République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Minisère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي والبحث العلمي

Ecole Nationale Polytechnique

Département d'Electronique

End of Studies Project Thesis
submitted for the fulfillment of:
State Engineer Degree in Electronics

Contribution to the Characterization of EEG data for Epileptic Seizures Detection

---

Author:
SADOUN Maria Sara Nour

Under the supervision of Dr Taous-Meriem Laleg, Professor
and Dr Mourad Adnane, Professor

Presented and defended on June 30^{th}, 2022 before the members of jury:

President	Cherif	LARBES	Prof.	Ecole Nationale Polytechnique, Algiers
Examiner	Mohamed O.	TAGHI	MAA	Ecole Nationale Polytechnique, Algiers
Supervisor	Taous Meriem	LALEG	Prof.	KAUST, Thuwal/INRIA-Saclay, Paris
Co-Supervisor	Mourad	ADNANE	Prof.	Ecole Nationale Polytechnique, Algiers

ENP 2022

République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Minisère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي والبحث العلمي

Ecole Nationale Polytechnique

Département d'Electronique

End of Studies Project Thesis
submitted for the fulfillment of:
State Engineer Degree in Electronics

Contribution to the Characterization of EEG data for Epileptic Seizures Detection

---

Author:
SADOUN Maria Sara Nour

Under the supervision of Dr Taous-Meriem Laleg, Professor
and Dr Mourad Adnane, Professor

Presented and defended on June 30^{th}, 2022 before the members of jury:

President	Cherif	LARBES	Prof.	Ecole Nationale Polytechnique, Algiers
Examiner	Mohamed O.	TAGHI	MAA	Ecole Nationale Polytechnique, Algiers
Supervisor	Taous Meriem	LALEG	Prof.	KAUST, Thuwal/INRIA-Saclay, Paris
Co-Supervisor	Mourad	ADNANE	Prof.	Ecole Nationale Polytechnique, Algiers

ENP 2022

République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Minisère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي والبحث العلمي

Ecole Nationale Polytechnique

المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Département d'Electronique

Mémoire de Projet de Fin d'Etudes
pour l'Obtention du Diplôme:
d'Ingénieur d'Etat en Electronique

Contribution à la Caractérisation de données EEG pour la Détection de Crises Epileptiques

---

Auteur:
SADOUN Maria Sara Nour

Sous la supervision de Dr Taous-Meriem Laleg, Professeur
et Dr Mourad Adnane, Professeur

Présenté et soutenu publiquement le 30 Juin 2022 devant les membres du jury:

Président	Cherif	LARBES	Prof.	Ecole Nationale Polytechnique, Algiers
Examinateur	Mohamed O.	TAGHI	MAA	Ecole Nationale Polytechnique, Algiers
Promoteur	Taous Meriem	LALEG	Prof.	KAUST, Thuwal/INRIA-Saclay, Paris
Co-Promoteur	Mourad	ADNANE	Prof.	Ecole Nationale Polytechnique, Algiers

ENP 2022

الملخص يعد اكتشاف نوبة الصرع مشكلة صعبة تتمثل في تحديد النوبة بين نشاط الدماغ الطبيعي باستخدام إشارات مخطط كهربية الدم (ييج) ، إما بواسطة طبيب أعصاب متمرس أو باستخدام أطر هندسية. في هذا العمل ، نهدف إلى المساهمة في مجال الكشف التلقائي عن نوبات الصرع لمساعدة الخبراء في المرافق الطبية وتحسين حياة المرضى بأمان واستقلالية ؛ وعلى المدى الطويل ، المساهمة في للنهوض بإعادة التأهيل وتقنيات المساعدة. من وجهة نظر هندسية ، سنسعى لفهم تأثيرات ومساهمة كل الميزات. سنقوم بتضمين نوعين من الميزات: أحدهما يتدفق من تحليل الإشارة شبه الكلاسيكية والثاني من السلوك الديناميكي غير الخطي للنشاط العصبي. سوف تستغل تنوع التردد البيانات ونساهم في تحسين المعلمات للسمات الديناميكية. في وقت لاحق ، بمعالجة مشكلة البيانات غير المتوازنة من خلال تقديم شبكات الخصومة التوليدية لزيادة البيانات. تظهر النتائج التجريبية موثوقية سير العمل وتحسين الأداء مقارنة بالدقة والحساسية والنوعية تقترب هذه المقاييس الثلاثة من الدرجات شبه المثالية ، 0.99 . يرجع هذا الأداء إلى جزأين رئيسيين: مقدمة ، لأول مرة لـ صثصا لتوصيف نوبة الصرع وتحسين للمعلمات غير الخطية.
مصطلحات المؤشر : ييج ، اكتشاف نوبات الصرع ، هندسة الميزات ، صثصا ، الديناميكي غير الخطي ، التحسين ، زيادة البيانات ، جا ، التصنيف.

Résumé- La détection de crises épileptiques est une problèmatique qui consiste à identifier une crise parmi l'activité cérébrale normale à l'aide de signaux d'électroencéphalogramme (EEG), soit par un neurologue expérimenté, ou dans des cadres automatiques conçus. Par ce travail, nous visons à contribuer à ce dernier pour aider les experts des établissements médicaux et améliorer la sécurité et l'autonomie des patients. Nous nous efforcerons de comprendre les effets et contribution de tout descripteur. Nous en incluons deux types: SCSA et les descripteurs dynamiques non linéaires. Nous exploiterons la diversité fréquencielle de l'EEG et contribuerons à l'optimisation des hyper-paramètres pour les caractéristiques dynamiques. Plus tard, nous traitons les données déséquilibrées en introduisant des Generative Adversarial Networks pour augmentater les données. Les résultats expérimentaux démontrent la fiabilité du workflow et l'amélioration des performances en termes de précision, sensibilité et spécificité. Ces trois métriques atteignent des scores de $0,99$. Cela est dû à deux parties principales: introduire, pour la première fois, la SCSA pour caractériser les crises épileptiques et l'optimisation des hyper-paramètres des descripteurs non-linéaires.

Mots-clés: EEG, détection des crises épileptiques, ingénierie des descripteurs, SCSA, dynamique non-linéaire, optimisation, GAN, classification.

Abstract- Epileptic seizure Detection is a challenging problem which consists in identifying a seizure among normal brain activity using electroencephalogram (EEG) signals, either by an experienced neurologist or automatically engineered frameworks. In this work, we aim to contribute to the latter to help experts in medical facilities and improve the safety and autonomy of patients. We will strive to understand the effects and contribution of each and all features. We include two types of features: SCSA and nonlinear dynamical features. We will exploit the frequency diversity of EEG and contribute to the optimization of time-embedding hyper-parameters for the dynamical features. Later on, we tackle imbalanced data by introducing 2D-Generative Adversarial Networks for Data Augmentation. Experimental results demonstrate the reliability of the workflow and performance enhancement compared to state-of-the-art accuracy, sensitivity and specificity. The three metrics approach consist scores of 0.99. This is due to two main parts: the introduction, for the first time of the SCSA to characterize epileptic seizures and the careful optimization of the time-embedding hyper-parameters for the nonlinear features.

Index-terms: EEG, Epileptic Seizure Detection, Feature Engineering, SCSA, nonlinear dynamics, optimization, GAN, classification.

Acknowledgment

*I would like first and foremost to thank and acknowledge **King Abdullah University of Sciences and Technology, KAUST** for proposing this thesis project in the first place. Moreover, having the opportunity to work within an academic environment as welcoming and enriching has undoubtedly been the key factor for the success of the work presented and conducted within this project.*

*I would like to express my most sincere and genuine gratitude to Professor Tous Meriem Laleg and Professor Tareq El-Naffouri from the **CEMSE; the Computer, Electrical and Mathematical Sciences and Engineering Division** for welcoming me and granting me the opportunity to grow academically as a research intern within the **EMANG, the Estimation, Modelling and ANalysis Group** throughout the past four months.*

Finally, I would like to thank KAUST for their funding, and their providing of IT means and assistance throught the last months.

This project wouldn't be what it is today without them welcoming me as an intern.

Acknowledgment

First and foremost, I would like to take the opportunity to express my deepest gratitude and thanks towards my project supervisor Dr Taous Meriem Laleg, Professor with King Abdullah University and permanent researcher with INRIA-Saclay, for her constant guidance and support, for her unwavering kindness, gentleness and presence at all times, for the knowledge she always shares so abundantly and her contagious passion for engineering and mathematics which inspires quite easily admiration and inspiration.

I would like to thank her for the opportunity she offered me to work within the Estimation, Modeling and Analysis Group under her supervision and grow as a debutant researcher but moreover as a human, as I have during the course of this project. Working under her supervision is undoubtedly a highlight in my academic years.

I would like to express once again my sincere gratitude to my supervisor Mr Mourad Adnane, Professor with Ecole Nationale Polytechnique for his supervision and flexibility. As a Professor, he has been one of the firsts to introduce me to a field I've adopted to be my vocation: Biomedical Engineering, and for three consecutive years, shared assistance and knowledge with constant kindness.

I would like to thank the President of the jury Dr Cherif Larbes, Professor with Ecole Nationale Polytechnique and the Examiner Mr Mohamed Oussaid Taghi, Teacher-Researcher with Ecole Nationale Polytechnique for agreeing to be members of the reading committee and for their constructive analysis of the present work. A special thanks is dedicated to them for their time and teaching and support for the last three years.

I would like to address my gratitude to King Abdullah University of Science for its funding of the present project without which this contribution wouldn't be what it is today. Along comes my most sincere thanks to Professor Tareq El-Naffouri for making it possible in the most fluid way.

, this is an acknowledgment to all Teachers and Professors I've had the honor of crossing path with since elementary school and my early years. To each and all of you, I owe in great part where I am today, accompanied with my most sincere appreciation.

Dedication

This work could not be first dedicated to anyone but my Mom. The source of all things. She, who showed me through her own vocation, how to turn empathy and channel the incapacity to accept the condition of the most disadvantaged of us in today's society, into a vocation of my own. Thank you for an unwavering gentleness that could not be anyone's but yours. Thank you for supporting and championing me at all time. If I am anything today, I owe it to you completely.

to my Father for bearing with me and the long talks that came along traffic, for in his own ways, taught me everything he knew.

This is a dedication to my sister Yousra and my brothers Mehdi, Youcef and Yacine. You are each and all rays of the sunshine that rises on me each morning and makes everyday worth the struggle.

I might add a special thanks to my aunt, my second mother, and her husband for their constant support and championing and making life better, a little easier everyday and so, in all types of ways.

This thanks is to all my friends and classmates for the past five years within Ecole Nationale Polytechnique, and most especially my Electronics class and the vic. Being a part of this community has been the honor of my young years. Thank you for the laughter, the hard work, sharing the good, the bad and less tasty meals. May we all together look back on these years of our lives with nothing but light nostalgia and deep gratitude.

Thank you to the EMAN-Group and Family, for welcoming me with open arms and minds. Joining you all on this oh beautiful journey and having the honor of meeting and befriending each of you from the four corners of the globe has been the highlight of my year. Here's to Manuel, Maria, Peihao, Mohamed, Ibrahima, Yasmine, Fahad, Evangelos, Ania and last but not least: Meriem.

I would like to take the time and thank my friends of heart Lilia and Yousra, without whom I wouldn't be here today. Thank you for your love, support, patience and for being, the three of us together, "us". Last year was yours, now it's my turn.

Zainab, this is not simply a thesis dedication payback but you already know that. May we reunite in Ist. and escape jet-lag and the limits of screens. Again, it is the East and Juliet is the Sun.

Finally, this work is dedicated to all epileptic patients, who have been put aside, marginalized, who missed job opportunities and the privilege of safe autonomy. May the scientific progress we are dedicating our days and lives to, make your lives a tat easier, a tat more just and equal to daily privilege we may forget about.

Contents

List of Figures

List of Tables

General Introduction

Motivation and Challenges

Epilepsy is a common brain disorder that, according to the World Health Organization (WHO) and the International League Against Epilepsy (ILAE) 2015, affects over 65 million people around the world, making it one of the most common neurological diseases globally [1]. One more relevant statistic to introduce is the fact that 70% of them live in low-income, third worlds countries [1]. In Algeria, it was estimated to have 400.000 people with epilepsy between children and adults [18] but we consider that given the lack of medical institutions in rural zones for diagnosis, to be largely superior to this rate.

Epilepsy is characterized by recurrent, unprovoked seizures. Seizures are the result of a transient and unexpected electrical disturbance of the brain and excessive neuronal discharge of the electrical activity of the brain [40]. They can vary from the shortest failure of attention to intense and extended convulsions and they can also range in frequency from less than 1 per year to several per day. Epilepsy is a source of constant anxiety for patients due to the unexpected onset of seizures involving the triggering of manifestations of which they lose consciousness. Therefore, it is a considerable undermining of the autonomy and safety of the patients. Indeed, epileptic patients, when unassisted, tend to have physical issues such as break and injuries as a result of seizures, which decrease the patient's quality of life. In addition, epileptic patients are socially marginalized and have reduced access to educational opportunities and professional positions, and official withholding of the possibility of obtaining a driving license; barriers to enter particular occupations, and excessive need to health and life insurance and the increase of their costs. Currently there is no cure for epilepsy but few drugs and medications to inhibit and control specific types of seizures, but which have at the same time significant collateral side effects.

A second option for drug-resistant patients, is the candidacy for surgical intervention given they are inherently forced to deal with it on a daily basis [54]. Among these patients, 7% to 8% may profit from epilepsy surgery but at the end, despite all these medical alternatives, about 25% of people with epilepsy will continue to experience seizures even with the best available treatment.

An additional aspect of the dangers epilepsy may present for the ones affected by this disorder is the occurrence of epileptic seizure when put under anesthesia for the sake of surgery or during artificial, or natural somatic states.

And last but not least is the growing interest for real-time, autonomous intervention in the context of assistance and rehabilitation technologies for neurological disorders. The latter research axis we explain in the previous paragraph is directed towards the development of neuro-stimulation and neuro-feedback devices that relies on the detection of the occurrence of an epileptic seizure in order to intervene with electrical stimulation to assist the brain in his effort of instantly recovering its

ordered, normal state.

As for the case of the majority of the epileptic population which live in low-incomes countries; the lack of facilities to diagnose rightfully Epilepsy is fueled mainly by the need of experimented neurologists and very long period of time to read EEG recording of patients. This represents a budget that goes beyond the financial scope rural communities can afford or be granted by authorities.

The globality of these statistics, the thorough understanding of the patients and neurologists' struggle is coupled to the reality that the majority of the world's cases struggle in third world countries are either un- or misdiagnosed because of the lack of resources needed. Adding to it last but not least, the promising research axis to tackle epilepsy in real-time and autonomously; push us to direct our attention towards the possibility of detecting epileptic seizures autonomously by monitoring the electrical neural activity of the patients.

During the last two decades, many frameworks have been proposed to tackle the problem of autonomous epileptic seizure detection in the prospect of contributing to one of the alternative previously explained, and as we will develop and present further within this work. However, given EEG signals are subject-specific and are different across types of EEG devices. Consequently, most seizure detection frameworks are overly patient-specific and fail when transitioning to clinical use in medical facilities. Additionally, deep learning frameworks fail mostly despite their performances on paper that approach the 100% because the rise in abstraction proposed by these architectures fail to consider the difference in the information distribution they will be fed with, compared to the training and test sets their development was based upon. In parallel, Machine-Learning framework encompass a multitude of milestones and at the level of each can limitations be identified. An extensive presentation of the limitations of the multiple machine learning approaches of the state-of-the-art will be presented later on in this dissertation. Qualitatively, the principal disadvantages of these latter methods are:

- The excessive use of manually extracted features that rely directly on human manipulation.
- The challenges of identifying the most discriminative features that best embody the differences between each class.

Contribution

A new tendency within the research field of epileptic seizure detection is to strive on designing frameworks that encompass autonomous Feature Engineering and a classification phase during which hyper-parameters are mildly tuned, so that they are at the service of the workflow performance and not the other way around.

The aspiration and aim of this project is to develop and design an automatic and autonomous framework that embodies this newer tendency that overcomes the principal limitations of past research work enunciated above.

The main contributions of our work can be summarized as follows:

- Clearly enunciate the relevance of nonlinear dynamical features with respect to the medical characterization of epileptic seizures and the electrical behavior of the brain of epileptic patients.

- Design an exhaustive Feature Engineering phase with its multiples pans: Extraction, Correlation study, Ranking and Selection of these features. It encompasses:
  - Introduce a novel kind of features to characterize EEG data: the Semi-Classical Signal Analysis, which proved performing on pulse-shaped,physiological signals.
  - Optimize the time-embedding hyper-parameters of the nonlinear dynamical features that capture the chaotic character of electrical neural activity.
- Propose a compromise between efficiency and computational simplicity through the design of a two-dimensional Generative Adversarial Networks for data augmentation in order to tackle Imbalanced datasets inherent to epileptic data given seizure last seconds or minutes and seizure-free states last the rest of the time.
- Validation and testing of the last proposition.
- Evaluate the performance of the whole workflow by performing binary classification with minimum hyper-parameter tuning for performance evaluation and comparison of results.

Organization of the Dissertation

This dissertation is organized as follows:

- Part **I** provides an overview of the background knowledge in the context of defining and epileptic seizures from a medical and engineering point-of-view; and classifying them, followed by an exhaustive presentation of Electroencephalography (EEG) and electrical neural activity. We also conduct a review of the state-of-the-art directed in the same research axis we consider for eventual performance comparison.
- Part **II** presents the core of the contribution: **Feature Engineering** and encompasses:
  - The extraction of the features with its two types: Semi-Classical Signal Analysis features and nonlinear dynamical ones.
  - The optimization of the time-embedding parameters of the nonlinear features.
  - The correlation study between newly introduced and well-established features.
  - The features' ranking and selection phase.
- Part **III** lays out our solution to Imbalanced datasets and explanation of our approach of developing a Generative Adversarial Network that we then validate.
- Part **IV** summarizes the performance of the Machine-Learning classification algorithms as a final step.
- Part **V** reflects on the contribution of the present dissertation and projects us towards eventual, future directions to carry on this research work.

Part I

Background knowledge

Chapter 1

Introduction

This part, as a whole, is dedicated to presenting in-depth explanation of fundamental concepts directly related to our case study. We present characteristics of *Epilepsy* as a disorder, its physical symptoms, neurological manifestation. Further, we present epileptic seizures types given the importance of taking in consideration this categorization in our case study due to the subject-specificity of biophysical signals. We, then carry on with a presentation of Electroencephalography (EEG) signals as a main acquisition system of electrical neural activity. Clinical specificities in the EEG collected on epileptic patients are defined which allows us to further define each notion which represents a case study by itself; from *epileptic spikes, (epileptiform discharges)* to *epileptic seizure detection*, and its *onset* and finally its *prediction*; and distinguish each from each and how they are also related and linked. Finally, this introduces the practical and engineering formulation of the subject of our project: *epileptic seizure detection* as a binary classification problem. The process of classification is one of predicting categorical variables where the output is restricted to two classes, in our case: seizure data and seizure-free data. Most efficient and well-established classification solutions utilize Deep Learning and Machine Learning algorithms. We'll present an brief explanation of the principle of classification in Artificial Intelligence, illustrated rhrough a state-of-the-art and literature review of the latest work done for *epileptic seizure detection*. A crucial point which justifies our motivation to conduct this project is the poor performance of most algorithms when transitioning to clinical use and explain multiple limitations in these frameworks in need to be addressed.

Chapter 2

Epilepsy disorder

In this chapter, we ought to introduce the medical definition of epilepsy and epileptic seizure followed by a formulation of the epileptic seizure problem from an engineering point of view along with an overview of the classification of epileptic seizures according to the International League Against Epilepsy in 2017.

2.1 Conceptual definitions of Epileptic Seizure and Epilepsy disorder

Epilepsy is the name of a brain disorder characterized predominantly by recurrent and unpredictable interruptions of normal brain function, called epileptic seizures. It is not a singular disease entity but a variety of disorders reflecting underlying varying brain dysfunction typically which is associated with clinico-electrical correlate on the electroencephalogram (EEG). Therefore, the epilepsies -a term preferred by many specialists- encompass a heterogeneous group of neurological disorders and syndromes.

It has proven necessary, multiple times through years; to state, adapt and correct definitions concerning the electrical activity of the brain during epileptic seizure episodes and how the disorder of epilepsy is defined based on the latter. These definitions affect both clinical diagnosis and treatment, understanding of communications and are crucial when formulating and structuring this medical issue as an engineering one.

We base our definitions on the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE) 2005 official report [40].

- An **epileptic seizure** is a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in the brain.

- **Epilepsy** is a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures and by the neurobiologic, cognitive, psychological, and social consequences of this condition. The definition of epilepsy requires the occurrence of at least one epileptic seizure.

To be sharper in our definition of epileptic seizures -being the core subject of our study case-, there are three main elements in clinically defining epileptic seizures that alter also the engineering perception of the subject of the conducted study:

1. **Mode of onset and termination:** An epileptic seizure is "transient" demarcated in time, with a clear start and end which can both be determined on EEG grounds. Its termination is less evident to determine because the remaining symptoms post-seizure can blur the latter.

2. **Clinical manifestations:** Signs and symptoms with a wide range of possible manifestations, feature prominently in the definition of epileptic seizure. On one side, seizure presentation depends on location of onset in the brain, patterns of propagation, maturity of the brain and medications. On another hand, seizure affects sensory and motor function; consciousness; memory; cognition. There are additional "psychic" manifestations, more complex and internal which consist of complex perceptual distortions. All previous signs represent facets to detect epileptic seizures but moreover, present eventual future extension to this study case.

3. **Abnormal enhanced synchrony:** Including the requirement for an uncontrolled electrical discharge as a feature of the definition of an epileptic seizure is necessary to distinguish it from many other clinical events that are not epileptic seizures but could be interpreted as such in EEG. EEG discharges during epileptic seizures are orderly and relatively stereotyped and neurons firing may involve inhibition as well as excitation. A feature more common to epileptic seizures is abnormal enhanced synchrony of neurons.

These facets of defining epileptic seizure can be exploited to discriminate healthy state from epileptic seizure state. Nevertheless, discussion of systems giving rise to epileptic seizures properly doesn't fall within the scope of the 2005 ILAE [40] definition since it doesn't directly ascribe neural behaviors noted, to the occurrence of epileptic seizures.

The consideration previously mentioned are missing from our actual frame although necessary in our context. Therefore, we introduce a few new concepts to technically define epilepsy beyond linear, standards systems and features; to fit better the aim of this study case of exploiting EEG signals for epileptic seizure detection.

The standard methods for time series analysis (e.g., power analysis, linear orthogonal transforms, and parametric linear modeling) not only fail to detect the critical features of a time series generated by an autonomous (no external input) nonlinear system, but may falsely suggest that most of the series is random noise.

2.2 Problem formulation

A great part of recent research interest in neuroengineering has turned to the mathematical concept of chaos as an explanation for a variety of complex processes we can illustrate in epilepsy disorder. Multiple properties of chaotic, nonlinear systems proved to be a turning-point in the modeling of physiological systems [47]:

1. Chaotic systems exhibit abrupt intermittent transitions between highly ordered and disordered states.

   > - In the same context, epilepsy is a group of disorders characterized by recurrent electrical discharges of the cerebral cortex that result in intermittent disturbances of brain function.

   This introduces an additional property of chaotic systems which is the ability to show self-organization: to evolve toward ordered temporal and spatial patterns. The transition from chaotic to ordered behavior, or the reverse, can occur as an abrupt phase transition with a minute change in the control parameters.

