017), 233–298.

[52] Law, M. H., and Jain, A. K. Incremental nonlinear dimensionality reduction by manifold learning. *IEEE transactions on pattern analysis and machine intelligence 28*, 3 (2006), 377–391.

[53] Leitinger, M., Beniczky, S., Rohracher, A., Gardella, E., Kalss, G., Qerama, E., Höfler, J., Lindberg-Larsen, A. H., Kuchukhidze, G., Dobesberger, J., et al. Salzburg consensus criteria for non-convulsive status epilepticus–approach to clinical application. *Epilepsy & Behavior 49* (2015), 158–163.

[54] Liang, S.-F., Chang, W.-L., and Chiueh, H. Eeg-based absence seizure detection methods. In *Neural Networks (IJCNN), The 2010 International Joint Conference on* (2010), IEEE, pp. 1–4.

[55] Lowenstein, D. H., and Alldredge, B. K. Status epilepticus. *New England Journal of Medicine 338*, 14 (1998), 970–976.

[56] Megiddo, I., Colson, A., Chisholm, D., Dua, T., Nandi, A., and Laxminarayan, R. Health and economic benefits of public financing of epilepsy treatment in india: An agent-based simulation model. *Epilepsia 57*, 3 (2016), 464–474.

[57] Minasyan, G. R., Chatten, J. B., and Harner, R. N. Detection of epileptiform activity in unresponsive patients using ann. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on* (2009), IEEE, pp. 2117–2124.

[58] Neligan, A., Sander, J., et al. The incidence and prevalence of epilepsy. *London: UCL Institute of Neurology* (2009).

[59] Orabona, F., and Crammer, K. New adaptive algorithms for online classification. In *Advances in neural information processing systems* (2010), pp. 1840–1848.

[60] Orosco, L., Correa, A. G., and Laciar, E. Review: a survey of performance and techniques for automatic epilepsy detection. *Journal of Medical and Biological Engineering 33*, 6 (2013), 526–537.

[61] Pang, S., Ozawa, S., and Kasabov, N. Incremental linear discriminant analysis for classification of data streams. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 35*, 5 (2005), 905–914.

[62] Petersen, E. B., Duun-Henriksen, J., Mazzaretto, A., Kjær, T. W., Thomsen, C. E., and Sorensen, H. B. Generic single-channel detection of absence seizures. In *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE* (2011), IEEE, pp. 4820–4823.

[63] Pronobis, A., Jie, L., and Caputo, B. The more you learn, the less you store: Memory-controlled incremental svm for visual place recognition. *Image and Vision Computing 28*, 7 (2010), 1080–1097.

[64] Qu, H., and Gotman, J. A seizure warning system for long-term epilepsy monitoring. *Neurology 45*, 12 (1995), 2250–2254.

[65] Rodríguez Aldana, Y., Hunyadi, B., Marañón Reyes, E., Rodríguez Rodríguez, V., and Van Huffel, S. Nonconvulsive epileptic seizures detection using multiway data analysis. In *Proc. of the 25th European Signal Processing Conference* (2017), pp. 1–5.

[66] Rodríguez Aldana, Y., Hunyadi, B., Marañón Reyes, E., Rodríguez Rodríguez, V., and Van Huffel, S. Nonconvulsive epileptic seizure detection in scalp eeg using multiway data analysis. *IEEE Journal of Biomedical and Health Informatics* (2018).

[67] Rodríguez Aldana, Y., Sanabria Macías, F., Marañón Reyes, E., Rodríguez Rodríguez, V., Van Huffel, S., and Hunyadi, B. Usando mínimos cuadrados parciales para perfeccionar la detección de crisis epilépticas no convulsivas. *Medisan 0*, 0 (2018), Submited September 27, 2018. Under review.

[68] Roughgarden, T., and Valiant, G. Cs168: The modern algorithmic toolbox lecture# 9: The singular value decomposition (svd) and low-rank matrix approximations. *https://web.stanford.edu/class/cs168/l/l9.pdf* (2017).

[69] Sackellares, J. C., Shiau, D.-S., Halford, J. J., LaRoche, S. M., and Kelly, K. M. Quantitative eeg analysis for automated detection of nonconvulsive seizures in intensive care units. *Epilepsy & Behavior 22* (2011), S69–S73.

[70] Scheffer, I. E., Berkovic, S., Capovilla, G., Connolly, M. B., French, J., Guilhoto, L., Hirsch, E., Jain, S., Mathern, G. W., Moshé, S. L., et al. Ilae classification of the epilepsies: position paper of the ilae commission for classification and terminology. *Epilepsia 58*, 4 (2017), 512–521.

[71] Schomer, D. L., and Da Silva, F. L. *Niedermeyer's electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2012.

[72] Sharma, P., Khan, Y. U., Farooq, O., Tripathi, M., and Adeli, H. A wavelet-statistical features approach for nonconvulsive seizure detection. *Clinical EEG and neuroscience 45*, 4 (2014), 274–284.