2. An essential, yet unaddressed in the previous definition, feature of epilepsy is the fact that seizures come and go over time - that seizures occur intermittently. This intermittency is difficult to explain according to the concepts of linear dynamics because in linear systems sudden transitions in state occur only in response to external input or trigger and not inherently to the system, in contrast to nonlinear ones.

3. Chaotic systems exhibit strong dependence on initial conditions: small differences in the initial value x_1, will result in big differences in the subsequent values x_n over time. This strong dependence on initial conditions means that predicting the long-term behavior of chaotic systems is difficult.

Therefore, a promising avenue for research into *epileptogenesis* is within the domain of nonlinear systems.

Historically, mathematical models of brain function have been applied to individual neurons or to relatively small networks of neurons. However, in the case of human epileptogenesis, our understanding of the neuronal mechanisms and circuitry involved is insufficient to formulate mathematical models on which to base a conclusive dynamical analysis.

We are interested, now, in an alternative to mathematical modeling, which aims to obtain empirical measures of the behavior of the system as a whole over time. This macroscopic approach is particularly useful for biological systems, such as the brain, where exact knowledge of the system is lacking. Analysis along these lines can provide insight into the global dynamical properties of the system. In this optic, the recent application of this approach to the analysis of EEG recordings in epileptic patients has provided exciting discoveries. These observations and previous results justify and support our approach in this work.
To this end, we formulate and conceptualize our electroencephalograms (EEG) as a series of numerical values (voltages) over time: a **time series**".

Definition 1 *A univariate time series $X = [x_1, x_2, \ldots, x_T]$ is an ordered set of real values. The length of X is equal to the number of real values T.*

A seizure detection methodology can be defined as follows: we select a database of EEG signals categorized according to the state of the brain: **seizure state** and **seizure-free state**. We opt for a binary classification methodology using a fitting binary classifier to exploit the differences and discrimination between the two states. Therefore the problem can be treated as a standard classification problem where windows of the EEG time-series are processed into features and labeled to reflect the neural state of the patient.

2.3 Types of epileptic seizures

Figure 2.1 presents the most important seizure classes given in a tree structure and presented in the most recent communication [41], the **ILAE 2017 Basic Classification of Seizures Types**. Seizures are first categorized by type of onset.

1. **Focal-onset** seizures are defined as "originating within networks limited to one hemisphere and may be discretely localized or more widely distributed."

2. **Generalized-onset** seizures are defined as "originating at some point within, and rapidly engaging, bilaterally distributed networks."

3. A seizure of **unknown onset** may still evidence certain defining motor (e.g., tonic–clonic) or nonmotor (e.g., behavior arrest) characteristics. These seizures can further be reclassified into focal or generalized-onset categories with more information.

The next level of focal seizure classification is by level of awareness which is defined as knowledge of self and environment and is a marker which determine whether level of consciousness is impaired. It specifically refers to awareness during a seizure, and not to awareness of the occurrence of the seizure.

1. a **focal aware seizure** implies the ability of the person having the seizure to later verify retained awareness.

2. If awareness of the event is impaired for any portion of the seizure, then the seizure is classified as a **focal seizure with impaired awareness**.

The seizure type **"secondarily generalized tonic–clonic"** is in a category on its own because of its common occurrence and importance and is reflective of a propagation pattern of seizure activity rather than a unique seizure type.

Level of awareness is not used as a classifier for generalized seizures, since the large majority of generalized seizures are associated with impaired awareness.

Seizures are further optionally, in this extended classification, characterized by:

1. **Motor onset signs and symptoms**: atonic, automatisms, clonic, epileptic spasms, or hyperkinetic, myoclonic, or tonic activity.

2. **Nonmotor-onset seizures** can manifest as autonomic, behavior arrest, cognitive, emotional, or sensory dysfunction.

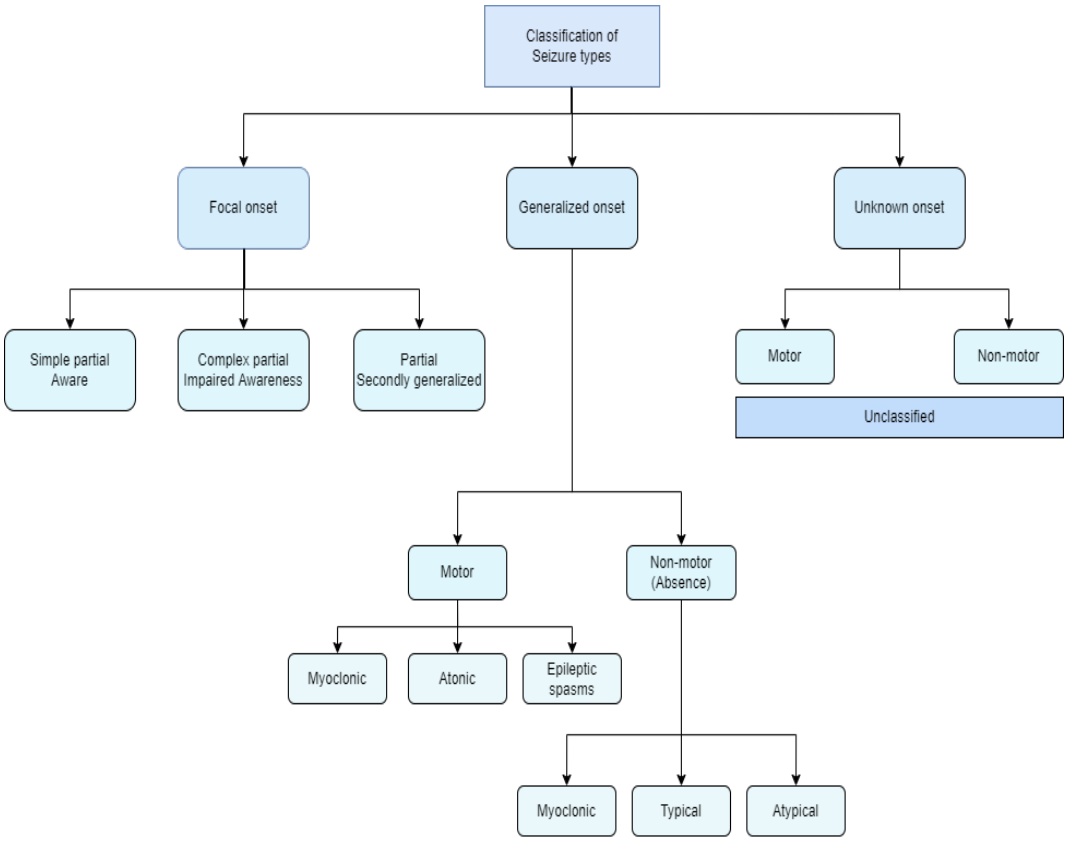


Figure 2.1: Classification of epileptic seizure according to ILAE 2017

Conclusion

In this chapter, we were able to define Epilepsy according to its latest medical definitions and furthermore explicit characteristics of epileptic seizures and facets which we are capable of exploiting in order to tackle **Epileptic Seizure Detection** from a neuroengineering and biomedical point of view and position. Additionally, the classification we present previously opens newer research axis to us: the possibility of dealing with different types of epilepsy according to the specificity of each, rather than solely direct our research towards more specific pathologies.

Chapter 3

Electroencephalography (EEG)

In this chapter, we propose an extensive presentation of the device of Electroencephalography (EEG) with the possible electrode placements according to the baseline considered and the area of the brain ought to monitor. Another crucial section is the presentation of the different brain-wave frequency sub-bands'. We also explain stages of EEG related to epilepsy disorder and how other devices can be coupled with EEG to tackle epilepsy. There is also a presentation, to a certain extent, of the state-of-the-art of epileptic seizure detection.

3.1 What is EEG

Electroencephalography (EEG) is a measurement of potentials that reflect the electrical activity over time of the human brain [81]. The EEG is widely used to study brain functions and to diagnose neurological disorders. The study of the brain's electrical activity, through EEG records, is one of the most important tools for the diagnoses of neurological diseases, such as epilepsy, brain tumors, head injury and dementia. It is also helpful for the treatment of abnormalities, behavioral disturbances (e.g., Autism) and attention disorders. The first EEG recording machine was introduced to the world by Hans Berger in 1929 who was a neuropsychiatrist from the University of Jena in Germany, and who described the graphical representations of the electric currents generated in the brain by the German term *"elektrenkephalogramm"*. The emergence of this revolutionary device helped create a new branch of medical science called ***neurophysiology***.

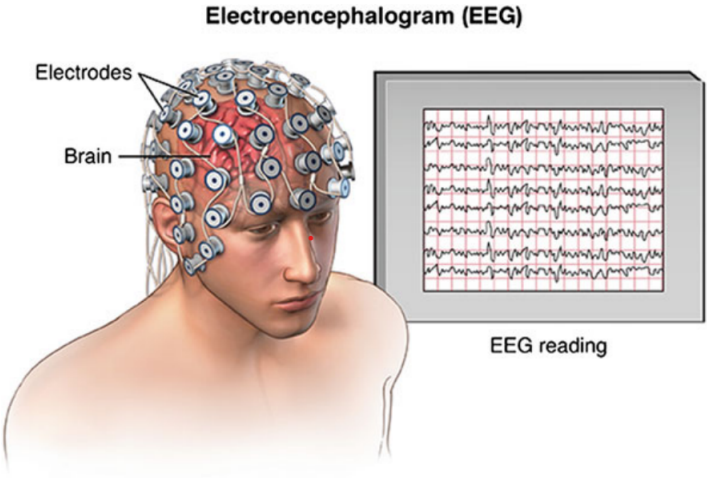


Figure 3.1: An illustration of EEG recording [81]

During the EEG test a number of small disks called electrodes are placed in different locations on the surface of the scalp with temporary glues.

Each pair of electrodes is connected to an amplifier and an EEG recording machine to display the wavy lines the electrical signals are converted to.

Fig. 3.1 presents an example of electrodes placement on the scalp during the recording of EEG signals while they are displayed on a computer screen. The electrodes detect tiny electrical charges that result from the activity of the brain cells, which are amplified to appear on a graph.

There are two types of EEGs, depending on where the signal is taken in the head:

- **scalp EEG:** small electrodes are placed on the scalp with good mechanical and electrical contact.
- **intracranial EEG:** Special electrodes are implanted in the brain during the surgery result.

The amplitude of an EEG signal typically ranges from about 1 to 100 μV in a normal adult, and it is approximately $10 - 20mV$ when measured with subdural -needle- electrodes. Since the architecture of the brain is nonuniform and the cortex is functionally organized, the EEG can vary depending on the location of the recording electrodes.

3.2 Electrodes placement

The question of the electrodes placement is important, because different lobes of cerebral cortex are responsible for processing different types of activities. The standard method for the scalp electrode localization is the international $10 - 20$ electrode system. The "10" and "20" stand for the actual distances between neighboring electrodes which are are either 10 or 20% of the total front-back or right-left distance of the skull. The positions are determined by the following two points:

- **nasion**, which is the point between the forehead and the nose, level with the eyes.
- **inon** which is the bony prominence at the base skull on the midline at the back of the head.

Fig. 3.2 presents the electrode position on the brain according to the **international 10–20 system**.

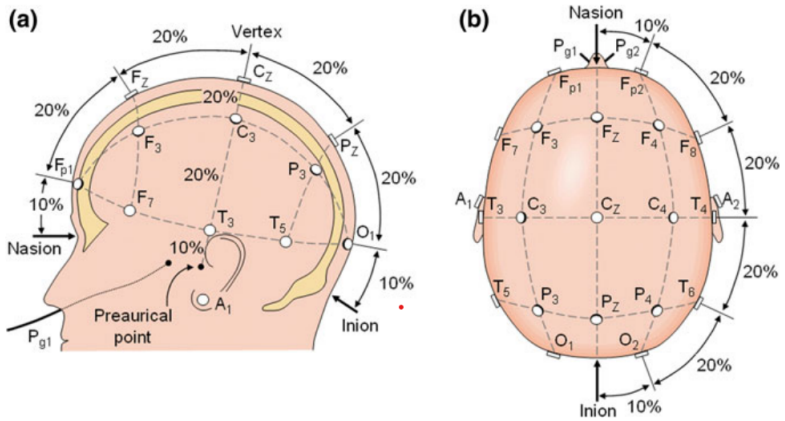


Figure 3.2: The international 10–20 electrode placement system [81]

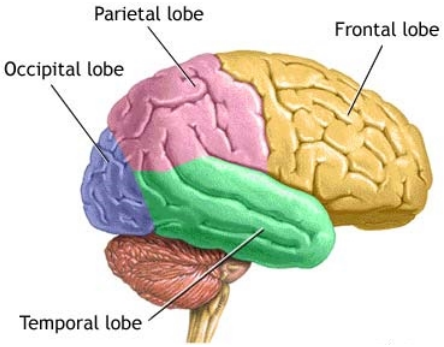


Figure 3.3: Division of each lobe in the brain hemisphere

Brain region/Cerebral lobe	Channel letter
Frontal	F
Parietal	P
Occipital	O
Temporal	T

Table 3.1: Division of electrodes according to cerebral lobe

Each location uses a letter to identify the brain lobes and regions as presented in Figs. 3.3 3.1
A number accompanies the letter in order to identify the hemisphere location. Odd numbers are used to refer to the electrodes positioned on the left hemisphere while the even numbers are for those on the right hemisphere. Instead of a number associated with the lobe letter, a **"z"** refers to an electrode placed on the *midline*.

Since an EEG voltage signal represents a difference between the voltages at two electrodes, the display of the EEG for the reading EEG machine may be set up in several ways. The placement of the electrodes is referred to as a ***montage***.

3.3 Common EEG montages

There are multiple available EEG montages according to the way the potential difference - displayed on the device- between two electrodes is calculated. Among these montages, we count:

3.3.1 Bipolar montage

One pair of electrodes makes up a channel and each channel (waveform) represents the difference between two adjacent electrodes and the entire montage consists of a series of these channels. Fig. 3.4 presents a diagram of a bipolar montage, where the channel $''Fp1^{\smile}F3''$ represents the difference in the voltage between the $Fp1$ electrode and the $F3$ electrode. The next channel in the montage, $''F3^{\smile}C3''$, is the voltage difference between $F3$ and $C3$, and so on, through all electrodes.

3.3.2 Referential montage

Each channel represents the difference between a certain electrode and a designated reference electrode. There is no standard position for this reference (in Fig. 3.5, we use the electrode $A2$ as the referential electrode). Midline positions are often used because they do not amplify the signal

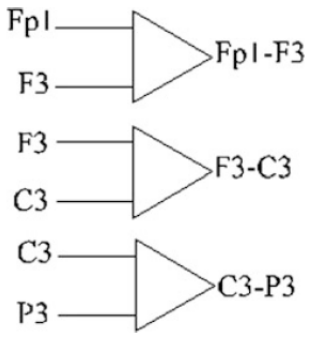


Figure 3.4: An illustration of the bipolar montage [81]

in one hemisphere versus the other. Another popular reference is *"linked ears"*, which is a physical or mathematical average of electrodes attached to both earlobes and mastoids.

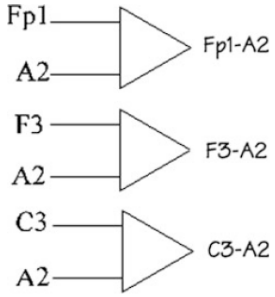


Figure 3.5: An illustration of the referential montage [81]

3.3.3 Average reference montage

The outputs of all of the amplifiers are summed and averaged, and this averaged signal is used as the common reference for each channel. An illustration of the average reference montage is given by Fig. 3.7.

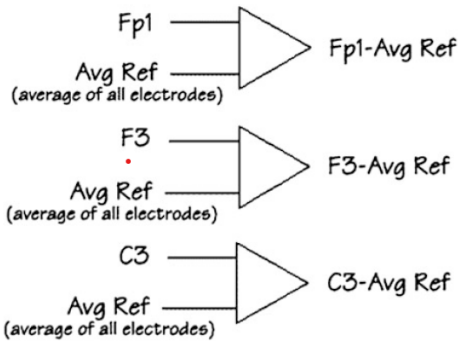


Figure 3.6: An illustration of the average referential montage [81]

3.4 Fundamental brain-waves frequency sub-bands

The neural signals discharge rhythmically and spontaneously which indicates their spontaneous characteristics and this certainly became a novel direction of EEG research. In addition, the rhythmicity of spontaneous EEG is the most important feature of EEG. The frequency diversity -along with the spatial one- is a core . At present, EEG can be divided into the following categories according to the frequency of rhythm [8]:

3.4.1 Delta Waves

Contains both the **delta** 1 rhythm **(1-2Hz)** and the **delta** 2 rhythm **(2-4Hz)** and are considered as "non-major" rhythms which mostly occur in deep sleep or anesthesia and organic lesions in the brain. The amplitude range is $20 - 200\ \mu V$.

3.4.2 Theta Waves

The frequency in the range of $(4 - 8\ Hz)$; it appears on both sides of the parietal and temporal lobes in the brains of younger people; and it appears when drowsiness or the central nervous system is in a state of inhibition.

3.4.3 Alpha Waves

The frequency range is $(8 - 13\ Hz)$. The alpha rhythm is the most important component of the EEG signal, and the main frequency is about 10 Hz. It is most obvious in the occipital region (see Fig. 3.3 of the human brain, and its amplitude ranges from 20 to 100 μV. The alpha rhythm reflects mainly a lot of thinking information of the human body and is easy to discover, so the alpha rhythm is a common rhythm of EEG research.

3.4.4 Beta Waves

The frequency range is $13 - 30\ Hz$, and the amplitude range is $5 - 20\ Hz$. In the frontal and temporal lobes, as the highest frequency of rhythm, it is often referred to as *"fast"*, appearing in the brain during excitement.

3.4.5 Gamma Waves

Gamma waves are fast oscillations (over $30Hz$) and are usually found during conscious perception. Due to small amplitude and high contamination by muscle artifacts, gamma waves are underestimated and not widely studied as compared to other slow brain waves. High gamma activity at temporal locations is associated with memory processes, but also with attention ,working memory, and long term processes. And it is, furthermore, involved in psychiatric disorders such as schizophrenia, Alzheimer's disease, and epilepsy. **However, it is harder to exploit Gamma frequency sub-bands using scalp EEG and is rather not included in research on EEG data except when invasive electrodes are included.**

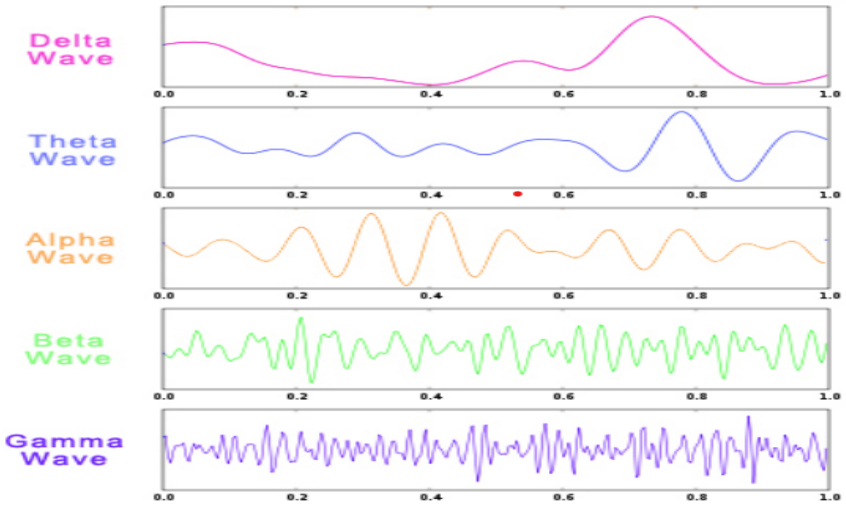


Figure 3.7: Brain waves visualization based on different frequency bands [48]

3.5 Stages of epilepsy in EEG signals

The characteristics of epileptic seizures can be studied by the analysis of the recorded EEG signals. EEG signals recorded just before and during seizures contain patterns which are different from those in a normal EEG signal recorded from a normal non-epileptic person. EEG analysis cannot only differentiate epileptic from normal data, but also distinguish these different abnormal stages/patterns of a seizure, such as **pre-ictal** (EEG changes preceding a seizure) and **ictal** (EEG changes during a seizure) [6]. In cases where patients have more than one seizure within a small duration, there is a stage called the **interictal** stage. Below, brief descriptions of each stage are presented [6]:

3.5.1 Preictal segments

A pre-ictal state becomes apparent during a certain time period before the occurrence of a seizure. It might not necessarily be visually apparent, however, it will reflect changes in the underlying signals which would be predictive of seizures within a certain time period going from 20 minutes to an hour. For a pre-ictal state to be of use clinically in a warning system, it has to be detected early enough so that the time under false warning is minimised.

3.5.2 Ictal segments

The ictal state is a change in EEG signals during a seizure, characterized by surges of electrical activities and mainly defined by its **onset**.

3.5.3 Interictal segments

The interictal is the stage between two successive seizure onsets. For two different patients, the number of epileptogenic neurons, cortical region, and the span of seizure can be altered differently.

3.5.4 Postictal segments

It simply defines the stage following the end point of the occurence of the seizure.

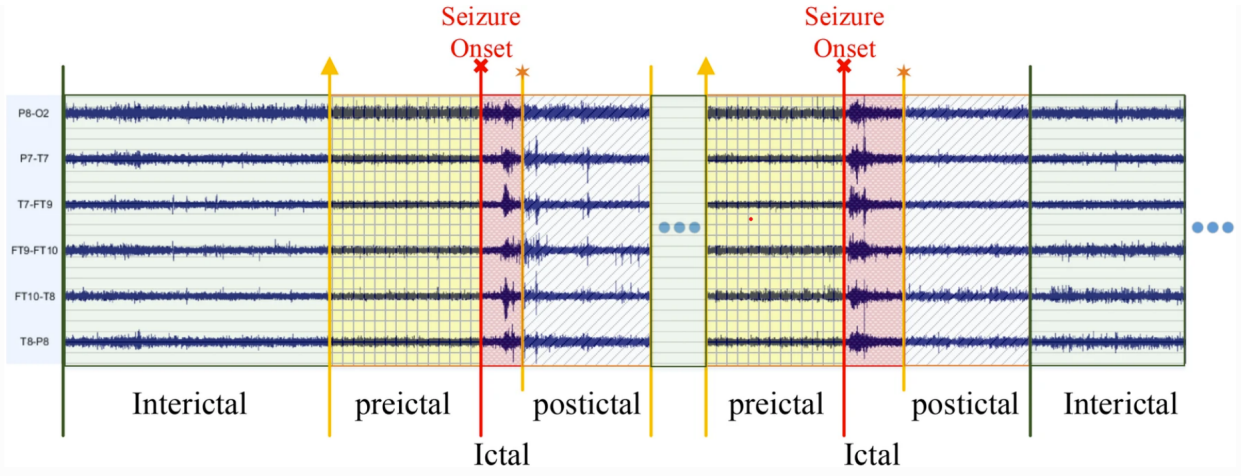


Figure 3.8: Showcase of the different stages of EEG signals of an epileptic patient [34]

The wave pattern, as we define it above, may hold valuable information about brain activity. Experienced neurologists can detect disorders by visually observing the EEG signals. However, this procedure is time-consuming and is prone to faulty detection due to high temporal and spatial aspects of the dynamic non-linear EEG data. Therefore, computerized techniques, EEG signal parameters extraction, and analysis can be profoundly beneficial in the diagnostics.