[73] Sidiropoulos, N. D., and Bro, R. On the uniqueness of multilinear decomposition of n-way arrays. *Journal of chemometrics 14*, 3 (2000), 229–239.

[74] Smith, G., Wagner, J. L., and Edwards, J. C. Ce: Epilepsy update, part 1 refining our understanding of a complex disease. *AJN The American Journal of Nursing 115*, 5 (2015), 40–47.

[75] Song, Y., Huang, J., Zhou, D., Zha, H., and Giles, C. L. Iknn: Informative k-nearest neighbor pattern classification. In *European Conference on Principles of Data Mining and Knowledge Discovery* (2007), Springer, pp. 248–264.

[76] Sørensen, M., and De Lathauwer, L. Tensor decompositions with block-toeplitz structure and applications in signal processing. In *Signals, Systems and Computers (ASILOMAR), 2011 Conference Record of the Forty Fifth Asilomar Conference on* (2011), IEEE, pp. 454–458.

[77] Sutter, R., Semmlack, S., and Kaplan, P. Nonconvulsive status epilepticus in adults-insights into the invisible. *Nature reviews. Neurology 12*, 5 (2016), 281–293.

[78] Tang, J., Alelyani, S., and Liu, H. Feature selection for classification: A review. *Data classification: Algorithms and applications* (2014), 37–66.

[79] Tatum, W., Rubboli, G., Kaplan, P., Mirsatari, S., Gloss, D., Caboclo, L. K., Drislane, F., Koutroumanidis, M., Schomer, D., Trenite, D. K.-N., et al. Clinical utility of eeg in diagnosing and monitoring epilepsy in adults. *Clinical Neurophysiology* (2018).

[80] Tatum IV, W. O. *Handbook of EEG interpretation*. Demos Medical Publishing, 2014.

[81] Theodoridis, S., Pikrakis, A., Koutroumbas, K., and Cavouras, D. *Introduction to pattern recognition: a matlab approach.* Academic Press, 2010.

[82] Treiman, D. M. Convulsive status epilepticus. *Current treatment options in neurology 1*, 4 (1999), 359–369.

[83] Trinka, E., Cock, H., Hesdorffer, D., Rossetti, A. O., Scheffer, I. E., Shinnar, S., Shorvon, S., and Lowenstein, D. H. A definition and classification of status epilepticus–report of the ilae task force on classification of status epilepticus. *Epilepsia 56*, 10 (2015), 1515–1523.

[84] Trinka, E., and Leitinger, M. Which eeg patterns in coma are nonconvulsive status epilepticus? *Epilepsy & Behavior 49* (2015), 203–222.

[85] Tsai, C.-H., Lin, C.-Y., and Lin, C.-J. Incremental and decremental training for linear classification. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining* (2014), ACM, pp. 343–352.

[86] Tzallas, A. T., Tsipouras, M. G., Tsalikakis, D. G., Karvounis, E. C., Astrakas, L., Konitsiotis, S., and Tzaphlidou, M. Automated epileptic seizure detection methods: A review study. In *Epilepsy-Histological, Electroencephalographic and Psychological Aspects.* InTech, 2012.

[87] Van Esbroeck, A., Smith, L., Syed, Z., Singh, S., and Karam, Z. Multi-task seizure detection: addressing intra-patient variation in seizure morphologies. *Machine Learning 102*, 3 (2016), 309–321.

[88] Varshney, S., Khan, Y. U., Farooq, O., Sharma, P., and Tripathi, M. Latency study of non-convulsive seizures. In *Multimedia, Signal Processing and Communication Technologies (IMPACT), 2013 International Conference on* (2013), IEEE, pp. 108–111.

[89] Vespa, P. M., Nenov, V., and Nuwer, M. R. Continuous eeg monitoring in the intensive care unit: early findings and clinical efficacy. *Journal of Clinical Neurophysiology 6*, 1 (1999), 1–13.

[90] Wall, M. E., Rechtsteiner, A., and Rocha, L. M. Singular value decomposition and principal component analysis. In *A practical approach to microarray data analysis.* Springer, 2003, pp. 91–109.

[91] Webb, A. R. *Statistical pattern recognition.* John Wiley & Sons, 2003.

[92] Wilson, S. B. A neural network method for automatic and incremental learning applied to patient-dependent seizure detection. *Clinical Neurophysiology 116*, 8 (2005), 1785–1795.

[93] Wold, S. Simca: a method for analyzing chemical data in terms of similarity and analogy. *Chemometrics* (1977), 243–282.

[94] Woolson, R. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials* (2007), 1–3.

[95] Xanthopoulos, P., Pardalos, P. M., and Trafalis, T. B. *Robust data mining*. Springer Science & Business Media, 2012.

[96] Xanthopoulos, P., Rebennack, S., Liu, C.-C., Zhang, J., Holmes, G. L., Uthman, B. M., and Pardalos, P. M. A novel wavelet based algorithm for spike and wave detection in absence epilepsy. In *BioInformatics and BioEngineering (BIBE), 2010 IEEE International Conference on* (2010), IEEE, pp. 14–19.