3.6 EEG signal characterization and analysis for epilepsy

An epileptic seizure, as defined by the ILAE -and presented in Chapter 2-, is a temporary event of symptoms due to synchronization of abnormally excessive activities of neurons in the brain. It has been estimated that approximately 65 million people around the world are affected by this disorder. Nevertheless, it is still a time-consuming process for neurologists to review continuous electroencephalograms (EEGs) to monitor epileptic patients. Therefore, several researchers have developed different techniques that help neurologists identify an epileptic occurrence. In this optic, we base off extensively our review of the literature and the methodologies explored based on the article presented in [22] and the proposed series of articles and approaches to study and review. This section focuses on a selection of features commonly used in the literature, including:

1. **Statistical parameters**: mean, variance, skewness,and kurtosis.
2. **Amplitude-related parameters** -energy, nonlinear energy, line length, maximum and minimum values.
3. **Entropy-related measures**.

These features can be categorized based on their interpretation or the domain from which they are calculated **- time, frequency or time-frequency domain**. While some studies consider specific groups of features according to the classification scheme they chose to go with, others have applied various groups extracted from different domain as: time, frequency, and time-frequency domains, all in one. For instance [2] proposes a framework of extracting 132 features, treating

feature redundancy and relevance analysis to reduce descriptors' vector dimension and as a result, 30-optimal features selected by $ReliefF$ achieved the best performance with 91% sensitivity and 95% specificity. These results show the need to assess feature selection in the state-of-the-art of the field given how crucial and directly related it is to the seizure classification performance.

Moving on to the nonlinear features illustrated when using entropy for epilepsy analysis, [4] reported on it extensively. Many types of entropy have been utilized in this case and so, both jointly and independently. [4] clearly stated entropies mathematically along with their advantages and disadvantages. The authors also illustrated differences of each feature between the normal (preictal, interictal) and ictal groups using the F-test that checks the equality of variances. The three highest ranked features were **Renyi's, sample, and spectral entropies**.

Moreover, the seizure detection techniques of the whole process, including feature extraction, feature selection, and classification methods, were summarized [7]. Feature extraction methods were divided in two spectral domains: time-frequency analysis and temporal domain. However, only a brief demonstration of each method with no mathematical explanation was depicted and a lack of the implementation description of the method subject of the paper. Later, characteristics of focal and non-focal seizure activities observed from EEG signals were also reviewed in . Later on, [3] provided a characterization of focal and non-focal seizure EEG signals retrieved respectively from the epilepsy-affected and unaffected areas. Nonlinear features, such as Hjorth parameters for instance, entropy, and fractal dimension, were used to characterize and compare the focal and non-focal EEG signals based on the Bern-Barcelona EEG database described in [12]. Yet, once again none of these features were expressed and clearly formulated and stated mathematically.

In addition, some **AESD** application studies have involved channel selection techniques for processing EEG signals because considering every channel may cause over-fitting problems. The channel selection techniques include filtering, wrapper, embedded, hybrid, and human-based techniques and were found to show a high accuracy and low computational cost function.

On another note, the use of deep learning, including ANNs, recurrent neural networks (RNNs), and Convolutional Neural Networks (CNNs), in biological signal processing, EEG for instance, were also reviewed within [37]. The author summarized the existence of two studies in this area of expertise focused on Autonomous Epileptic Seizure Detection and their prediction. Furthermore, methods for non-stationary signal classification using time-frequency and time-scale representations, and their differences were discussed in depth for seizure detection [21]. Several features were extracted: **spectrogram, Extended Modified B-Distribution, compact kernel distribution (CKD), Wigner-Ville distribution (WVD), scalogram, Affine WVD (AWVD), pseudo-smoothed AWVD, and CKD with feature selection, fed to a random forest (RF)**. The results showed that this framework achieved an accuracy superior to 80%, while CKD with 50 optimal selected features yielded an accuracy of 86.41%.

The paper review [38] mainly focuses on its part on the use of the wavelet-based method for seizure detection using different types and kinds of wavelets: **discrete wavelet transform (DWT) and continuous wavelet transform (CWT)**, with the use of **nonlinear-based** features and **chaos-based** measurements. The authors showed that the ***Daubechies*** wavelet proved to be the most efficient and relevant wavelet for seizure detection.

Additionally, in the feature extraction phase of the review developed in [14], the authors categorized features into four groups: linear univariate and bivariate, nonlinear univariate and bivariate

features. Along this line, a nonlinear measure is defined as an attribute of the dynamic states of the brain.

In conclusion, and from the past literature review, we conclude that many previous papers have applied a combination of features, however, while providing neither a clear conclusion about which feature has the highest contribution nor an intuitive meaning of using such features [83], [9], [10]. Additionally, mathematical expressions of some complicated features, such as entropies, and Hurst exponent (HE), were not consistently stated and so may lead to an incorrect implementation [6].

A more general synoptic scheme is presented in Fig. 3.9 and summarizes the standard worfklow for epileptic seizure detection utilizing Artificial Intelligence:

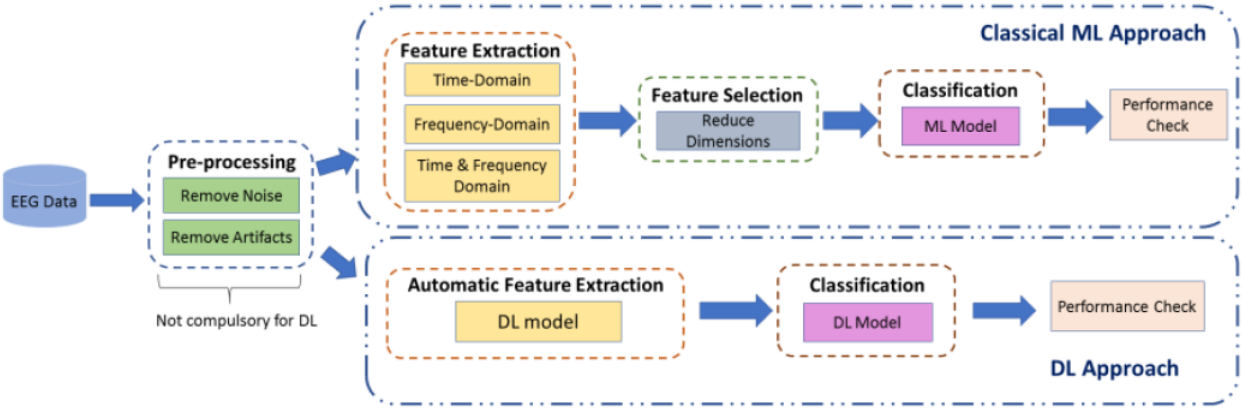


Figure 3.9: Process of epilepsy detection using EEG data and classification algorithm

3.7 Alternatives to EEG for brain activity study

With the impressive rise of interest in Brain-Computer Interfaces and a common scientific effort to further explore and exploit the human brain, multiple devices are now available to access cerebral activity, measure it and record its patterns through different types of technologies. Each has shown its ability to tackle a specific angle and range of patients according to their ages, disabilities or disorder and activity.

While we extensively discussed EEG, there are many alternatives that have been exploited, and proved their relevance as **Magneto-encephalography (MEG), Functional magnetic resonance imaging (fMRI)** and **Functional near-infrared spectroscopy (fNIRS)**. We provide a concise overview of each technology given how each can provide additional information, when coupled with EEG, for a more efficient epileptic seizure detection workflow [70].

3.7.1 Magneto-Encephalography (MEG)

It is a non-invasive technique allowing the measurement and analysis of the tiny magnetic fields generated outside the scalp by the working brain. MEG is obtained through very sensitive magnetometers-like superconducting quantum interference devices (SQUID). Each SQUID sensor contains a superconducting wire magnetically coupled to the SQUID, and produces a voltage proportional to the magnetic field received. A computer program converts the SQUID data into maps of the currents flowing throughout the brain as a function of time.

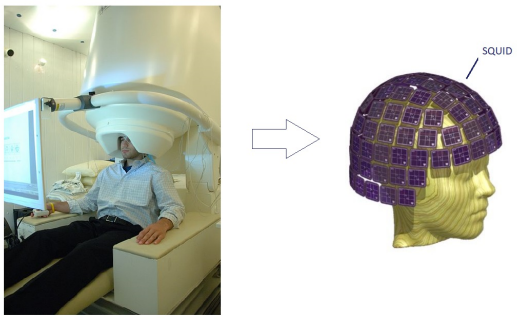


Figure 3.10: Magneto-Encephalography (MEG) recording

MEG has been used and exploited for epileptiform discharges detection [59] and for the detection of Alzheimer disease in its early stages [62]. Moreover, it is considered largely more efficient for epilepsy diagnosis and source localization of epilepsy [28] and more [23].

3.7.2 Functional magnetic resonance imaging (fMRI)

It is a non-electric technique for measuring brain activity by detecting the changes in blood oxygenation that occurs in response to neural activity. When neurons increase their activity with respect to baseline level, modulation of deoxyhemoglobin concentration is induced, generating the **blood oxygen level dependent (BOLD)** contrast. These techniques measure changes in the inhomogeneity of the magnetic field. Deoxyhemoglobin is of interest here; it introduces the said inhomogeneity into the nearby magnetic field. Hence, an increase in its concentration would cause an increase in image intensity. So, fMRI can be used to produce activation maps showing which parts of the brain are involved.

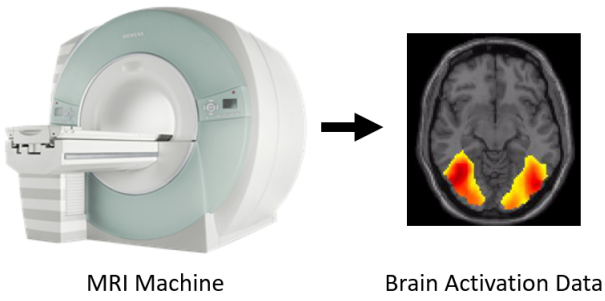


Figure 3.11: Functional magnetic resonance imaging (fMRI)

3.7.3 Functional near-infrared spectroscopy (fNIRS)

It is a non-invasive assessment of brain function by detecting changes in the associated blood hemoglobin concentrations. It provides information about oxygenation based on the optical properties of oxyhemoglobin and deoxyhemoglobin which have distinct absorption properties. The degree of oxygenation is determined by shining light and measuring the amount of light emerging. fNIRS serves the purpose of neuroimaging relying on the concept that it takes advantage of the optical window of $700 - 900nm$ in which all other components of the human body are mostly transparent, while hemoglobin and deoxygenated-hemoglobin absorbs light strongly. Differences in the absorption allow the measurement of relative changes in hemoglobin concentration through the use of light attenuation at multiple wavelengths.

Figure 3.12: Functional near-infrared spectroscopy (fNIRS)

fNIRS proposes a better temporal resolution than EEG, and a lesser spatial one and it is also portable. This offers concrete horizons to coupling fNIRS and EEG as it has been proven in other studies [61] for BCIs in its large sense. In general, fNIRS are used principally to measure the cerebral activity of the frontal lobe of the brain, but other models propose a measure on the whole scalp. fNIRS is gaining much interest when dealing with infantile disorders as for epilepsy or autism [33].

This section puts into perspective horizons for future work and possibilities to tackle problematics that go along with the present study with additional neurosphysiologic data sources.

Conclusion

From this chapter, we had the opportunity to explore Electroencephalography in all its extension and establish its crucial role in tackling Epilepsy with its different electrode models and placements and referential baselines. Moreover, the most important point in this chapter is the extensive explanation and presentation of the brain-wave frequency sub-bands and how important the place it can hold in biomedical engineering given how each sub-band has been identified to certain mental effort. On another note, the definition of the different stages of EEG in epileptic EEG data is offering precious tools for all axis of research in epilepsy; including epileptic seizure detection, epileptic spike detection, epileptic seizure onset detection and prediction. In our case, Ictal data is of upmost importance and is at the heart of our study and based on its distinction from the rest of the brain normal activity allow us to carry on more thorough investigations. As for the use of other different monitoring and recording devices along with EEG, this proves to be a hot-topic in the research field giving how, as we explained, each device has proved to be able to retrieve neural information differently.

Chapter 4

Review of the state-of-the-art algorithms

Introduction

In the latest decades, research efforts have been multiplied to design more advanced devices, imagine more fitting technologies to deal with epilepsy in medical environments and provide efficient contributions to assist medical experts and neuroscientists, and reduce the time taken to detect, diagnose and treat epilepsy as a whole disorder. Research communications in biomedical engineering highlights impressive results and performance over the years. Nevertheless, the transition of these designed devices into clinical use have shown a considerable fail and fall in their performance and their ability to efficiently generalize over biophysical data if provided with different distributions from the initial one they were trained with. In this chapter, other mistakes in the commonly shared approach have been diagnosed in order to better the transition from Engineering research and conception to daily clinical use. We are willing to first enumerate to our best knowledge issues and flaws that have been recognized within the biomedical engineering community concerning detection devices. Secondly, we will establish a diagnosis to a certain extent of the field to this day in terms of performance and naturally point out lacks we identify in each protocol. Finally, a synoptic diagram that summarizes our framework will be presented in order to lead the way for the upcoming parts of this report.

4.1 The nature of the datasets utilized

Most approaches involve mostly -often exclusively- **carefully selected short duration** EEG segments, result of a thorough preselection of overtly demonstrative electrographic seizure patterns.

4.2 Validation metrics

An additional aspect to take into consideration in biomedical and physiological problems is the **metrics** and **scores** used for validation of the approaches that are themselves too abstract and consider EEG signals the same way they would, an telecommunication signal. Most papers focused on a particular group of features that use the classification accuracy as the main indicator to conclude the effectiveness of those features. However, we believe that this cannot be a fair indicator for a method comparison. Indeed, an unrealistic validation regime such as testing on training data or excluding the worst performing records instead of considering them for in-depth microscopic studies to calibrate the algorithms accordingly and even better, identify the pathological specificities of the patient whose data seems to not go by our framework and therefore avoid

misplaced generalization of designed algorithms and workflows.

4.3 Seizure Detection: Detection Algorithm Assessments Using a Large Dataset

This upcoming section was designed to assess performance of commercially marketed (**Persyst 13, or** $P13$) automated seizure detector in comparison with the performance of multiple expert readers. The design of commercialized devices naturally include the incorporation of many new features and methodologies; including new artifact reduction technologies, by-channel processing to aid in the identification of low amplitude, focal seizure patterns, use of empirical null methodology to better capture the non-normal statistics of EEG patterns, improved and expanded training sets. Therefore, we summarize and further develop the comparison and evaluation results obtained and developed through the article [79].

The obtained results were summarized in Table 4.1.

Reader(Threshold)	Sensitivity% $\pm SD$ $(Range)$	FP/Day $\pm SD$ $(Range)$
A	84.9 $\pm$30.8 $(0.0 - 100.0)$	0.978 $\pm$2.914 $(0.00 - 28.06)$
B	73.7 $\pm$37.8 $(0.0 - 100.0)$	0.375 $\pm$1.800 $(0.00 - 20.49)$
C	72.5 $\pm$38.0 $(0.0 - 100.0)$	1.038 $\pm$3.417 $(0.00 - 28.08)$
Persyst 13	82.5 $\pm$33.2 $(0.0 - 100.0)$	11.32 $\pm$11.65 $(0.00 - 61.10)$

Table 4.1: Sensitivity and FPs for Readers and Two Seizure Detection Algorithms (Mean, SD, and Range) [79]

FP stands for ***False Positive*** and A, B and C stands for three distinct readers, experts in neurosciences and epilepsy. The concept of FP is of great importance in biomedical binary classification given it quantifies the rate of **false alarms** the detection system will emit.

We first do not omit to mention that as explained in [79] that the neuroscientist experts are considered mildly focused with the assurance that each reading will be proof-read and peer-reviewed by the other specialists members of the experiment protocol. Therefore, the sensitivity they exhibit and the FP/Day noted are to be taken with larger ranges.

We can first of all observe an average FP/Day rate considerably higher for the commercialized device $Persyst13$ in comparison with the performance of neuro-experts reading of EEGs. As for the $Sensitivity$, that stands for the **True Positive** rate predicted, It is narrowly higher for $Persyst13$ than most performances from neuroscientists. There are two points to deduct from this observation:

- First, it legitimizes the rise in research interest concerning autonomic and automatic physiological anomaly detection and the need to offer patients better, more precise treatment in facilities, but also open steadier paths for neuro-feedback systems and closed-loop neurostimulation; or simply support the reading of neurosurgeons and neuroscientists.
- Secondly, this performance comparison justifies the need to enhance research content already available, and, or search for new methodologies; and definitely supports the premise that the research axis of epileptic seizure detection is still wide and open to exploration for amelioration.

4.4 Performance comparison

In order to be capable of estimating how our framework performs in comparison with a certain number of research work conducted with a similar type of approach and relatively recent in order to be up-to-date. Moreover, we'll strive on emphasizing the lacks we note in each protocol.

First, the thorough study of [5] shows a clear similarity in the theoretical approach. Table 4.2 summarizes the performance, in various quantitative aspects of the designed workflow. The accuracy shows to be very satisfying with a majority of the accuracy rates above 90% along with high sensitivity and specificity values. However, as we can note, the number of samples that make the test data set is 90. This can be very easily considered too drastically limited for validation of the method or the affirmation that the method, as designed can encompass more generality present within EEG signals.

Classifiers	TN	FN	TP	FP	Accuracy	Positive predictive	Sensitivity	Specificity
Fuzzy	30	0	60	0	98.1	100	value (%)	(%)
SVM	30	2	58	0	95.9	100	99.4	100
KNN	29	1	59	1	93	98.9	97.2	100
PNN	29	1	59	1	93	98.9	97.8	97.8
DT	27	1	59	3	88.5	95.7	97.8	97.8
GMM	29	1	59	1	95.9	97.9	98.3	91.1

Table 4.2: Performance measures obtained using the seven classifiers [5]

In [50], the classification accuracy reached is 90%. The most evident remark to make here is the space for improvement. However, a little over 40 samples were used for testing and validating, which cannot account for generalization. Additionally, not the totality of the time-embedding hyper-parameters were optimized, rather only the embedding dimension m.

One equally interesting conducted and published work on the thematic is presented in [87]. Their average accuracy across different patients is 95.3%: though we can observe more detailed results in Fig. 4.1. In the latter, we can observe accuracy scores that can fall drastically for some patients (slightly above 70%) which is dangerous for the general performance estimation of the proposed workflow, specifically with no study of the "outliers", these subjects that show this "unusual" bad performance and understand further what are the pathological limitations of the designed approach. By their own iteration, this performance is considered a little below newer papers' performance. Additionally, as they also explain, there were no treatment of the intrinsic imbalanced characteristic of the epileptic seizure dataset and its relative poverty in terms of volume and number of samples.

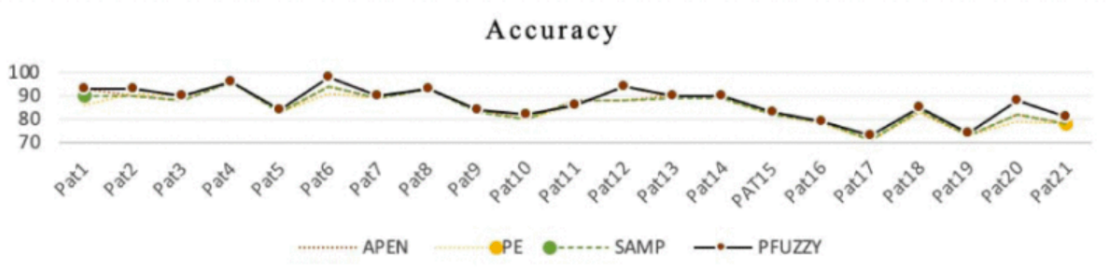


Figure 4.1: Accuracy measure for all entropy variants of all 21 patients [87]

The last paper we considered in this comparative study is [31]. The approach described in this paper is led exhaustively and converges with our own on many angles. First and foremost, Table 4.3 summarizes the performance of the designed protocol on the different classifiers:

Methods	Accuracy (%)	Sensitivity (%)	Specificity (%)
LS-SVM	99.50	100.00	99.40
KNN	97.90	99.80	94.00
LR	99.00	100.00	98.00
LDA	99.00	100.00	99.00
NB	91.00	98.00	84.00
RF	97.00	99.00	95.00

Table 4.3: Classification accuracy, sensitivity, and specificity presented by the six classifiers by using the 18 selected features for training and testing [31]

Furthermore, the article also provides a performance summary gathered from relevant papers to the axis chosen -and therefore relevant to our own process. It's presented in Table 4.4

Methods	Year	Accuracy (%)	Sensitivity (%)	Specificity (%)
Proposed method	2019	99.50	100.00	99.40
Wang et al. [86]	2017	99.25	97.98	99.56
Jaiswal et al. [49]	2017	99.07	98.82	99.32
Swami et al. [82]	2016	98.00	96.39	98.57
Samiee et al. [78]	2015	96.50	95.60	97.40

Table 4.4: Comparison of classification accuracy, sensitivity, and specificity obtained by the proposed method and other existing state-of-art methods using the same dataset [31]

One important limitation in [31] is first the lack of optimization of the hyper-parameters for the computation of the nonlinear features. This is absolutely different from tuning classifiers' parameters given the confidence in our features at the heart of any ML approach, in its ability to generalize lies in avoiding the hard-coding of parameters for features specifically. On another note, their approach to tackle Imbalanced Learning is undoubtedly interesting and worth more attention by opting for an approach on the algorithm-level though its computational cost may be up to debate. In conclusion, the results noted in both tables give us enough content and data for future performance comparison.

4.5 Framework summary

Having established a state-of-the-art in terms of types of features utilized and in terms of accuracy and performance of cited methodologies, we can summarize and present synoptically the protocol we mean to carry out in this research case. There are central points and peripheral interests in correlation between features we consider crucial to understand our way of quantifying information in signals and more.

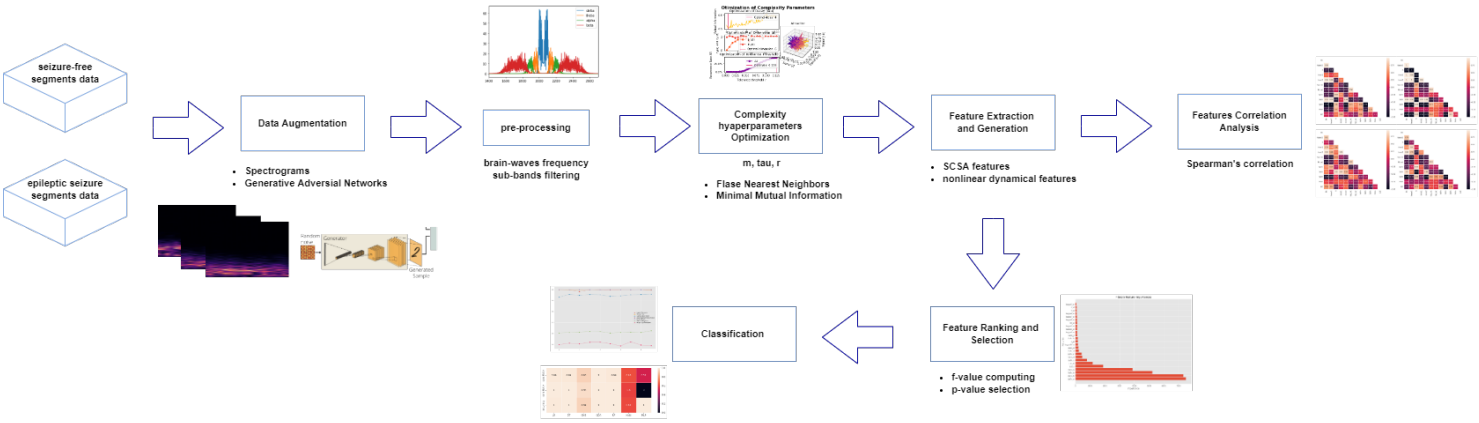


Figure 4.2: Synopsis of the proposed methodology

This process will be extensively developed, explained and explored in the following report and document through its various chapters.

Conclusion

We are capable at the conclusion of this chapter to confidently carry on the development of our framework knowing limitations that have been identified and the accuracy and performances noted to this day and with a background explicit this way, we can compare and position ourselves in the future, within the state-of-the-art of the field.

Chapter 5

Datasets Description

Introduction

In this chapter, we present the most crucial part of our study and what our results revolve around: the datasets of epileptic EEG. We will present and offer an exploration of each of the database we considered during our study case with an organized showcase of the specificities of each. This chapter is critical in our report and reflexion.
We consider two datasets for the sake of international generalization: the Bonn University dataset and the CHB-MIT dataset collected at the Children's Hospital Boston from pediatric subjects with intractable seizures

5.1 Bonn University epilepsy Database

Five sets, denoted **A** to **E**, and each containing 100 single channel EEG segments of $23.6 - sec$ duration, were composed for the study [13]. These segments were selected and cut out from continuous multichannel EEG recordings after visual inspection for artifacts -due to muscle activity or eye movements: electrodes containing pathological activity for C, D and E, and those exhibiting strong eye movement artifacts in A and B were omitted. The data were then written continuously onto the acquisition system at a sampling rate of $173.61Hz$. Additionally, band-pass filter settings were $0.53 - 40Hz$ $(12dB/oct)$.
Five patients were selected given they all achieved complete seizure control after resection of one of the hippocampal formations, which was therefore correctly diagnosed to be the epileptogenic zone.

Sets **A** and **B** consist of segments taken from surface EEG recordings that were carried out on five healthy volunteers using a standardized electrode placement scheme while sets **C, D**, and **E** originated from our EEG archive of presurgical diagnosis.

The table 5.1 summarizes the type of EEG utilized to extract and record each subset and the condition of the patient during the experimental protocol.

subset	EEG type	state of the patient
A	single-channel scalp EEG	relaxed and eyes open
B	single-channel scalp EEG	relaxed and eyes-closed
C	invasive single-channel EEG	seizure-free, hippocampal formation of the opposite brain hemisphere
D	invasive single-channel EEG	seizure-free, from the diagnosed epileptogenic zone
E	invasive single-channel EEG	epileptic seizure activity

Table 5.1: Summary of the Bonn University dataset

Exemplary EEGs from each subset of the database are depicted in Fig. 5.1:

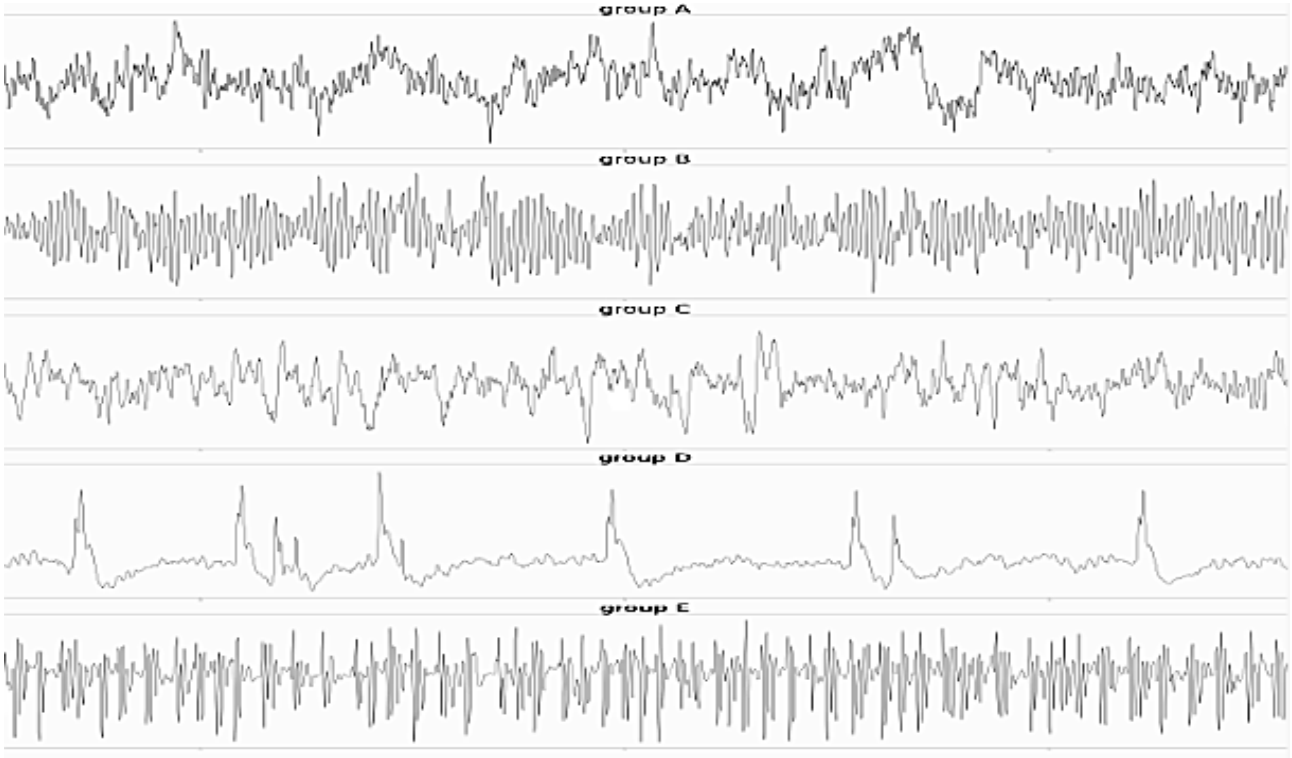


Figure 5.1: Exemplary time-series of each subset A to E

We consider in our study the four first subsets A, B, C and D as **non-epileptic** data and the last subset E exclusively composed of **ictal** data will be the **epileptic** data.

5.2 CHB-MIT Scalp epilepsy Database

This **CHB-MIT** database, collected at the **Children's Hospital Boston**, consists of EEG recordings from **pediatric** subjects with intractable seizures [80]. Subjects were monitored for up to several days following **withdrawal of anti-seizure medication** in order to **characterize their seizures** and assess their candidacy for surgical intervention.
Recordings, grouped into 23 cases, were collected from 22 subjects (5 males, ages $3-22$; and 17

females, ages $1.5 - 19$). a file is joined to the database and contains the gender and age of each subject.

Each case (*chb*01, *chb*02, etc.) contains between 9 and 42 continuous files from a single subject. Naturally, in order to protect the privacy of the subjects, all protected health information (PHI) in the original files has been replaced with surrogate information. In most cases, the files contain exactly **one hour** of digitized EEG signals,with some exceptions as its is the case for *chb*10 with recording up to two hours long, and those belonging to cases *chb*04, *chb*06, *chb*07, *chb*09, and *chb*23 are four hours long. Files within which we can find seizures recorded proved to be shorter.

All signals were sampled at 256 samples per second and most files contain 23 EEG signals (24 or 26 in a few cases). The International $10 - 20$ system of EEG electrode positions and nomenclature was used for these recordings. In a few records, other signals are also recorded, such as an ECG signal and a vagal nerve stimulus (VNS) signal.

To summarize the content of the database, we have 664 files included in this collection, and the file and among them, 129 contain one or more seizures. In all, these records include 198 seizures; with the beginning and end of each seizure annotated. In addition, a set of files named provided present information about the montage used for each recording.
We can exhibit for the sake of illustration the occurrence of an epileptic seizure in the middle of interictal EEG data for two distinct patients *chb*07 in Fig. 5.2 and *chb*09 in Fig. 5.3.

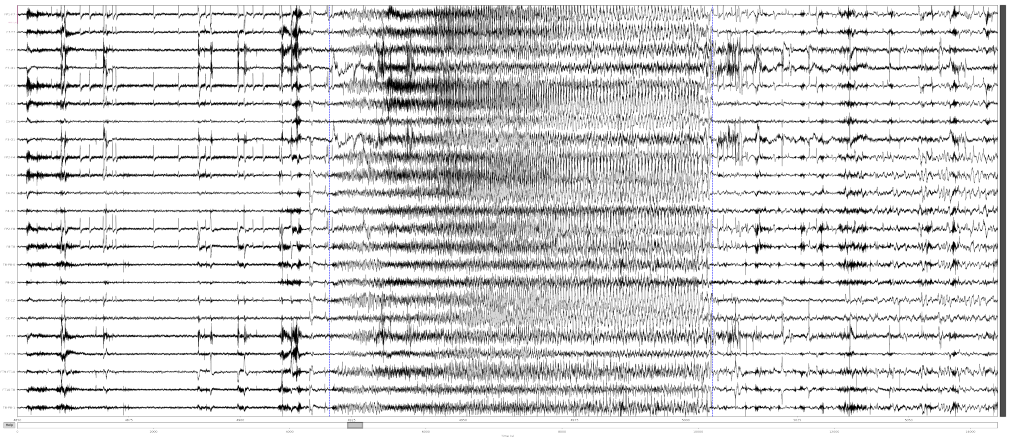


Figure 5.2: Plotting of EEG data segments for patient n7

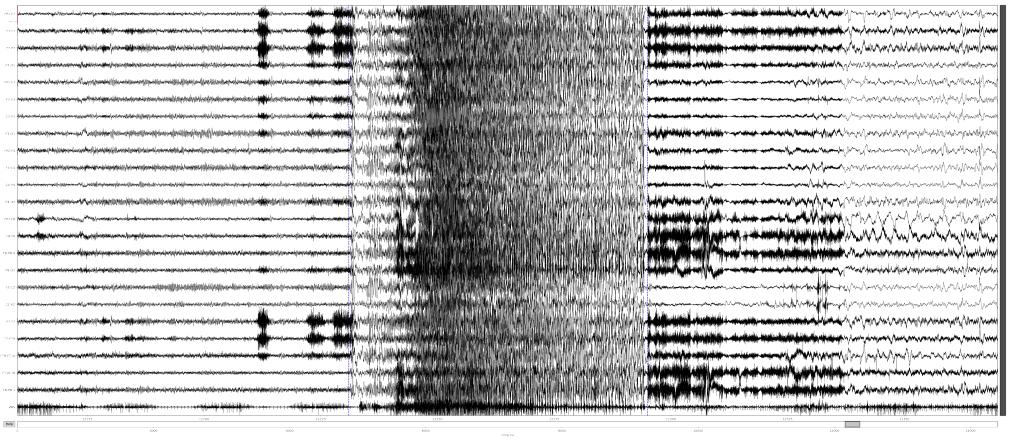


Figure 5.3: Plotting of EEG data segments for patient n9

Conclusion

Now that we have an acceptable understanding of our datasets, we must state that due to the extreme volume of the database, we have not been able to exploit the CHB-MIT database yet in our study due to computational constraints. Future work includes a run of our workflow using Ibex from KAUST that provides heterogeneous nodes to run our nodes by putting at our disposal greater computing resources.

Definition 2 *Ibex is a resource cluster that contains different architectures of CPUs like Intel Cascade Lakes, Skylakes, AMD Rome. These different CPUs are accessed (for source code compilation and job submission) using the following login node.*
The IBEX cluster CPU compute nodes are accessed by the SLURM scheduling using the constraint by directly specifying the CPU type as shown below.
Therefore, it provides resource remotely accessible to put less computational pressure on our machines and run our codes (Matlab, C, C++, Python and Jupyter Notebooks) faster.

Part II

Features Engineering

Chapter 6

Semi-Classical Signal Analysis

Introduction

In this chapter, we present a novel and recent algorithm for signal representation based on the Schrodinger operator: the **Semi-Classical Signal Analysis**, with its formulation and the mathematics behind, its properties and moreover results of the practical application and simulation we obtain. an upmost interesting characteristics of the SCSA is its ability to be derived for other signal processing applications as denoising. We will also tackle this important facet of SCSA and explain our interest in the method for feature extraction from physiological signals.

The Semi-Classical Signal Analysis (SCSA) introduces a novel signal presentation and analysis method, based on a semi-classical approach [55] that suggests the decomposition -and reconstruction- of a signal $y(t)$ into -and through- a set of functions given by the squared eigenfunctions of the Schrodinger operator, and the negative eigenvalues associated; is possible.

Definition 3 *Let $y(t) > 0$, $t \in \mathbb{R}$ be a real-valued function and $-y(t)$ a Schrodinger potential. The reconstruction of $y(t)$ according to the Semi-Classical Signal Analysis (SCSA) decomposition is $y_h(t)$, and is defined as:*

$$y_h(t) = 4h \sum_{n=1}^{N_h} \kappa_{nh} \psi_{nh}^2(t), \quad x \in \mathbb{R} \tag{6.1}$$

Where ψ_{nh} are the L^2-normalized eigenfunctions of the spectrum of the Schrodinger operator $H_h(y)$ and κ_{ih} the square root of the negative eigenvalues associated, and are denoted:

$$-\kappa_{nh}^2 \quad with \quad \kappa_{nh} > 0 \quad and \quad \kappa_{1h} > \kappa_{1h} > \cdots > \kappa_{nh} \quad n = \overline{1 : N_h} \tag{6.2}$$

and $N_h(y)$, the number of negative eigenvalues of the operator and most importantly h, is the semi-classical -quantum- parameter.

The key parameter in the reconstruction equation 6.1 is h, **the semi-classical parameter**. We will discuss further into the chapter later how the **quantum parameter** h plays a crucial role within this reconstruction approach: indeed, the approximation $|y - y_h|$ improves as h decreases.

We present now the basic principles behind the development of this novel signal representation and the mathematical background to it.

The signal $y(t)$, at the heart of the SCSA method and to be reconstructed using 6.1, is under some assumptions presented in [55] considered as a **multiplication operator** on some function

space:

$$\phi \to y.\phi \tag{6.3}$$

This condition on y now established, the discrete spectrum of the Schrodinger operator $H_h(y)$ is defined, for a small h, by:

$$H_h(y)\psi(t) = -h^2\frac{d^2\psi(t)}{dt^2} - y(t)\psi(t), \quad \psi \in H^2(\mathbb{R}), \quad h > 0 \tag{6.4}$$

where $H^2(\mathbb{R})$ denotes the Sobolev space of order two -which happens to be the Hilbert space.

H_h is the spectrum of the Hamiltonian operator $\hat{H}$ the equation is constructed around:

$$\hat{H} = -h^2\frac{d^2}{dt^2} + V(\vec{r}) \tag{6.5}$$

Where $V(\vec{r})$ is a Schrodinger potential -in our case, we consider $-y(t)$ by identification.

The SCSA owes its name to what we refer to as the "Semi-Classical" concept the method is, therefore, based on:

> ***" In quantum mechanics, semi-classical is a concept used to assess the fact that the actions of the studied system are large compared to the quantum h".***

Mathematically, this can be interpreted as the asymptotic expansions of the actions in a neighborhood of $h = 0$.

6.1 SCSA properties

Proposition 1 *The number N_h of negative eigenvalues is a decreasing function of the semi-classical parameter h based on the equality:*

$$\lim_{h\to 0} hN_h = \frac{1}{\pi}\int_{-\infty}^{+\infty}\sqrt{y(t)}dt \tag{6.6}$$

Proposition 2 *One interesting property of the negative eigenvalues is the notification that:*

$$-y_{\max} \leq -\kappa_{nh}^2 < 0 \tag{6.7}$$

Hence, for a fixed value of h, κ_{nh}^2 can be interpreted as particular values of y as illustrated in Fig 6.1. This is therefore how the SCSA appears to be a novel approach for quantifying a real-valued signal.

Proposition 3 *The study of the convergence of the momentums of κ_{nh} when $h \to 0$, which are referred to in semi-classical literature as invariants of the KdV equation and which are in more common mathematical terms ; Rietz sums proved to be interesting in the analysis of our signal of interest. Hence, under further hypothesis, we have:*

$$\lim_{h\to 0} h\sum_{n=1}^{N_h}\kappa_{nh} = \frac{1}{4}\int_{-\infty}^{+\infty} y(t)dt \tag{6.8}$$

$$\lim_{h\to 0} h\sum_{n=1}^{N_h}\kappa_{nh}^3 = \frac{3}{16}\int_{-\infty}^{+\infty} y^2(t)dt \tag{6.9}$$

the first and second order invariants for the KdV equation.

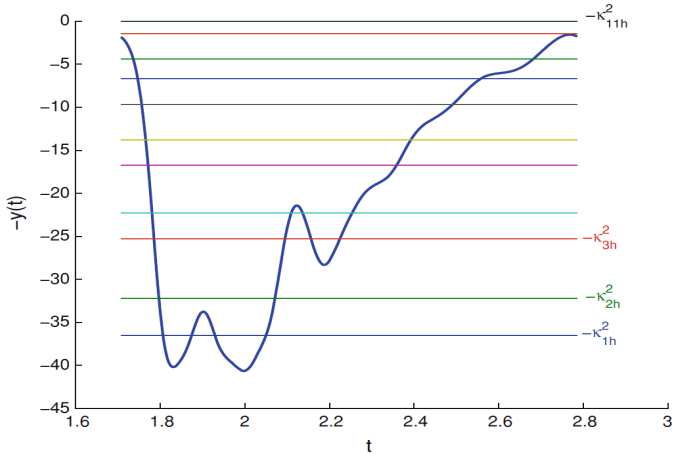


Figure 6.1: Signal quantization using the negative spectrum of Schrödinger operator $H_h(y)$ [55]

6.2 Implementation of the SCSA

In this section, we are interested in the practical implementation of the Semi-Classical Signal Analysis and its validation through numerical examples that include simulated and real data.

This section includes a presentation of the SCSA algorithm, a set of observations which ensure a better comprehension of the practical aspect of the SCSA and further remarks.

The problem considered is the one associated to $H_1\left(\frac{y}{h^2}\right)$ instead of $H_1(y)$ because it proved to be more convenient.
For a simplified notation, we denote $\chi = \frac{1}{h^2}$, $N_h = N_\chi$ and $\kappa_{n\chi} = \frac{\kappa_{nh}}{h^2}$ and we proceed with replacing it in the formula 6.1.

6.2.1 Numerical scheme

As we developed earlier in this chapter, the one-dimensional Schrödinger equation is brought to the solving of the spectral problem stated in 6.4. The discretization of this operator leads to an eigenvalue problem of a matrix.

We consider a grid of M equidistant points t_j, $j = \overline{1:M}$ such that:

$$a = t_1 < t_2 < \cdots < t_{M-1} < t_M = b \tag{6.10}$$

for which we denote $y_j = y(t_j)\,, \quad \psi_j = \psi(t_j)\,, \quad j = \overline{1:M}$ the values at the grid points and the distance between two consecutive points, the step: $\Delta t = \frac{b-a}{M-1}$.

Now, we can translate our Schrödinger problem as an eigenvalue matrix problem formulated:

$$(-\boldsymbol{D}_2 - \chi\,\mathrm{diag}(\boldsymbol{Y}))\,\underline{\psi} = \lambda\underline{\psi} \tag{6.11}$$

Where:

– $\mathrm{diag}(\boldsymbol{Y})$ is a diagonal matrix of the values y_j, $j = \overline{1:M}$.

– $\underline{\psi} = \begin{bmatrix} \psi_1 & \psi_2 & \cdots & \psi_{M-1} & \psi_M \end{bmatrix}^T$.

– $\boldsymbol{D}_2$ is the second order differentiation matrix given by:

– If M is even:

$$D_2(k,j) = \frac{\Delta^2}{(\Delta t)^2} \begin{cases} \frac{-\pi^2}{3\Delta^2} - \frac{1}{6} & \text{for } k = j \\ -(-1)^{k-j} \frac{1}{2} \frac{1}{\sin^2\left(\frac{(k-j)\Delta}{2}\right)} & \text{for } k \neq j \end{cases} \tag{6.12}$$

- If M is odd:

$$\boldsymbol{D}_2(k,j) = \frac{\Delta^2}{(\Delta t)^2} \begin{cases} \frac{-\pi^2}{3\Delta^2} - \frac{1}{12} & \text{pour } k = j \\ -(-1)^{k-j} \frac{1}{2} \frac{1}{\sin\left(\frac{(k-j)\Delta}{2}\right)} \cot\left(\frac{(k-j)\Delta}{2}\right) & \text{pour } k \neq j \end{cases} \tag{6.13}$$

with $\Delta = \frac{2\pi}{M}$ and according to 6.12 and 6.13, $\boldsymbol{D}_2$ is symmetric and definite negative.

The SCSA implementation is also a matter of optimization in order to find and localize the optimal value of h - or χ - that ensures a robust approximation with a minimal number of negative eigenvalues, based on a simple cost function of mean squared error- and that, in order to undertake a minimal computational cost.

6.2.2 The SCSA algorithm

Here is presented the Semi-Classical Signal Analysis algorithm:

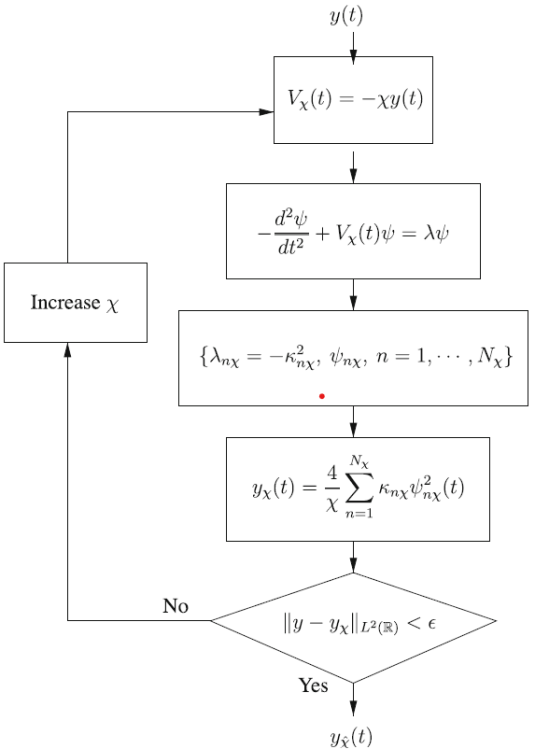


Figure 6.2: The Semi-Classical Signal Analysis algorithm [55]

6.2.3 Illustrative Examples

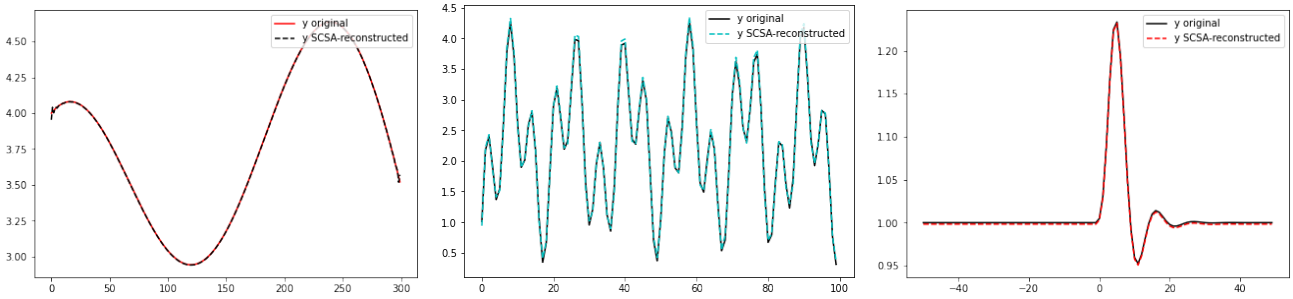
In order to illustrate the reconstruction capacity of the SCSA on the simplest signals and observe the reconstruction performance on signals with different morphologies.

An important characteristic of the SCSA to consider before applying it to any signal, is that it only performs and works with positive signals which makes it obligatory to perform a simple beforehand translation $y - y_{min} > 0$ and the Schrodinger operator potential to be considered is $-\chi(y - y_{min})$.