[97] Yamada, M., Saha, A., Ouyang, H., Yin, D., and Chang, Y. n^3 lars: Minimum redundancy maximum relevance feature selection for large and high-dimensional data. *arXiv preprint arXiv:1411.2331* (2014).

[98] Yang, X., Tang, K., and Yao, X. The minimum redundancy–maximum relevance approach to building sparse support vector machines. In *International Conference on Intelligent Data Engineering and Automated Learning* (2009), Springer, pp. 184–190.

[99] Zhang, Y., Xu, G., Wang, J., and Liang, L. An automatic patient-specific seizure onset detection method in intracranial eeg based on incremental nonlinear dimensionality reduction. *Computers in biology and medicine 40*, 11-12 (2010), 889–899.

[100] Zink, R., Hunyadi, B., Van Huffel, S., and De Vos, M. Tensor-based classification of an auditory mobile bci without a subject-specific calibration phase. *Journal of Neural Engineering 13*, 2 (2016), 1–10.

Curriculum Vitae

Yissel Rodríguez Aldana was born on February 17th 1986 in Santiago de Cuba, Cuba. She received the B.Sc. degree in Informatics Sciences Engineering at the University of Informatics Sciences, Havana, Cuba in 2008.

From 2008 to August 2012 she worked as junior researcher in the Neuroscience Center of Cuba (CNEURO), located in Havana, Cuba. In parallel she worked as associated professor at the Universidad de Oriente, Santiago de Cuba, Cuba.

In September 2012, she begins to work as a full-time professor at Universidad de Oriente, achieving the category of assistant professor in 2013. In December 2013 she obtained the degree of master of science in Biomedical Engineering from Universidad de Oriente.

She started the PhD in 2014 in Universidad de Oriente under supervision of Prof. Enrique J. Marañón Reyes and Prof. Sabine Van Huffel. She was admitted to the Arenberg Doctoral School of Science, Engineering & Technology in February 2016 to start a joint PhD. at the KU Leuven and the Universidad de Oriente. Her current research interests include machine learning and tensor methods applied to biomedical signal processing.

List of publications

Papers in (Inter)national Journals

- Rodríguez Aldana, Y., Hunyadi, B., Maranón Reyes, E.J., Rodríguez Rodríguez, V. and Van Huffel, S. "Nonconvulsive Epileptic Seizure Detection in Scalp EEG Using Multiway Data Analysis." IEEE Journal of Biomedical and Health Informatics.(2018), to appear.
- Rodríguez Aldana, Y., Sanabria Macías, F., Maranón Reyes, E.J., Rodríguez Rodríguez, V., Van Huffel, S. and Hunyadi, B. "Usando Mínimos Cuadrados Parciales para perfeccionar la Detección de Crisis Epilépticas no Convulsivas." Medisan.(2018). Submited September 27, 2018. Under review.

Papers in Proceedings of International Conferences

- Rodríguez Aldana, Y., Maranón Reyes, E.J., Puebla Iglesias, T. M. "Ocular artifact correction in the electroencephalogram using empirical modes." VII International Conference on Electrical Engineering FIE 2014, Santiago de Cuba. CD-Rom. Santiago de Cuba.(2014).
- Rodríguez Aldana, Y., Hunyadi, B., Maranón Reyes, E.J., Rodríguez Rodríguez, V. and Van Huffel, S. "Nonconvulsive Epileptic Seizures Detection Using Multiway Data Analysis." In Proc. of the 25th European Signal Processing Conference,(EUSIPCO 2017), Kos, August 28- September 3, 2017. 5 pages.(2017)

Abstracts in Proceedings of (Inter)national Conferences

- Rodríguez Aldana, Y., Hunyadi, B., Maranón Reyes, E.J., Rodríguez Rodríguez, V. and Van Huffel, S. "Semiautomatic Non-Convulsive Epileptic Seizure Detection Using Multi-Way Data Analysis". Abstracts 15th Belgian Day on Biomedical Engineering. Brussels, Belgium. (Nov. 2016).
- Rodríguez Aldana, Y., Hunyadi, B., Maranón Reyes, E.J., Rodríguez Rodríguez, V. and Van Huffel, S. "Nonconvulsive Epileptic Seizure Detection in Scalp EEG with Tensor Analysis." Abstracts PhD Scientific Workshop. Santiago de Cuba, Cuba. (Nov. 2017). Awarded with the best scientific pesentation.
- Rodríguez Aldana, Y., Hunyadi, B., Maranón Reyes, E.J., Rodríguez Rodríguez, V. and Van Huffel, S. "Tensor-based model for nonconvulsive epileptic seizures detection." Abstracts EURASIP summer school on tensor-based signal processing. (Aug. 2018). Awarded with the best poster pesentation.



FACULTY OF ENGINEERING SCIENCE
DEPARTMENT OF ELECTRICAL ENGINEERING
BIOMED
Celestijnenlaan 200A box 2402
B-3001 Leuven
first.last@cs.kuleuven.be
http://www.XXXXX.cs.kuleuven.be