The first set of reconstructions are carried on standard functions we enumerate as:

1. A sine wave function $y_1(t) = cos(\frac{t}{40}) + 3$
2. A combination and sum of trigonometric functions $y_2(t) = sin(t) + cos(\frac{20t}{3}) + 0.3tanh(6t) + 2$
3. A pulse-shaped wave function $y_3(t)$.

In this first stage of results, the optimization problem we explained earlier for the tuning of the quantum parameter h is put aside and h is quickly manually tuned for observation sake.



(a) sinusoidal wave

$h = 0.8$ $N_h = 231$
$\hat{mse} = 1.89.10^{-6}$

(b) sum of trigon. functions

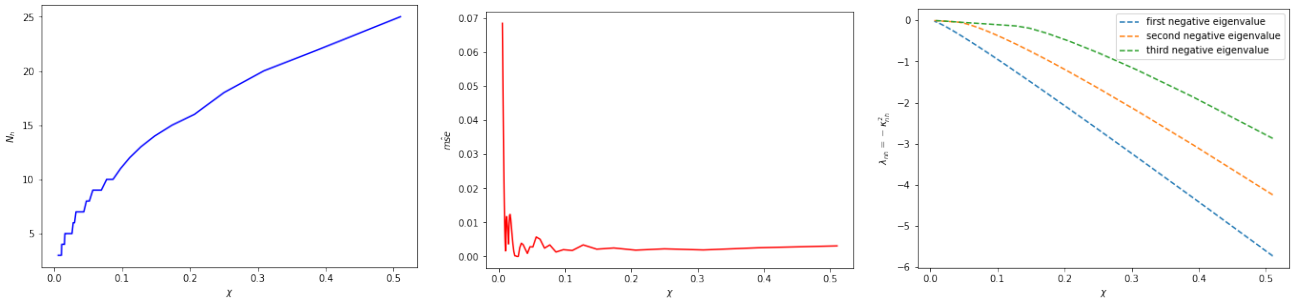
$h = 0.7$ $N_h = 134$
$\hat{mse} = 7.82.10^{-5}$

(c) pulse-shaped signal

$h = 0.5$ $N_h = 63$
$\hat{mse} = 2.92.10^{-6}$

Figure 6.3: Reconstruction results using the Semi-Classical Signal Analysis

More in-depth investigations are conducted on the pulse-shaped signal. We scan over a wide range of values for $\chi = \frac{1}{h^2}$ and plot N_χ, the number of negative eigenvalues, the mean square error and the evolution of the values of the first three $-\kappa^2_{n\chi}$.



(a) number of negative eigenvalues N_χ

(b) mean square error registered $\hat{mse}$

(c) The two first eigenvalues $-\kappa^2_{ih}$ for $i = 1, 2, 3$

Figure 6.4: Variations of different parameters according to χ for pulse-shaped signal

6.2.4 Remarks and results

Based on the simulation results above, further conclusions have been made regarding the Semi-Classical Signal Analysis method:

- N_χ is an increasing function of χ as stated and demonstrated in the equation 6.6 and from 8.6 (a), it is a step-by-step function that diverges when h attains relatively very small values.
- In 6.1, we stated that the reconstruction of the signal using SCSA improves as h decreases towards 0. This statement is more relevant when taking into consideration the following remarks made:
  - The rank of the matrix subject of the eigenvalue problem 6.11 is M. Therefore, the maximum number of eigenvalues N_h attainable when h decreases is M. If h decreases beyond a certain bound, the necessary number of eigenvalues to consider will exceed the number available ($i.e$ M) and the reconstruction becomes, hence, impossible.
  - We note from Fig. 6.3 that the magnitude of h is highly related to the amplitude of the peaks and the signal in general. Therefore, the smaller h is relatively to the amplitude of the signal of interest, the better is the SCSA-approximation.
  - A too small value of h, with respect to the previous point, results in a flat reconstruction of the signal.
- From 8.6 (b), there are specific values for which the mean-square error registered is minimal and therefore the reconstruction optimal. These are, in fact, the values for which the Schrodinger potential considered is what we call: a reflectionless potential.
- One of the SCSA's most attracting characteristics is its portability between domains: the method is applicable in the exact same way in any domain we might be tempted to use it in. The time and frequency domains, being more frequently used when working with raw time-series biomedical signals, SCSA has proved to be just has efficient and relevant in both domains.

6.2.5 Quantum-based h-interval selection for the SCSA method

The SCSA algorithm's design parameter, h, has for now, been solely tuned heuristically with no defined guidelines on the interval of searching for the optimal value of h, in the sense of the error between the original signal and the decomposed one 6.1 and the knowledge of the quantum mechanical principles residing in the SCSA formulation is left unused. The SCSA framework is now extended by calculating the bounds of the reconstruction parameter and providing an interval for h. By identification and parallelism, the derived bounds are effectively the sampling theorem for SCSA, which is of paramount importance for the application of the theory.

Now, both the minimum and maximum value of h are derived. The definition of these two bounds solves the problem of an undefined greedy searching for the optimal value of the representation.

As SCSA is heavily based on the Schrodinger eigenvalue problem, the interval is deduced from quantum mechanical principles.

Proposition 4 *The lower bound -the minimum value of h-, is derived of the sampling theorem. The reasoning behind the calculation of the maximum frequency was to derive the value of h so that no aliasing is introduced in the eigenfunctions:*

$$\frac{1}{T_s} \geq 2\beta \frac{\omega_{\max}}{2\pi} \tag{6.14}$$

with $\beta \geq 1$ being the over-sampling factor and T_s, the sampling frequency. The lower bound is retrieved as:

$$h_{\min} = \beta \frac{T_s}{\pi} \sqrt{y_{\max}} \tag{6.15}$$

Proposition 5 *As for the approximation of the upper bound, it is assumed that the choice of h brings the first eigenvalue to the maximum value of the signal (negated). As the perturbation theory collapses when the perturbation is large (the smaller the h, the larger the effective y from 6.1), the result will yield the smallest of the upper bounds of h. From (13):*

$$h_{\max} = \frac{T_s}{\pi} M \sqrt{\frac{\sum_{n=1}^{M/2} \frac{|c_n|^2}{2n^2}}{y_{\max} - c_0}} \tag{6.16}$$

Where M is the number of points and c_n are the complex Fourier coefficients.

Now the bounds mathematically defined, it is interesting to give each value and formula more contextual sense. h_{min} is directly bound to the maximal amplitude within our signal of interest.

As for h_{max}, its is directly bound to the complex Fourier coefficients of $y(t)$. Intuitively and knowing the greater h is, the lesser number of negative eigenvalues of the spectrum $H_h(y)$ are used for the reconstruction, limiting h means the greater bound acts as a low-pass filter and therefore allows a minimal amount of noise to pass through when reconstructing for the quantum value of h_{max}.

The most important bound of the two to determine is the lower bound h_{min} and that, because it is the value that can give the best possible reconstruction -which is not necessarily optimal given the computational cost can be less for a value of h_{opt} existing within the interval we defined in this section, that respects the constraint defined for the $m\hat{s}e$.

6.2.6 Signal denoising based on the curvature constraint and the SCSA

In this section, we propose to extend the SCSA formalism to the denoising of different signals, in particular **pulse shaped signals**, by including an **optimization** that considers **curvature constraints** [58]. The performance of the method has been measured by comparing its performance with other denoising methods. Results have indicated that the proposed method not only produces good denoising performance but also guarantees the peaks are well preserved in the denoising process.
The C-SCSA proposes a novel cost-function J:

$$\bar{J} = \sum_{i=i}^{N} \left[y_\delta \left(t_i \right) - y_h \left(t_i \right) \right]^2 + \mu \sum_{i=1}^{N} k \left(t_i \right) \tag{6.17}$$

where $k(t_i)$ is a certain **smoothness penalty term** which operates on the reconstructed signal y_h. $\sum_{i=1}^{N} k \left(t_i \right)$ describes the **"wiggliness"** of the reconstructed signal y_h, μ is a **non-negative smoothing parameter** that needs to be properly selected. It depends on the characteristic of

the input signal. Larger values of μ force y_h to be smoother. In this method, we define smoothness penalty term $k(t)$ to be the curvature:

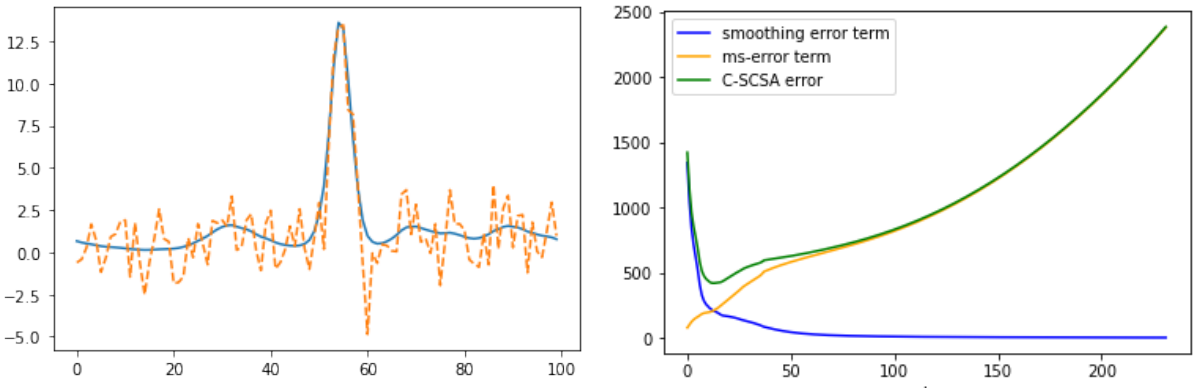
$$k(t) = \frac{|y_h''(t)|}{(1 + y_h'(t)^2)^{\frac{3}{2}}} \tag{6.18}$$

Where y_h' is the first and y_h'' the second derivative of the reconstructed signal.

The definition of this cost function allows us to propose a scanning method which iteratively scans h to minimize the cost-function J and so by balancing both terms in the cost function: the fidelity term and the smoothness term.

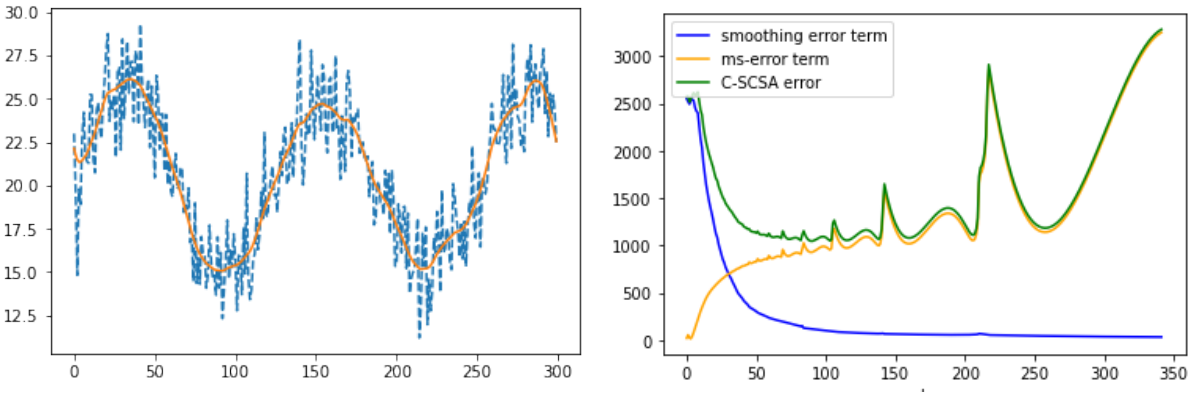
The most challenging aspect of the C-SCSA is the tuning of the hyper-parameter μ which allows to balance between a SCSA reconstruction that is loyal to the original signal and a smoothing of the noise that pollutes the local smooth variations within y.

We decide to run simulations on types of



(a) reconstruction and denoising results (b) cost function with both its terms

Figure 6.5: C-SCSA denoising simulations on a pulse-shaped noisy signal



(a) reconstruction and denoising results (b) cost function with both its terms

Figure 6.6: C-SCSA denoising simulations on a sine-wave noisy signal

6.3 SCSA features for epileptic seizure detection

Now that the Semi-Classical Signal Analysis theory and analytic formulation have been thoroughly developed and discussed in the section above, the existence of various features deriving from this approach being available pushes us to utilize them in both time or frequency domains. Secondary features are also deducted from the ones already defined with the purpose of discriminating epileptic seizure and seizure-free signal data.

SCSA features have successfully been used in the characterization of a wide range of biophysical signals and is gaining wider when processing with electrocardiograms [56, 57], electroencephalography [59], MEG signals [29] and also MRI [19] and MRS [20, 30].

These results motivate us further in including them in our framework with their multiple variants:

Table 6.1: SCSA features

Feature	Feature definition	Choice motivation
κ_{1h}	first squared root eigenvalue	coincides with the amplitude of the highest peak in the considered frame
N_h	number of eigenvalues	N_h is dynamical and depends on the morphology of the signal
h	semi-classical reconstruction parameter	relative to the signal, its peaks, its local variations, amplitudes $\ldots etc$
$I_1(y(t))$	first invariant	the integral of the signal on the frame and proved to hold relevant information in [56]
$I_2(y(t))$	second invariant	integral of the square signal and refer to [56]
N_h_rat	$N_h\_rat(i) = \frac{N_h(i+1)}{N_h(i)}$	conveys the variation in the number of samples necessary for reconstruction between 2 consecutive windows
κ_1_{rat}	$\kappa_{1h}\_rat(i) = \frac{\kappa_{1h}(i+1)}{\kappa_{1h}(i)}$	conveys the variation in the amplitude of the highest peak between 2 consecutive windows

Conclusion

With a clear idea of what the SCSA is and what it proposes for physiologic signal processing; as in our case neural epileptic data; we were able to define the set of SCSA features to exploit and have mathematical background that allow us confidence in their manipulation.

Chapter 7

Nonlinear dynamical features

Introduction

Based on the problem formulation proposed and developed in earlier chapters -most definitely in section 2.2 - and the definition of epileptic seizures from a dynamical, chaotic, nonlinear point of view, we are motivated and capable to legitimize our interest in nonlinear dynamical features which are already well established within this research axis. Failing to tackle epilepsy through a rigorous, reliable mathematical model to this day, these features allow us to quantify the dynamics of the signal we process.

7.1 Entropy

Multiple feature extraction and data transformation techniques have been utilized in order to build more efficient and relevant frameworks for epileptic seizure detection. Many recent studies and frameworks, for that matter, relied on one single feature: entropy and by combining its multiple sub-types [5], sometimes up to eight of them [31]. A number of different entropy estimators, with both its categories spectral and embedding, have been used in the context of quantifying the complexity of the signal [50] and handling nonlinear and chaotic systems [67], that by relying on the definition of entropy itself.

Entropy is defined as a thermodynamic quantity describing the amount of disorder in a system [50] because it ultimately measures the randomness and amount of uncertainty associated with a variable X when only its distribution is known. That's also why -and how- entropy is a function of the probability distribution $\{p_1, p_2, \dots\}$ and not a function of the values of the variable in the particular series $\{x_1, x_2, \dots\}$.

In other words, entropy is a nonlinear measure suitable for studying chaotic behavior and randomness, suits well to analyze perceptual time series data such as EEG data [87].

7.1.1 Spectral Entropy

According to information theory, Spectral Entropy (SpEn) describes the complexity of a system and is applied to the power spectrum of a signal. The feature extraction by spectral entropy consists of four (4) steps:

1. **Step 1:** Calculate the spectrum of the signal $X(\omega_i)$

2. **Step 2:** Calculate the Power Spectral Density (PSD):

$$P(\omega_i) = \frac{1}{N} |X(\omega_i)|^2 \tag{7.1}$$

3. **Step 3:** Normalize the calculated PSD so that it can be viewed as as a Probability Density Function so that all components are between zero to one so that the sum of the normalized power spectral components p_i over the considered frequency range in equal to one.

$$p_i = \frac{P(\omega_i)}{\sum_i P(\omega_i)} \tag{7.2}$$

4. **Step 4:** The Power Spectral Entropy is now calculated using a standard formula for an entropy calculation:

$$SpEn = -\sum_{i=1}^{n} p_i \log p_i \tag{7.3}$$

7.1.2 Approximate Entropy

Heuristically, Approximate entropy (ApEn) measures the logarithmic probability that nearby pattern runs remain close in the next incremental comparison. ApEn quantifies the concept of variable complexity and is an appropriate algorithm for studying its evolution and classifying systems with no need it for complete reconstruction of the dynamics of the system to classify it [35].

The more regular the data of interest is, the smaller the value of the Approximate entropy is -keeping in mind ApEn is non-negative and its minimal bound is 0.

The ApEn algorithm requires to choose -or tune- a pair of hyperparameters

- m: the embedding dimension and is the length of the template (length of the window of the different vector comparisons)
- r: the scaling parameter (or noise filter) for which a superposition of noise much smaller in magnitude does not affect the calculation of the entropy.

We present below the algorithm to compute the Approximate Entropy of a time-series $u(t)$ of length N

Therefore, mathematically, ApEn measures the logarithmic frequency with which blocks of length m that are close together stay together for the next position.

7.1.3 Sample Entropy

According to [74], the bias registered when computing ApEn makes the results suggest more regularity than there is in reality.

By allowing each vector to count itself, all functions $C_m^i(r)$ remain positive, avoiding having to calculate a $log(0)$ because the self-counting allows at least one match and one possible vector. Eliminating this bias by preventing each vector from being counted with itself would make ApEn unstable in many situations, leaving it undetermined if each vector does not find at least one match.

Algorithm 1 ApEn Algorithm

Input $u = \{u(1), u(2), \ldots, u(N)\}$, $(m \geq N, r)$
Output ApEn

$i \leftarrow 1$
while $i \leq N - m + 1$ **do**
$x(i) \leftarrow \{u(i), u(i+1), \ldots, u(i+m-1)\}$
$x(j) \leftarrow \{u(j), u(j+1), \ldots, u(j+m-1)\}$

$d[x(i), x(j)] \leftarrow \max_{k=1,2,\ldots,m}(|u(i+k-1) - u(j+k-1)|)$
$C_i^m(r) \leftarrow$ (number of $j \leq N - m + 1$ such that $d[x(i), x(j)] \leq r)/(N - m + 1)$
$i \leftarrow i + 1$

$\phi^m(r) \leftarrow \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log C_i^m(r)$

$\mathrm{ApEn}(m, r, N)(u) \leftarrow \phi^m(r) - \phi^{m+1}(r)$

To avoid bias implications retained from ApEn, [74] defines Sample Entropy (SaEn), as a statistic which doesn't have self-counting with the justification that entropy being a mean for information production measure, comparing data with itself is irrelevant.

Mathematically, it is the negative value of the logarithm of the conditional probability that two similar sequences of m points remain similar at the next point $m + 1$, counting each vector over all the other vectors except on itself.

In order to mathematically illustrate the theoretical differences between SaEn et ApEn, [35] has presented the two following formulas:

$$\mathrm{SaEn}(m, r, N) = -\log \frac{\sum_{i=1}^{N-m} \sum_{j=1, j\neq i}^{N-m} [\text{ number of times that } d\,[|x_{m+1}(j) - x_{m+1}(i)|] < r]}{\sum_{i=1}^{N-m} \sum_{j=1, j\neq i}^{N-m} [\text{ number of times that } d\,[|x_m(j) - x_m(i)|] < r]} \quad (7.4)$$

$$\mathrm{ApEn}(m, r, N) \simeq -\frac{1}{N-m} \sum_{i=1}^{N-m} \log \frac{\sum_{j=1}^{N-m} [\text{ number of times that } d\,[|x_{m+1}(j) - x_{m+1}(i)|] < r]}{\sum_{j=1}^{N-m} [\text{ number of times that } d\,[|x_m(j) - x_m(i)|] < r]}. \quad (7.5)$$

These expressions allow us to formulate the key differences between both entropies:

- SaEn eliminates self-counting ($j \neq i$) in comparison to ApEn.
- The sum of all template vectors is inside the logarithm in SaEn and outside in ApEn which implies that SaEn considers the complete series and if a template finds a match, then SaEn is already defined, while ApEn needs a match for each template. And given the Jensen's inequality - $log(\sum) > \sum log$ -, that term is greater in SaEn than in ApEn.
- ApEn includes a factor ($\frac{1}{N-m}$) which makes this statistic dependent on the size of the series, whereas SaEn doesn't include it.

7.1.4 Phase Entropy

Second order measures such as auto-correlation function or power spectrum provides a partial explanation of a random process. The principles of power spectrum are extended to higher orders

called **Higher Order Spectra (HOS)** to determine the subtle changes and the shape of a waveform [4].

HOS are the spectral representations of higher order moments of an irregular signal. It detects the **nonlinearity**, **deviations from Gaussianity** and **preserves the phase characteristics** of a signal. Two approaches can be employed in the estimation of HOS: indirect and direct approaches.

In the direct approach, data is divided into segments with equal number of overlapping blocks. The Fast Fourier transform (FFT) is computed for each block and the product of spectral coefficients results in 'raw' HOS. The weighted average of raw HOS gives rise to the third order moment called the bispectrum. It contains two frequencies (f_1 and f_2) and is given by:

$$B\left(f_1, f_2\right) = E\left[X\left(f_1\right) X\left(f_2\right) X\left(f_1 + f_2\right)\right] \tag{7.6}$$

where $X(f)$ is the Fourier transform of a signal. The frequency f can be normalized to lie between 0 and 1 the Nyquist frequency.

On another note, Phase Entropy ($PhEn$) was a novel entropy method For analyzing the complexity of heart rate variability (HRV) and has been proposed as a quantification of the phase space [76] and was formulated based on the above explanations and expressions and is as follows:

$$\begin{gathered}
\mathrm{PhEn} = \sum_{n=0}^{N-1} p\left(y_n\right) \log p\left(y_n\right) \\
\text{where } p\left(y_n\right) = \frac{1}{L} \sum_{f_1, f_2 \in \Omega} l\left(\phi\left(B\left(f_1, f_2\right)\right) \in y_n\right) \\
y_n = \left\{ \varphi \mid -\pi + \frac{2\pi n}{N} \leqslant \varphi < -\pi + \frac{2\pi(n+1)}{N} \right\}, n = 0, 1, \ldots N-1
\end{gathered} \tag{7.7}$$

where N = an integer, L = number of points within Ω, φ = the phase angle of the bispectrum, l = indicator function ($l = 1$ when phase angle is within the y_n range).

7.1.5 Fuzzy Entropy

The Fuzzy entropy is a modification of the Sample Entropy by introducing fuzzily defined exponential functions for comparison of vectors' similarity instead of the $Chebyshev$ [32]:

$$\begin{gathered}
\boldsymbol{X}_m(i) = [x(i), x(i+1), \ldots, x(i+m-1)] - x0(i) \\
D_{ij,m} = \mu\left(d_{ij,m}, r\right) = \exp\left(-\left(d_{ij,m}\right)^n / r\right) \\
\phi_m(n, r) = \frac{1}{N-m} \sum_{i=1}^{N-m} \left(\frac{1}{N-m-1} \sum_{j=1, j \neq i}^{N-m} D_{ij,m} \right) \\
FuzzyEn(m, n, r, N) = \ln \phi_m(n, r) - \ln \phi_{m+1}(n, r)
\end{gathered} \tag{7.8}$$

7.2 Chaos Analysis features

Chaotic analysis is a relatively novel area in the study of physiological signals. Chaotic features of electroencephalogram have been analyzed in various disease states with a direct involvement of the brain such as: epilepsy, Alzheimer's disease, sleep disorders, and depression [36].

The basic theory behind conducting nonlinear analysis on electroencephalograms is that EEG signal is produced fundamentally due to a nonlinear deterministic process between multiple neurons which are highly dynamic by nature [46] and which, at their own level, exhibit nonlinear property [11]. And, given that brain activity is nothing but the sum of interactions of millions of neuronal populations, nonlinear analysis is evaluated to be an apt approach for EEG study. Chaotic analysis, which forms a subset of nonlinear analysis, is therefore well suited for studying EEG signals.

In this optic, various chaotic features have been utilized and tested efficient for quantifying the complexity within the human brain electrical activity and either diagnosing, detecting, locating or distinguish disorders from normal cognitive activity.

7.2.1 Largest Lyapunov Exponent

An essential concept of deterministic chaos is that small differences over time in the system's initial values are prone to diverge to very large differences. The reason behind this behavior is the exponential growth of these differences, which is principally a feedback effect. This is, implicitly, a different approach to one of the fundamental properties of chaos: the existence of strong sensitivity to a change of the initial conditions [15].

> ***" Differences grow in respect to the size they have reached in the previous moment of time. The larger they are, the faster they grow."***

The Lyapunov exponent indicates the speed with which two initially close dynamics:

– diverge -if the Lyapunov exponent is positive.

– converge -if it is negative- in phase space.

If the difference of the initial values -initial distance between neighboring points- is d_0, the mean divergence between neighboring trajectories at time t can be estimated as:

$$d(t) = d_0 e^{\lambda t} \tag{7.9}$$

Where λ is the Lyapunov exponent.

The Lyapunov exponent, therefore, indicates how fast a complex system is prone to diverge towards deterministic chaos and allows to perceive certain conclusions about the predictability of a system.

And, when the system is defined in more dimensions, there may be a whole spectrum of Lyapunov exponents and the largest value of them is the so-called Largest Laypunov Exponent (LLE).

Multiple algorithms and numerical approaches have been proposed and implemented to estimate the Largest Lyapunov Explonents with distinct approaches and formulations and have proven reliable; for instance: Wolf [88], Rosenstein [77] and Kantz [51].

In our case study, we decide to go with the Rosenstein method given its high computational speed [36] and simpleness of implementation. However, we should note that this method does not provide the value of the LLE dorectly, rather, it yields the following function:

$$y(i, \Delta t) = \frac{1}{\Delta t} \langle \ln d_k(i) \rangle , \quad d_k(i) = \min_{x_k} \|x_k - x'_k\| \tag{7.10}$$

where x_k is a given point within our time-series signal of interest and x'_k denotes its neighbor.

This algorithm is based on the relation between $d_j(i)$ and the Lyapunov Exponents: $d_j(i) \simeq e^{(\lambda_1(i\Delta t)}$. The largest Lyapunov exponent is then, computed by estimating the inclination of the most linear part of the function.

7.3 Optimization of hyper-parameters of Complexity Analysis features

Nonlinear time-series analysis is used to understand, characterize and predict dynamical information about human physiological systems. This is based on the concept of state-space reconstruction, which allows one to reconstruct the full dynamics of a nonlinear system from a single time series (a signal). We now, firsthand define basic concepts in phase space reconstruction and complex -nonlinear dynamical- systems [64].

Definition 4 *The phase space of time series is created in order to analyze the chaotic behavior of a time-series. It is a space that consists of state variables and their derivatives. Often, two or three dimensional spaces are used and a point is determined by Cartesian coordinates . The customization consisting of Cartesian coordinates x, y, z is inadequate when dealing with a dynamical system, because the latter doesn't stay fixed to place points ie, the point position depends on the time in motion.*

Definition 5 *The state of a dynamical system is a vector of variables and which is represented by a point in a phase space. This state aims to provide a complete description of the system (in this case, health) at some point in time.*

Definition 6 *The state space represents the set of all possible states and the way in which the system evolves over time (e.g., change in person's health state over time) can be referred to as trajectory.*

Definition 7 *The attractor is the common subset of state space different trajectories may evolve or converge towards after a sufficiently long time. The presence and behavior of attractors gives intuition about the underlying dynamical systems. This attractor can be a fixed-point where all trajectories converge to a single equilibrium point or it can be periodic where trajectories follow a regular path (i.e., cyclic path).*

The purpose of predicting of function's attractor and topological properties is to re-establish the phase space of the system. One standard method for state-space reconstruction is **time-delay embedding** -also known as delay-coordinate embedding. It aims to identify the state of a system at a particular point in time by searching the past history of observations for similar states, and by studying how these similar states evolve, in turn predict the future course of the time series [64].

In conclusion, the purpose of time-delay embedding is to reconstruct the state and dynamics of an unknown dynamical system from measurements or observations of that system taken over time.

7.3.1 Embedding dimension m

The embedding dimension parameter sets the number of vectors to be compared. Practically speaking, it dictates the number of additional signals successively delayed by τ, and therefore its sets how many axes will be included in the reconstruction space of the system's dynamics. It decides how much of the system's history is shown.
Dimensionality must be sufficiently high to generate relevant information and create a rich history of states over time, but also low enough to be easily understandable.

The optimal value for the embedding dimension m is computed using the **False Nearest Neighbors (FNN)**. The false nearest neighbors method is the most popular tool for the selection of the minimal embedding dimension.

False Nearest Neighbors (FNN)

This method is based on the assumption that two points that are near to each other in the sufficient embedding dimension should remain close as the dimension increases. However, if the embedding dimension is too small, then the points that are in reality far apart could seem to be neighbors (as a consequence of projecting into a space of smaller dimension). The various modifications of the method apply geometrical reasoning: one increases ED until the reconstructed image is unfolded. The method checks the neighbors in increasing embedding dimensions until it finds only a negligible number of false neighbors when going from dimension m to $m+1$. This m is chosen as the lowest embedding dimension, which is presumed to give reconstruction without self-intersections [52].

7.3.2 Time-delay τ

It's a certain time that sets basic delay to generate the respective axes of the reconstruction: $x(t), \quad x(t-\tau), \quad x(t-2\tau)\ldots$

E.g., if $\tau = 1$, the state $x(t)$ would be plotted against its prior self $x(t-1)$.

If τ is too small, constructed signals are too much alike and if too large, the reconstructed trajectory will show connections between states very far in the past and to those far in the future with no actual relationship, which might make the reconstruction extremely complex.

The optimal value of time-delay τ is set as being the **local minimum of the mutual information** $I(x(t), x(t-\tau))$ between the original time series $x(t)$ and the time series $x(t-\tau)$ shifted by τ. Since mutual information is computed for a times series and a time-shifted version of the same time series, this is called the auto-mutual information or average mutual information (AMI) [85].

7.3.3 Tolerance threshold r

As it names indicates, it sets the tolerance for accepting similar patterns. Its optimal value is obtained using the **Maximum Approximate Entropy**, given it has been found that this approach allows a correct assessment of the complexity of the signal [63].

Conclusion

With an overview of the features that captured our interest in this research axis, we are confident in the legitimacy of combining this approach to the one already in our hands -SCSA. With the

addition of the optimization of the hyperparameters, we are proposing an autonomic approach that do not hard-code the values of these parameters but rather, compute beforehand their optimal values and generate to their best, these features.

Chapter 8

Feature Engineering for epileptic EEG signal Characterization

Introduction

As we explained and presented the mathematical reasoning behind the choice of our multiple features in the previous chapter, we now propose in this one, a clear and succinct explanation of the workflow leading to the final result of the feature extraction phase including pre-processing milestones crucial to hope and expect any reliable performance following the feature generation. One additional major section concerns the optimization of the time-embedding hyper-parameters.

8.1 Normalization of the time-series

The normalization of time-series is conceptualized as changing the values of numeric columns in the dataset to a **common scale, without distorting** differences in the ranges of values.
Normalization, instead of standardization, is used when the distribution of the signal of interest is unknown/ not Gaussian.
We also remind ourselves that it is necessary to have time-series exclusively positive > 0 given the constraints when applying SCSA. It has been proven, also, that normalization of initial time-series data before the computing of the features enhances the representativeness.

The **min-max normalization** is performed following the sample mathematical operation:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{8.1}$$

8.2 Windowing and data-Segmentation

It is necessary to segment our time-series data. In any practical and research applications, EEG signals are split into small segments of length W, and with an overlap specified along the way to avoid any loss of information due to brutal shift from one window to another. The choice of these parameters are crucial for the representativeness and the relevance of the extracted dataset of features. Indeed, the choice of the window size when processing event-related (ER) time-series data can alter the results of the classification model given it is proper to the requirements of the problematic we aim to take on and solve.
Smaller windows are more fit when tackling a faster activity detection issue with reduced resources available, while larger windows are found when dealing recognition of complex activities with a requirement for greater resources. It's therefore a trade-off between detection time and accuracy.

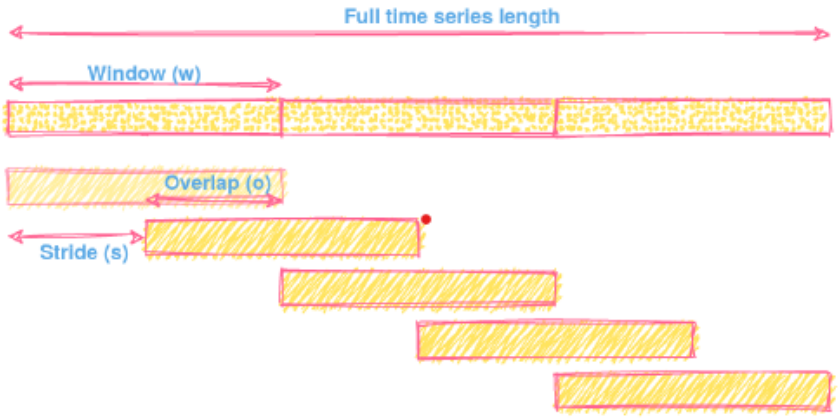


Figure 8.1: EEG time-series data segmentation

Segments that are too long, are time-consuming and resource-expensive but also induce the classification model to fail because it misses the onset seizure. A too small window is advised against so that the model is still handed with enough context to make informed decisions.
These size parameters can be computed through a tuning of hyper-parameters but that would show to be over resource-expensive and time-consuming process. Therefore, we opt for an approach based on literature review accompanied with a hard-on tuning.

We set our choice on a window size of $4s$ with an overlap of $2s$.

8.3 Time-series signal filtering

As we developed in details in Section 3.4, specific frequency sub-bands that hold relevant information were defined, decades back in neuroscientific literature and are thoroughly used in more recent research work for epilepsy [68], [84]. Based on this understandable tendency for "Quantified-EEG", we chose to perform filtering on each of our segments to obtain time-series data in each of the relevant scalp-EEG frequency bands: δ, θ, α, β.

We design a band-pass **Butterworth** filter of $8th$ order with cutoff frequencies according to which sub-bands we are retrieving:

Frequency sub-band	Frequency interval
Theta θ	$[0.5 - 4]Hz$
Delta δ	$[4 - 8]Hz$
Alpha α	$[8 - 12]Hz$
Beta β	$[12 - 30]Hz$

Table 8.1: Frequency interval for each brain-wave sub-band

Therefore, we obtain four time-series for one segment of data and we compute our set of features for each time-series and we obtain a dataset with the following columns (features):

$$with \quad \mathrm{x} = \{\theta, \delta, \alpha, \beta\} \tag{8.2}$$

Feature_x	Abbreviation
Number of eigenvalues	N_h
A set of squared-root eigenvalues	κ_i i=$\overline{1:10}$
Semi-classical hyper-parameter	h
First invariant	$I_1(h)$
Second invariant	$I_2(h)$
Ratio of two consecutive κ	κ_rat
Ratio of two consecutive N_h	N_h_rat
Approximate Entropy	$ApEn$
Sample Entropy	$SaEn$
Spectral Entropy	$SpEn$
Phase Entropy	$PhEn$
Largest Lyapunov Exponent	LLE
Fuzzy Entropy	$FzEn$

Table 8.2: Summary of the extracted features for each brain-wave sub-band

8.4 Designed approach to the tuning of the hyperparameters

For each EEG time-series, we proceed with applying filters according to the four main brain frequency sub-bands. We run the algorithms explained above on each window of the four filtered versions of each window of the 100 epileptic seizure data segments of the Bonn University dataset as the approach in [16] inspired us to. We count down the frequency of occurrence of all the values for each hyper-parameter and obtain the following results and the Figures 8.2 8.3 8.4 8.5 illustrate the system's dynamics reconstruction and the variations of the functions we base our optimization on.

$$pars_\theta = [\text{'}Dimension\text{'} : 5, \quad \text{'}Delay\text{'} : 6, \quad \text{'}Tolerance\text{'} : 0.16] \tag{8.3}$$

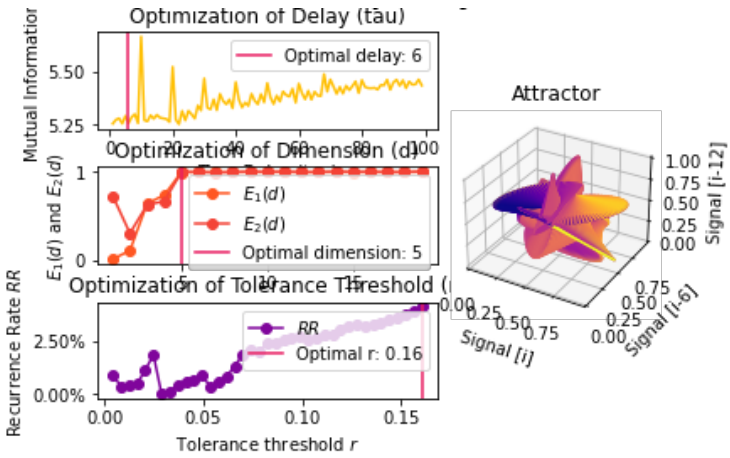


Figure 8.2: Reconstruction of the attractor of the dynamical system for $\theta-$filtered signals

$$pars_\delta = [\text{'}Dimension\text{'} : 5, \quad \text{'}Delay\text{'} : 4, \quad \text{'}Tolerance\text{'} : 0.16] \tag{8.4}$$

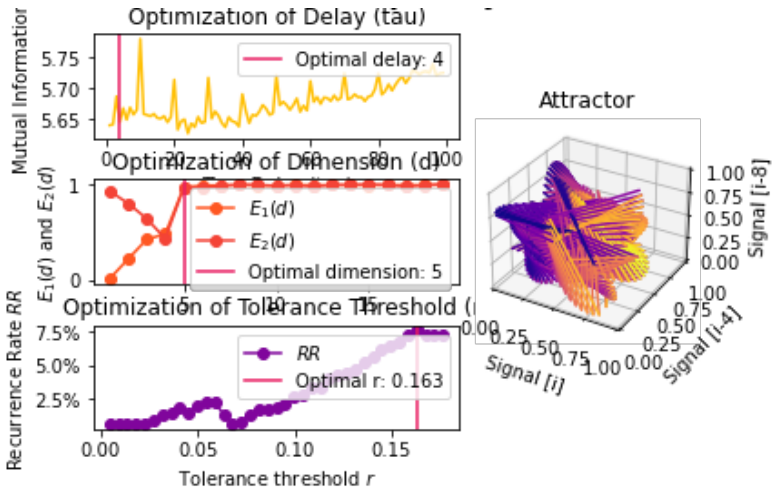


Figure 8.3: Reconstruction of the attractor of the dynamical system for $\delta-$filtered signals

$$pars_{\alpha} = ['Dimension' : 5, \quad 'Delay' : 4, \quad 'Tolerance' : 0.15] \tag{8.5}$$

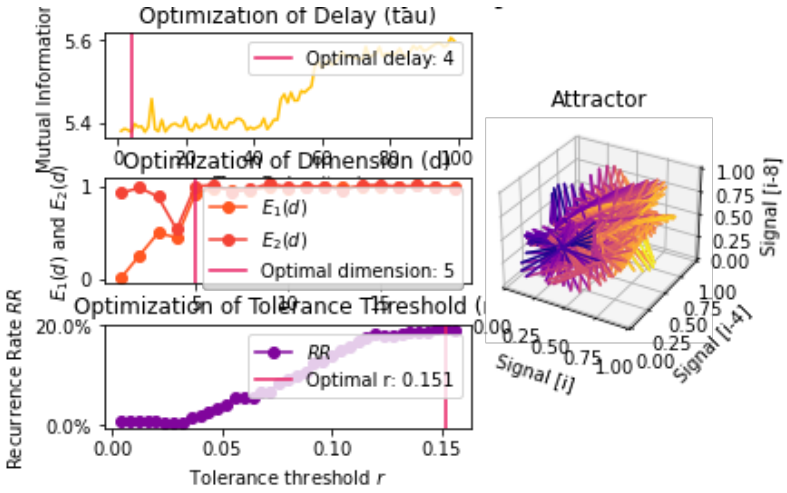


Figure 8.4: Reconstruction of the attractor of the dynamical system for $\alpha-$filtered signals

$$pars_{\beta} = ['Dimension' : 6, \quad 'Delay' : 4, \quad 'Tolerance' : 0.12] \tag{8.6}$$

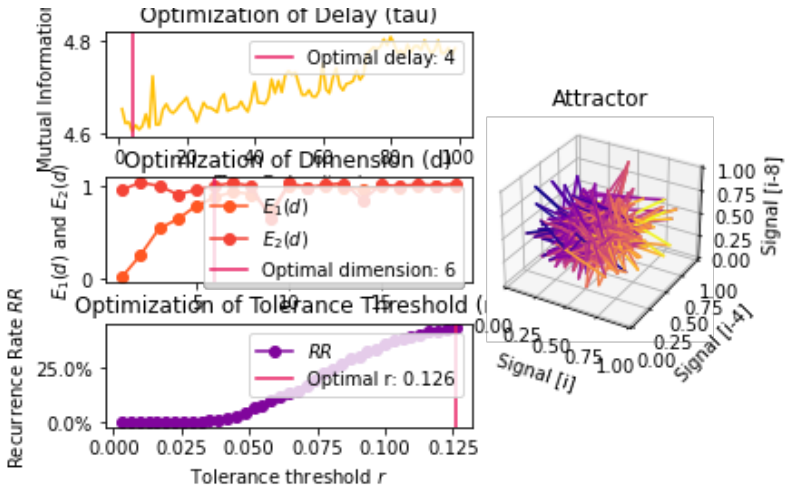


Figure 8.5: Reconstruction of the attractor of the dynamical system for $\beta-$filtered signals

While these statistical features quantify and identify specific, known behaviors within the signal of interest, the features extracted through the Semi-Classical Signal Analysis study are for the first time introduced in the present work, to assess the problematic of epileptic seizure detection.

8.5 Features correlation assessment and study

Entropies have been recurrently used for epileptic seizure detection [5] [31] [50] [87], and as we explained in details in the previous chapter, each expresses a certain characteristic or specificity within the signal processed: its regularity, chaocity, complexity ... etc.

It is therefore, in a first time, interesting to study the correlation between these features that result from two distinct mathematical backgrounds that are for the first time employed together.

From a second point of view, SCSA features have proven to be capable of acting as biomarkers. Hence, it is a facet worth exploiting in features resulting from an approach proven to be of good performance on pulse-shaped, biophysical signals.

In statistics, correlation is a method of assessing a possible two-way linear association, connection, or any form of relationship, link or correspondence between two continuous variables. Correlation is measured by a statistic called the correlation coefficient, which represents the strength of the putative linear association between the variables in question that takes a value in the range 1 to +1 [69]. The stronger the relation, the closer the correlation coefficient is to ± 1.

8.5.1 Spearman's rank correlation

In contrast with the standard definition of correlation as assessing the linear association between the two considered variables, the Spearman rank correlation coefficient evaluates the relationship between the rank of these observations -said X and Y-, and detects the existence of monotonic relationships (increasing or decreasing), regardless of or their precise form (linear, exponential, power, etc).

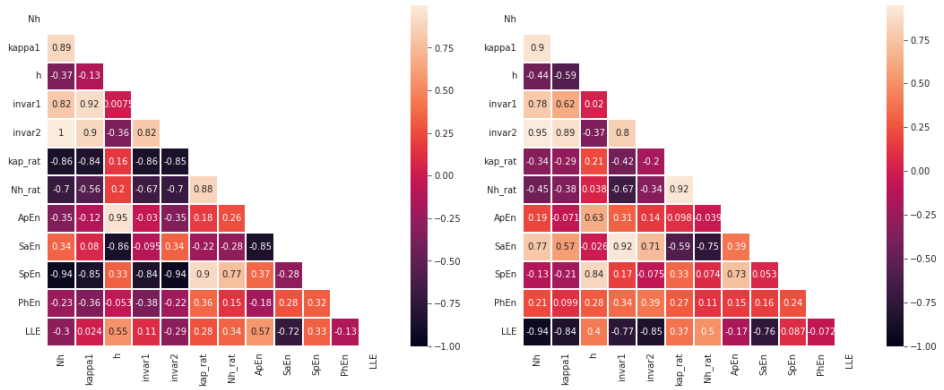
It's appropriate to opt for this coefficient when the variables are skewed or ordinal given it's proven robust when the latters show extreme values or outliers. For a correlation between variables x and y, the formula for calculating the Spearman's correlation coefficient is given by:

$$r_s = 1 - \frac{6 \sum d_i^2}{n\left(n^2 - 1\right)} \tag{8.7}$$

where $d_i = \mathrm{R}\left(X_i\right) - \mathrm{R}\left(Y_i\right)$ is the difference between the two ranks of each observation, n is the number of observations.

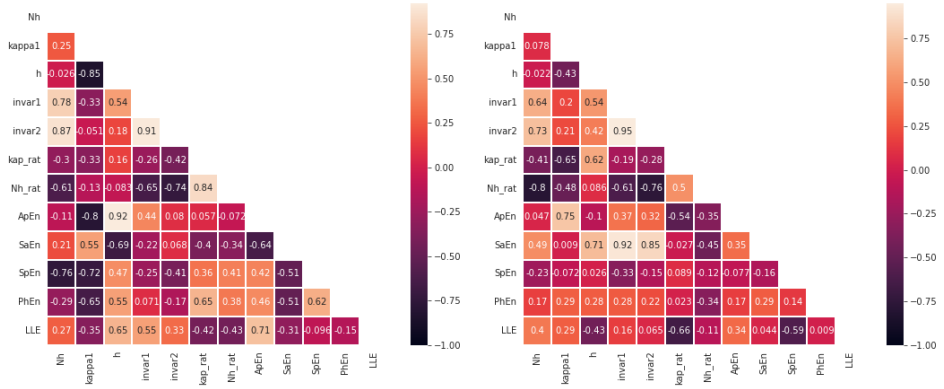
For the sake of this case study, we compute the Spearman's rank correlation coefficients heatmaps between the set of features combining both statistical and SCSA features.

For each frequency sub-band of interest, we obtain a correlation heatmap as follows:



(a) correlation heatmap for θ sub-band

(b) correlation heatmap for δ sub-band



(c) correlation heatmap for α sub-band

(d) correlation heatmap for β sub-band

Figure 8.6: Heatmaps of the Spearman's correlation between features

A very relevant observation to begin our reasoning with is correlations quantified through this approach can either be justified through a mathematical demonstration or by a correlation that cannot be further backed through mathematical formulas.

The most prominent correlation we were capable of assessing in this chapter was the -practically- linear relation between the **Sample Entropy** and the **First Invariant** within the β frequency sub-band, and that for a Spearman's correlation rate of $r_s = 0.92$ as we can observe in 8.6. The said linear correlation is illustrated in Fig. 8.7.

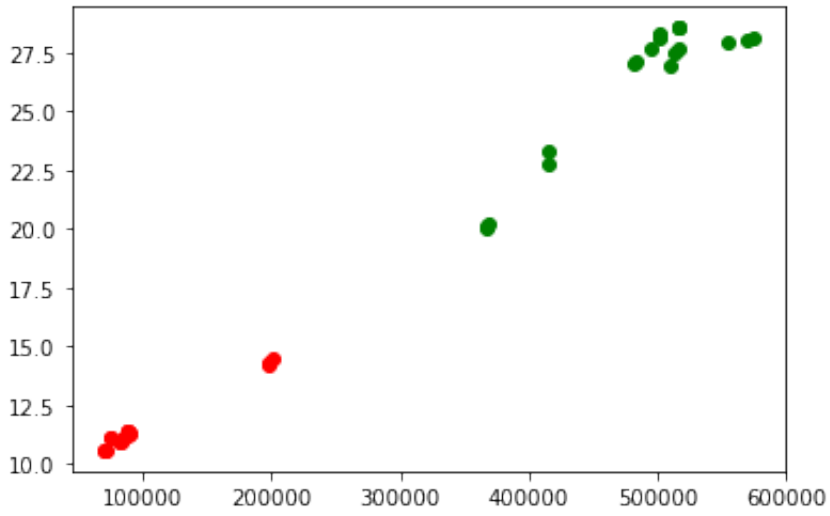


Figure 8.7: $SaEn = f(invar_1)$

We are digging further into more theoretical articles and research in order to demonstrate and back this correlation with more mathematical proof. WE have been able to identify promising axis in this quest given the first invariant is none but a ***Rietz sum***. We have been able to as a first step find track as in [42], [75] who tackle principally the concept of Kolmogorov-Sinai entropy -which is considered as the forefather of Approximate and Sample Entropy- in the quantum domain and until the ***semi-classical*** limit of it. This falls directly within our scope of work. However, as this could be, by itself, an independent axis of research, we will suffice for now in this report with these results and preliminaries.

8.6 Features ranking and selection

Selection of features is crucial to build an efficient workflow and is the most important process for creating a most efficient machine learning model. Adding redundant variables reduces the generalization capability of the model and may also reduce the overall accuracy of a classifier. Furthermore, adding more and more variables to a model increases the overall complexity of the model and introduces what we call the **Curse of Dimensionality** which is defined by Bellman in [17] as phenomenons that occur when attempting to describe, model or analyze data in Euclidean spaces more complex -with far more dimensions- than the initial one the phenomenons at the origin of the data occur in. This proved to largely impact the performance of Machine Learning models which can fall exponentially along with the need for more data samples growing exponentially with the number of features utilized.

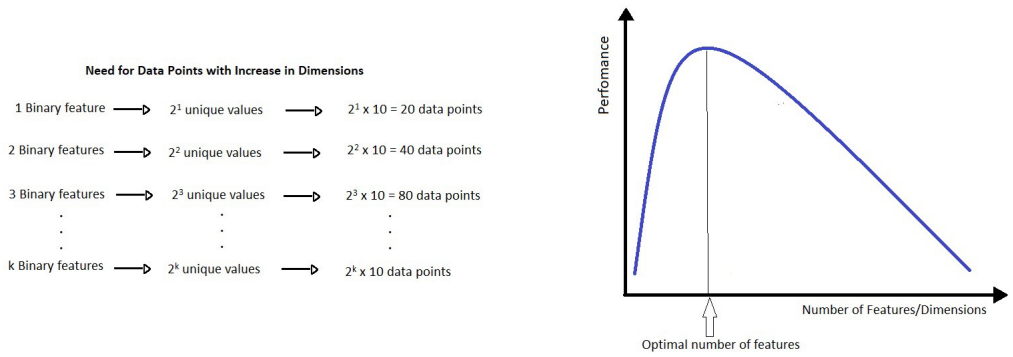


Figure 8.8: The Curse of Dimensionality

Feature selection is a process of selecting required features that have the most impact on the output variable; it means only the features which are independent and are highly related to the output variable.

Multiple methods have been used depending on the type of features and type of labels in the database considered. In Fig. 8.10, we present the **filter methods** which concern **Supervised Classification**.

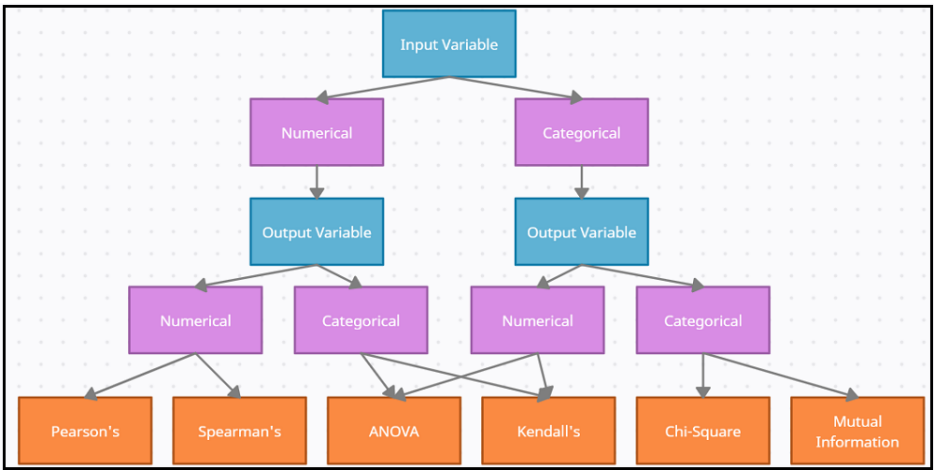


Figure 8.9: Filter methods for features selection

In our case, We deal with **numerical** input variables and **categorical** output. Therefore, we are going to utilize the **ANOVA** method to rank the features in a decreasing order of relevance and importance based on the **f-value** or **f-ratio**.

8.6.1 ANalysis Of VAriance (ANOVA)

Analysis of variance (ANOVA) is a statistical technique used to check if the means of two or more groups are significantly different from each other.

F-ratio shows how much of the total variation comes from the variation between groups and how much comes from the variation within groups and is defined as follows:

$$F = \frac{\text{Variation between groups}}{\text{Variation within groups}} \tag{8.8}$$

If the largest contribution to the variation comes from:

– the variation **between groups**: it is more likely that the mean of groups are different.

– the variation **within groups**, then we can conclude the elements in a group are different rather than entire groups.

it is more likely the elements in a group are different rather than entire groups.

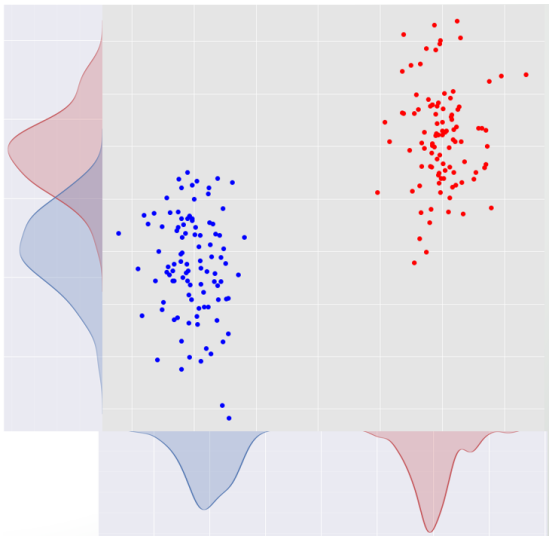


Figure 8.10: Illustration of the impact of variance between features on distribution of samples [66]

The mathematical formulation of the $F - ratio$ is as follows:

$$F = \frac{\text{MSB}}{\text{MSW}} = \frac{\text{SSB}/df_b}{\text{SSW}/df_w} \tag{8.9}$$

where

$$\text{SSB} = \sum_{j=1}^{k} \left(\bar{X}_j - \bar{X}\right)^2 \quad and \quad \text{SSW} = \sum_{j=1}^{k} \sum_{j=1}^{l} \left(X - \bar{X}_j\right)^2$$

$$df_b = \mathbf{n} - \mathbf{k} \quad and \quad df_w = \text{k} - 1$$

In our study case, we evaluate the relationship between y as $class \quad labels$ and computes for each feature of the dataset, a F-value. A large score suggests.

A small value suggests the feature is independent from the label y and a larger one would suggest the feature is non-randomly related to y, and therefore likely to provide important information.

We perform **ANOVA** on the dataset of features extracted in **the first section of Part II: Features Extraction**. We exclusively keep the 25 features with the highest importance rates (highest f-ratios) and the results, through the quantification of the importance and relevance of each feature towards the occurrence of the labels, are:

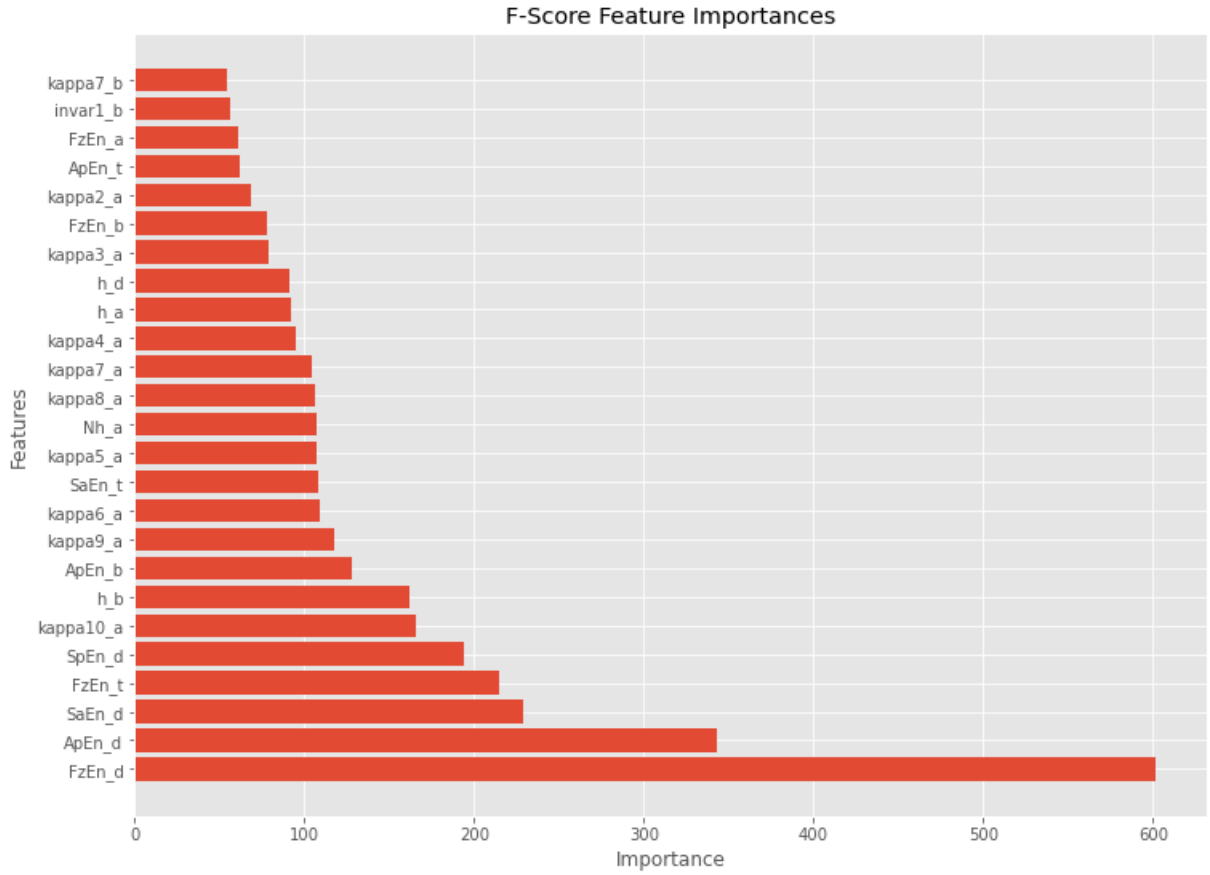


Figure 8.11: Ranking of the 25 most important features according to their f-value

It is possible to deduct from the f-value, what we call the **p-value**.

8.6.2 p-value and null-hypothesis

Definition 8 *A p-value is used in hypothesis testing to help support or reject the* ***null hypothesis****. It is the evidence against a null hypothesis.*

Definition 9 *The null hypothesis is a general statement that there is no relationship between two measured phenomena. Testing the null hypothesis —and thus concluding that there are or are no ground for believing that there is a relationship between two phenomenons.*

The smaller the p-value, the stronger the evidence that we should reject the null hypothesis (the less chance results could be random) and a large p-value means your results have a high probability of being completely random and not caused to anything related to the experiment at the origin of the dataset. Therefore, the smaller the p-value, the more important ("significant") your results.
We can observe a similitude and correlation in the logic we employ in both f-value and p-value. However, the f-distribution table provides critical values, not P-Values notice. Indeed, the f-distribution table simply gives us an **f-critical** value to compare our f-statistic to but does not directly provides us with a p-value.

When we specify our f-statistic as we do in the beginning as from feature ranking using ANOVA, and we of course have our two degrees of freedom specified beforehand, we would simply use what

we call **an f-distribution Calculator**.

Luckily, we are in no need to implement neither an f-distribution calculator neither the table of correspondence. Routines available on Python libraries, ready for use, summarize the work on this point and a simple choice of the number of features we decide on retrieving is enough to obtain the results illustrated in Fig. 8.11.

Conclusion

At the epilogue of this chapter, we can be first of all, satisfied with our intuition to focus on the optimization of the hyper-parameters prior to diving in feature extraction and trying to tune classifiers to obtain gratifying rates. An important addition we explain in this chapter is the use of the frequency diversity of brain-wave signals by introducing the notion of θ, δ, α and β in the larger picture and computing our features for each frequency sub-band given how by definition, each of them reflects a certain type of activity in the brain.

A second section of this chapter, and its results more importantly, emphasizes, on another note, the importance feature correlation study has and should hold more often. It is undeniably crucial to be capable of understanding our features, especially when we claim the introduction of novels features, novel analysis methods. It is also of out-most importance to choose the right criterion for feature ranking later ahead. Moreover, with a method relying in major part on the relevance of the features extracted and their understanding, this step has to be embedded in workflows. As in our case, it helps us be more confident in the introduction of the SCSA in this case. First of all, given according to the Figures 8.6, most SCSA features do not correlate to the point of disqualification. On the contrary, the correlation rates we note gives more credit to the SCSA features, because they are translated in some way to the ability of these features to capture, to a certain extent, the same type of information other well-established features have been defined to be able to.

Last but not least, feature ranking and selection is more than necessary. It may be the most important phase of the work. Much interest have grown towards it during the last two decades, and legitimately so. If not for a feature ranking and selection method adapted to the formulation of the problem we have, an unavoidable gap would form preventing us from passing through into classification and evaluation of the method as a whole. Indeed, we passed from 88 features to 25 features. In terms of computational cost, this is considered enormous.

Part III

Data Augmentation

Chapter 9

Problem Formulation

Introduction

Defining imbalanced learning is crucial to lay the foundation for the following development. In this chapter, we define imbalanced learning and proceed with providing a brief overview and definition of the multiple alternatives to tackle study case of inherently imbalanced datasets. And while the latter are at the core of different problematics as prediction, regression or clustering tasks, we are concerned with imbalanced learning as a **Classification** matter. In this Chapter, we provide a brief introduction to the problem formulation, research methods, and challenges and alternatives in this field.

Definition 10 *Imbalanced learning is defined as the learning process for data representation and information extraction with severe data distribution skews to develop effective decision boundaries to support the decision-making process.*

Definition 11 *Imbalanced classifications pose a challenge for predictive modeling given most the machine learning algorithms used for classification were designed around the assumption of an equal number of samples for each class. This results in models that have poor predictive performance, specifically for the minority class. This is a problem because typically, the minority class is more important and at the heart of the study we are doing and therefore the problem is more sensitive to classification errors for the minority class than the majority class but which is not reflected in the construction of the cost function and the classifier.*

Now, in the state-of-the-art of imbalanced learning, there are three main axes to work with imbalance learning as presented in Fig 9.1.

9.1 Data-level processing:

It is the most popular and dominant approach to remedy an imbalanced classification problem and so, by changing the composition of the training dataset. These techniques are generally referred to as sampling, or **resampling** techniques. They owe their popularity to their simplicity of implementation and therefore the possibility of using all the classifiers fit for balanced datatsets in our study case.

This resampling technique ignores the underlying cause of the imbalance in the problem domain and is only performed on the training dataset used by the algorithm to learn a model, and not performed on the holdout test or validation dataset. The intent is not the removal of the class bias

from the model fit but to evaluate the resulting model on data that is real and representative of the target problem domain [27].

– **Undersampling:** We remove random samples from the majority class until we reach a balance with a ratio of 1 : 1 between the population of the two classes.

  Yet, undersampling techniques cause a considerable loss of information with the removal of samples and thus, discard potentially relevant information.

– **Oversampling:** We duplicate random samples from the minority class to reach again the 1:1 ratio.

  This approach can cause overfitting with repetitive data samples because it is data replication within a class that is already very poor.

9.2 Algorithm-level:

Algorithm-level solutions are seen as an alternative approach to data preprocessing methods. Instead of focusing on modifying the training set in order to combat class skew, this approach aims at modifying the classifier learning procedure itself [39].

These methods concentrate on modifying existing learners to alleviate their bias towards majority class, and this requires a good insight into the modified learning algorithm and a precise identification of reasons for its failure in mining skewed distributions. Therefore, we must analyze the splitting mechanism.

One of the main tracks in this axis of methods is altering the regularization term in the formulation of the optimization problem.

Indeed, the assumption made by standard classifiers that all of training instances are of equal importance for the training process, is not valid in case of imbalanced data. To accommodate this, an weighted modification is proposed. The underlying assumption behind this approach is to assign different weights or penalties to each training instance. This may be achieved by re-writing the soft margin optimization task as:

$$\min\left(\frac{1}{2}w\cdot w+\sum_{i=1}^{l}C_i\xi_i\right)$$

$$\text{subject to } \underset{i=1,\cdots,l}{\forall}\ \underset{\xi_i\geq 0}{\forall}\ \underset{C_i\geq 0}{\forall} y_i\left(w\cdot\Phi\left(x_i\right)+b\right)\geq 1-C_i\xi_i, \quad (9.1)$$

The penalty term $\sum_{i=1}^{l}\xi_i$ depicts how many training instances may lie on the wrong side of the decision boundary during the classifier learning procedure and C_i is the penalty/weight associated with given training instance x_i. Values of weights may be given depending on the imbalance ratio between classes or individual instance complexity factors. C_i will be at higher values when the error is made on classifying on a sample from the minority class.

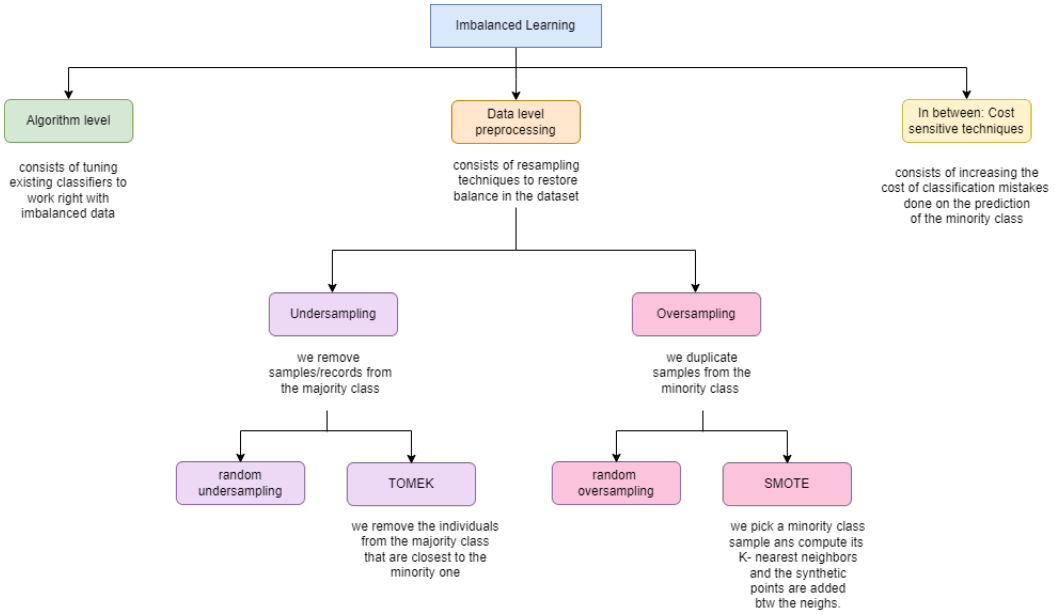


Figure 9.1: Imbalanced learning

9.3 In-between (cost-sensitive techniques):

As it names indicates, these methods are not exclusively within any of the previous categories. Yet, it is more similar and largely fall within the algorithm-level techniques.

A summary of these methods and a few others of their sub-types are clearly ordered and briefly explained in Fig 9.1.

While the alternative of being able to restore balance in our dataset seems to be the most interesting axis to carry on, oversampling techniques remain insufficient to satisfy the necessary condition of distinct samples to avoid overfitting. In 2014, the newest alternative to tackle imbalanced datasets through data augmentation was introduced to the research and data-science world: **Generative Adversarial Networks (GAN)**.

Chapter 10

Generative Adversarial Networks (GAN)

Introduction

Yann LeCun described Generative Adversarial Networks as ***"the most interesting idea in the last 10 years in Machine Learning"***. Such a judgment gives us perspective in the present and upcoming impact of this new approach in the Artificial Intelligence field. We start this chapter with an introducing section to the main concepts in the growing GAN field, followed by an exhaustive section on the training of 2D-GAN and why we opt for this type of architecture whilst unrolling the latter. Later on, we present the input and data representation we choose to feed the Generative Networks with: spectrograms. Finally, it is crucial to evaluate the perfomance of this section of our work as sepectrograms generated, time-series retrieved and features extracted to ensure no over or under-fitting from the model.

10.1 Definitions and Introduction

Definition 12 *Generative modeling is an unsupervised learning task in machine learning that involves automatically discovering and learning the regularities or patterns in input data in such a way that the model can be used to generate or output new examples that plausibly could have been drawn from the original dataset.*

Definition 13 *GANs are deep-learning-based models that clever enough to train a generative model by framing the problem* ***as a supervised learning problem****. It consists of two sub neural networks: the generator and the discriminator models which are trained together in a zero-sum game to mimic the distribution of a chosen dataset and augment this same data population.*

Definition 14 *The generator (G) is the part in the model that differentiate Generative Networks from the popular discriminative ones. It estimates the probability distribution of the real samples and learns to generate realistic fake samples by transforming latent/random noise data. The random samples are fed as input.*

Definition 15 *The discriminator (D) is the part of the Network in charge of distinguishing between real (original) and fake (generated) data. It estimates the probability that a given sample comes from real data or from the generator*

The graphical model in . 10.1 shows naively the difference in input and output of each of the NN of the GAN while the synoptic schematic illustrated in 10.2 shows a general block diagram of the GAN architecture.

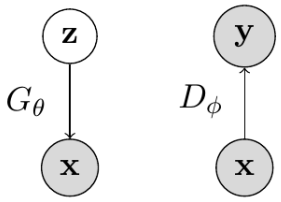


Figure 10.1: Graphical model of D and G

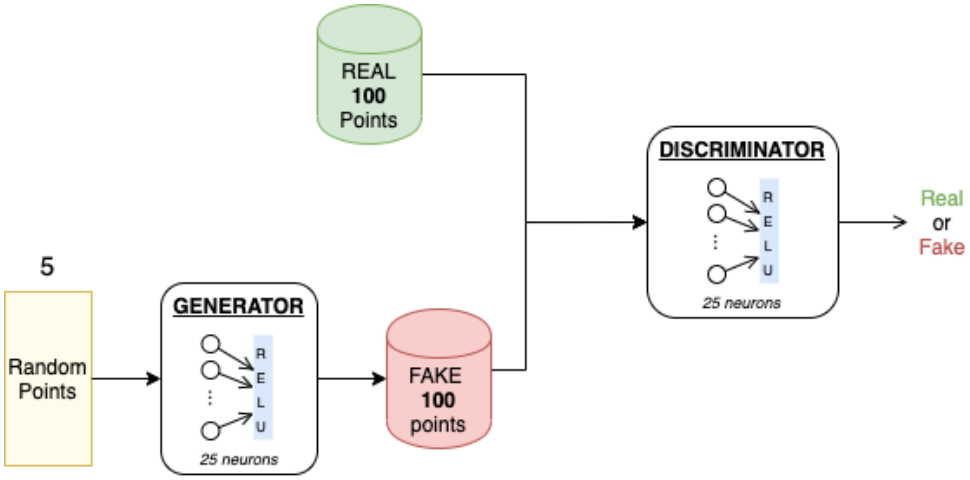


Figure 10.2: Synoptic schematic of Generative Adversarial Networks [65]

10.2 Organization of the training of the GAN

10.2.1 Minimax optimization problem

Simply put, GANs as proposed by Goodfellow [43], is similar to the idea of **game theory** where two players compete against each other with only one winner -our two players in this case are the generator and the discriminator.
As we explained earlier the roles of each, the end goal for the generator is to continuously keep on generating and improving the random noisy samples such that they match the real ones the discriminator is trained on, and therefore, causes the discriminator to label the artificial data from the generator as real. The opposing nature of the dynamics between these two networks justifies the use of adversarial and the mathematical perception of this dynamics as a ***minimax*** problem where the generator minimizes a two-sample test objective ($p_{data} = p_\theta$) and the discriminator maximizes the objective ($p_{data} \neq p_\theta$). Both the generator's and discriminator's training procedure can be formalized mathematically using a ***minimax*** decision rule/optimization problem [45]:

$$\min_\theta \max_\phi V\left(G_\theta, D_\phi\right) = \mathbb{E}_{\mathbf{x}\sim \mathbf{p}_{\text{data}}}\left[\log D_\phi(\mathbf{x})\right] + \mathbb{E}_{\mathbf{z}\sim p(\mathbf{z})}\left[\log\left(1 - D_\phi\left(G_\theta(\mathbf{z})\right)\right)\right] \tag{10.1}$$

Where G, D, x and z are respectively: generator parameters, discriminator parameters, real sample and generated sample (from noise).

Let's unpack this expression. We know that the discriminator is maximizing the function $D(x)$ function with respect to its parameters ϕ, where given a fixed generator G_θ it is performing binary classification: it assigns probability 1 to data points from the training set $x \sim p_{data}$, and assigns probability 0 to generated samples $x \sim p_G$.

The function $D(x)$ gives a probability as to whether the sample belongs to a real or generated data distribution. As an end goal, both the discriminator and generator are trying to maximize the $\log(D(x))$ function (the discriminator's performance on labeling), whilst minimizing the $\log(1 - D(G(z))$. The generator minimizes this latter term for a fixed discriminator D_ϕ.

So intuitively, the generator tries to fool the discriminator to the best of its ability by generating samples that look indistinguishable from p_{data} and maximize the probability that D misclassifies them as real ones. All the while, D adapts to minimize its error on classifying generated and real samples, and hence, not get fooled by the generator.

As a conclusion to this section, we can easily conclude that the direct link between the discriminator and the generator during the training of G to product samples as similar as possible to the original data distribution. Indeed, GANs are a practically a **closed-loop** system.

10.2.2 Feedback and Closed-loops in GANs architectures

Indeed, with the previous explanations and definitions taken into consideration, we can deduce that when the discriminator is good at detecting fake samples, the generator is updated more, and when the discriminator model is relatively poor or confused when detecting fake samples, the generator model is updated less.

Indeed, we can imagine that at the beginning of the training of the generator, the discriminator will classify the generated samples as not real (class 0) or a low probability of being real (0.3 or 0.5). The **back-propagation** process used to update the model weights will see this as a **large error** and will update the model weights **(i.e. only the weights in the generator)** to account for this error and correct it, in turn making the generator better at generating plausible fake samples. We can now understand how the performance of the discriminator modulates the update of the

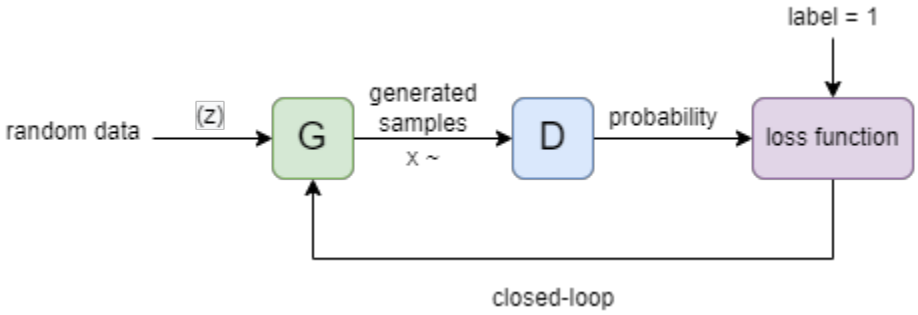


Figure 10.3: Modeling of the GAN with its feed-back loop

weights of the GAN.

There is a crucial point to take into consideration now and it is the need to freeze the discriminator (meaning its weights) and mark all its layers as untrainable. The reason behind this measure is to avoid they are updated or over-train on fake samples so that their weights don't update based on the discriminator's performance on fake samples [24].

We know that the discriminator is only concerned with distinguishing between real and fake examples; therefore, the discriminator model can be trained in a standalone manner on examples of each. The generator model is only concerned with the discriminator's performance on fake examples.

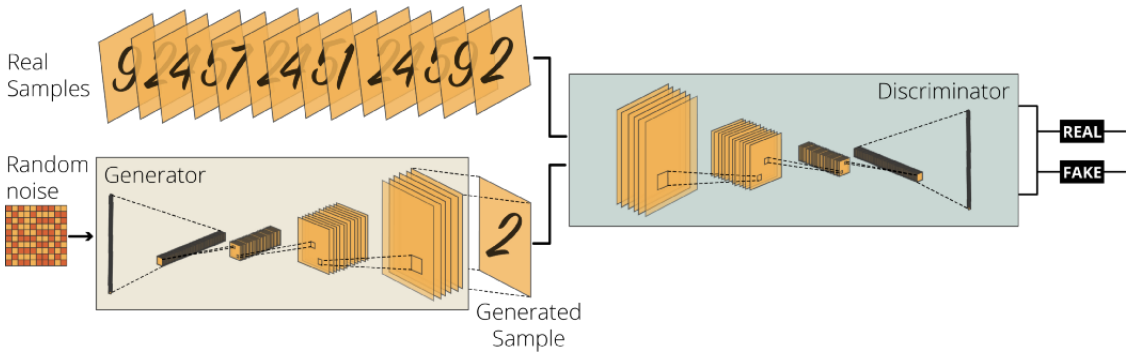


Figure 10.4: 2D-GAN with MNIST data [89]

Therefore, practically, the simplest approach is to create a new model that subsumes or encapsulates the generator and discriminator models. Specifically, a new GAN model can be defined that stacks the generator and discriminator such that the generator receives as input random points in the latent space, generates samples that are fed into the discriminator model directly, classified, and the output of this larger model can be used to update the model weights of the generator. It's just a logical third model that uses the already-defined layers and weights from the standalone generator and discriminator models.

When training the generator via this subsumed GAN model, there is one more important change. We want the discriminator to think that the samples output by the generator are real, not fake. Therefore, when the generator is trained as part of the GAN model and the generated fake samples are fed into the discriminator for classification, we will mark the generated samples as real (class 1).

10.3 2-Dimensional GANs

One of the most popular and democratized applications of GANs in research but also in entertainment are in image generation and two-dimensional data augmentation. In our case study, this approach is particularly interesting given it is instilled in the research field, and on another note, we can avoid dealing with the temporal dimension coherence in time-series signals. As images, we will feed the spectrograms of our epileptic data segments as real data input.

10.3.1 Spectrograms

A spectrogram is a visualization of the frequency spectrum -frequency range- of a signal. It's a tool extending the one-dimensional representation of a spectrum to a second dimension which can be time. In coordinate $(y - axis)$, the frequency is represented, while time is represented in abscissa axis $(x - axis)$. The color $(z - axis)$ represents the magnitude of the signal for a specific frequency at a specific time stamp.

The spectrogram is generated as a matrix output. Therefore, once normalized with respect to itself, is a good candidate to be fed into a Convolutional Neural Network (CNN) as a 2D-signal.

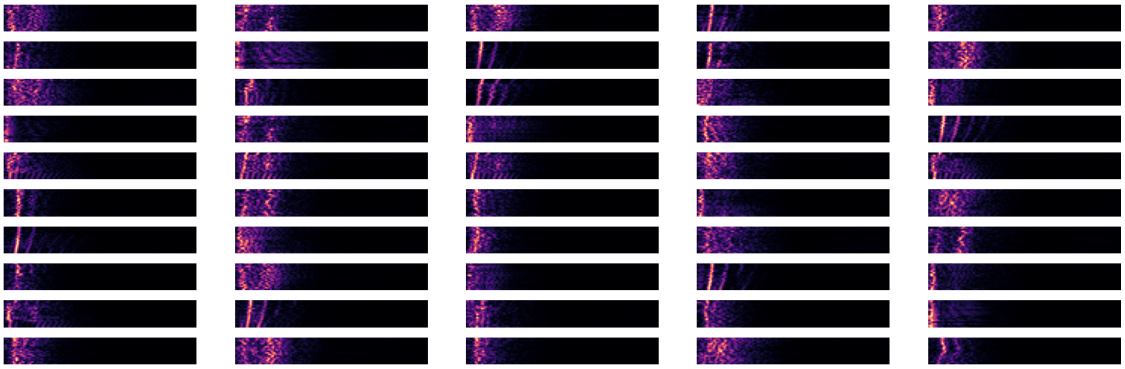


Figure 10.5: 50 spectrograms generated from 50 original segments of epileptic-seizure data

Practically, when we plot and fit this matrix information to the f and t scale bound to the duration of the time-series segment and to the sampling frequency. In our case, the duration of each segment is of $23.6s$ with a sampling frequency $173.61Hz$. What we obtain therefore is not used as the input of our CNN, but is illustrated in the example of Fig 10.6 for the sake of representation:

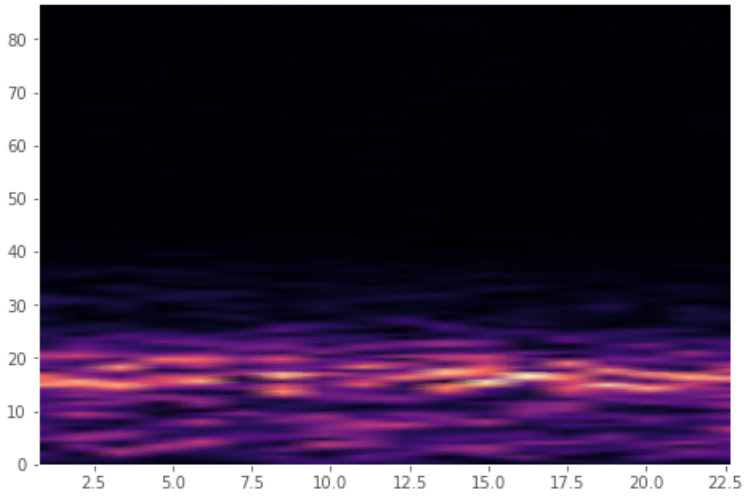


Figure 10.6: Spectrogram of segment $n°10$ for epileptic seizure data

In our case, we opt for mode ***magnitude*** which translates as retrieving the magnitude of the signal based on the Short-Term Fast-Fourier **STFT**. We also choose to apply a ***Hanning window***.

10.3.2 Architecture of the 2D-GAN model

In this section, we will go deeper in presenting the architecture designed for each of the Generator and the Discriminator of our 2D-GAN model.

Generator model

The architecture of the Generator is presented in Fig. 10.7 and inspired from [25], with its layers and dimensions of the input and output of each. It is the part of the model responsible for creating new, fake but plausible spectrograms: it operates by taking a point from the latent space as input and outputting a generated spectrogram.

Definition 16 *The **latent space** is an arbitrarily defined vector space of **Gaussian-distributed values**, e.g. 100 dimensions and by randomly drawing points from this space and providing them to the generator model during training, the latter will assign meaning to these latent points and at the end of training, the latent vector space represents a compressed representation of the output space that only the generator knows how to turn into plausible spectrograms.*

In our case, we opt for a latent vector of 100 element vector of Gaussian random numbers that the generator model transform to an outputs as a two-dimensional spectrogram matrix considered as an image of 129×18 pixels -*i.e.* 2322 values- with the pixel values in $[0, 1]$ - the spectrograms are normalized with respect to themselves.

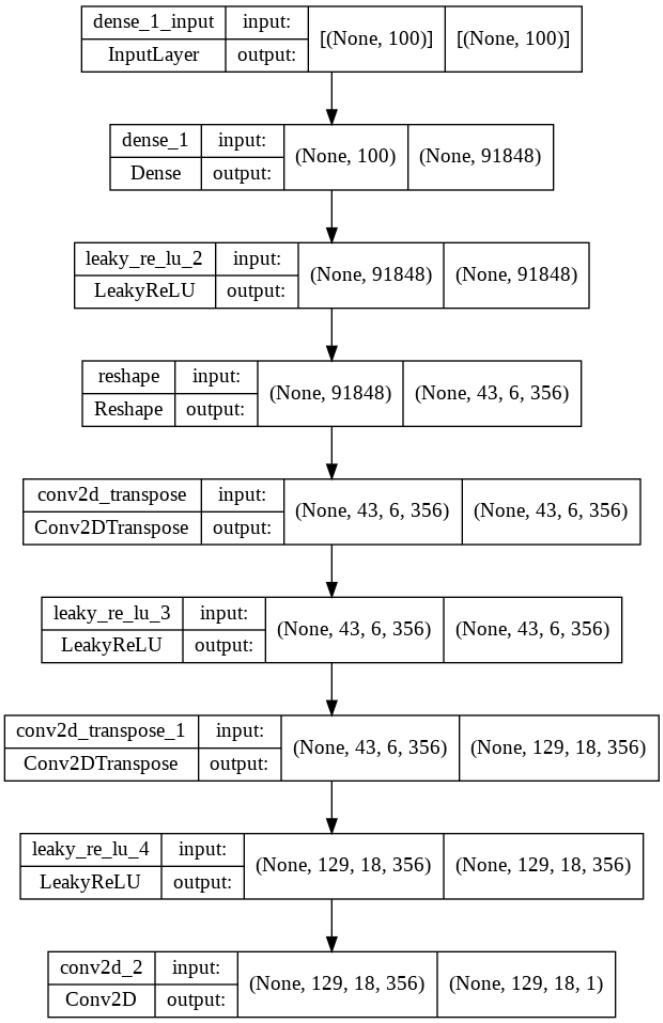


Figure 10.7: Architecure of the Generator model

There are multiple ways to achieve this but one approach has proven effective: deep convolutional generative adversarial networks. It involves two main elements.
The first is a **Dense layer** as the first hidden layer that has enough nodes to represent a low-resolution version of the output image. Specifically, our original image would be the size of 129×18 or 2322 nodes, and an image one third the size (one nineth the area) would be 43×6 or 294 nodes.

We don't just want one low-resolution version of the image; we want ***many parallel versions*** or interpretations of the input. This is a pattern in convolutional neural networks where we have many **parallel filters** resulting in multiple parallel activation maps, called feature maps. We naturally also want the same thing in reverse: many parallel versions of our output with different

learned features that can be collapsed in the output layer into a final image. Therefore, the first hidden layer, the Dense one, needs enough nodes for multiple low-resolution versions of our output image, such as 356.

Discriminator model

As for the Discriminator, we decide on the architecure summarized in Fig. 10.8. Indeed, the model must take a sample image from our dataset as input and output a classification prediction as to whether the model estimates the sample is real or fake: it's a scalar that reflects how likely the model estimates that the sample is real (or fake). This is a **binary classification** problem . The discriminator model we build is directly inspired from [25] and has **two convolutional** layers with 64 filters each, a small kernel size of 3, and larger than normal stride of 2. The model has no pooling layers and a single node in the output layer with the **sigmoid activation function** to predict whether the input sample is real or fake. The model is trained to minimize the binary cross entropy loss function, appropriate for binary classification.

We will use in defining the discriminator model, the **LeakyReLU** instead of ReLU, using Dropout, and using the **Adam version of stochastic gradient descent** with a learning rate of 0.0002 and a momentum of 0.5.

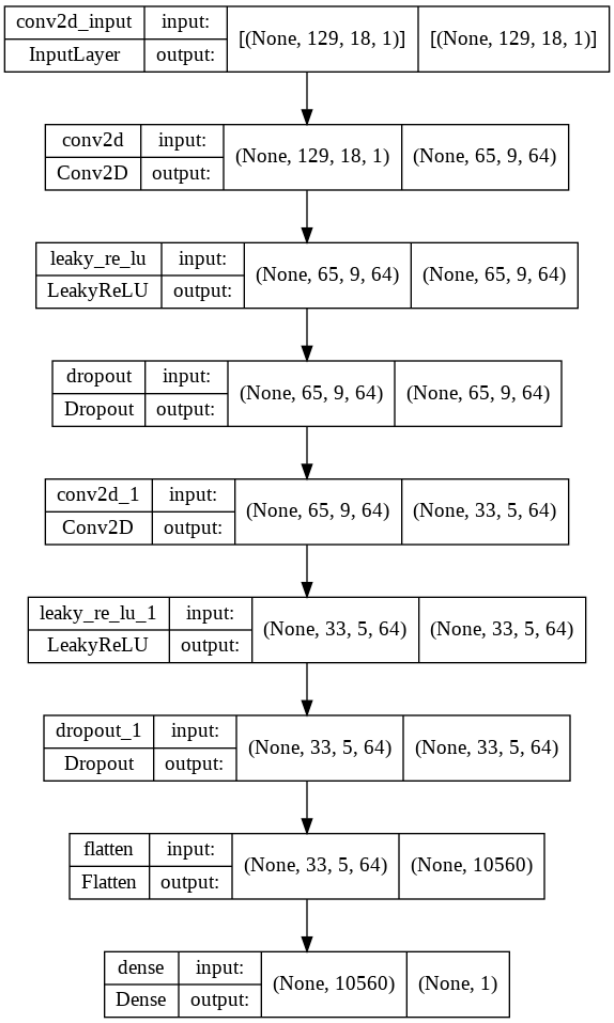


Figure 10.8: Architecture of the Discriminator model

10.3.3 Training the Generator Model

As we explained in the section 10.2.2, the weights in the generator model are updated based on the performance of the discriminator model.

When the discriminator is good at detecting fake samples, the generator is updated more, and when the discriminator model is confused concerning fake samples, the generator model is updated less. This defines the zero-sum or adversarial relationship between the two models.

Specifically, a new logical GAN model is defined by stacking the generator and discriminator where only the discriminator model can be trained in a standalone manner.

Given the generator is only concerned with the discriminator's performance on fake examples, we will mark all of the discriminator's layers as untrainable using **keras.API**.

For that, the GAN model uses the same parameters and settings as the discriminator in terms of loss-function, solver, learning rate and momentum.

The discriminator model was then marked as not trainable, added to the GAN model, and compiled. In this model, the weights of the discriminator cannot be changed when the GAN is updated. This change in the trainable property -which impacts the model after it's compiled- does not impact the training of standalone discriminator model [25].

Instead, what is required is that we first update the discriminator model with real and fake samples, then update the generator via the composite model.

There are a few things to note in this model training function:

- the number of epochs we train our model for.
- the number of batches within an epoch is defined by how many times the batch size divides into the training dataset.

The discriminator model is updated **once per batch** by combining **one half** a batch of fake and real examples into a single batch which will be faster over a long run, especially when training on GPU hardware.

Finally, we report the loss over each batch given how critical it is. The reason for this is that a crash in the discriminator loss indicates that the generator model has started generating unexploitable examples that the discriminator can easily discriminate.

We choose to opt for a total of 160 epochs and a batch-size of 25 which results in a total of $160 \times \frac{100}{25} = 640$ batches -100 stands for the number of samples we consider in our dataset which is 100 epileptic seizure spectrograms- for which we monitor the discriminator's and generator's losses in parallel and we obtain the following graph of the training losses in Fig. 10.9:

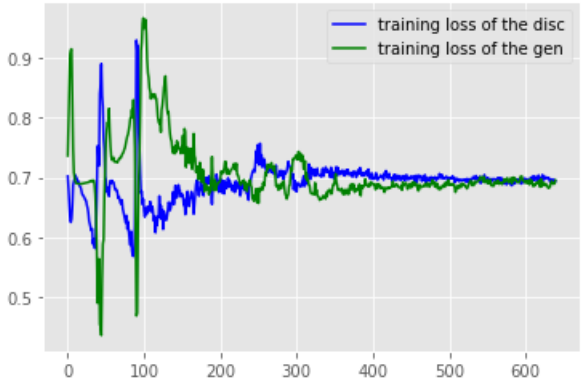


Figure 10.9: Training loss monitoring over batches

The discriminator loss while the generator loss is relatively less critical. From this graph, we can observe an obvious trait which is the convergence and stabilization of the training loss for both the Generator and Discriminator after a certain number of batches which backs the idea of minimax optimization and Game theory GANs were built upon.

10.3.4 Performance evaluation of the GAN model

Practically, there are no objective ways to evaluate the performance of a GAN model in literature because we cannot calculate an objective error score for generated images.

Instead, images must be subjectively evaluated for quality by a human operator. This means that we cannot know when to stop training without looking at examples of generated images [25]. In turn, the adversarial nature of the training process means that the generator is changing after every batch, meaning that once "good enough" images can be generated, the subjective quality of the images may then begin to vary, not necessarily towards improvement, but eventually towards possible degradation.

There are three ways to handle this complex training situation:

– Periodically evaluate the classification accuracy of the discriminator on real and fake images.

– Periodically generate many images and save them to file for subjective review.

– Periodically save the generator model.

We choose to perform all three of these measures together every 10 training epochs. The result will be the accuracy evaluation graph for each of the discriminator and generator as illustrated in Fig. 10.10 a saved generator model for which we have a way of subjectively assessing the quality of its output and objectively knowing how well the discriminator was fooled at the time the model was saved.

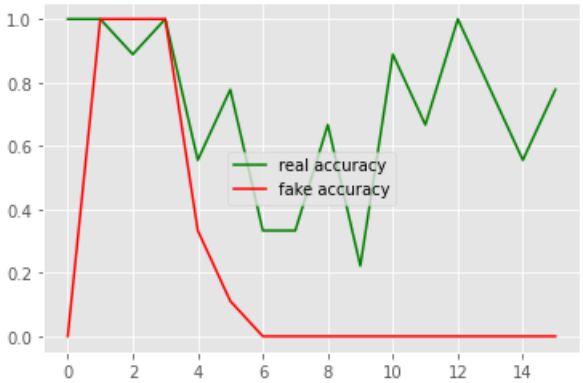


Figure 10.10: Caption

Training the GAN over many epochs, such as 160 epochs, will result in 16 snapshots of the model that can be inspected and from which specific outputs and models can be picked later. We evaluate the performance of the discriminator by retrieving a sample of real spectrogram images, as well as generating the same number of fake spectrograms images with the generator model, then evaluating the classification accuracy of the discriminator model on each sample and reporting these scores.

10.3.5 Inversion of spectrograms: from spectrogram to time-series signal

The output of the GAN are as explained previously, matrix of the spectrogram information of the segments as illustrated in Fig. 10.11.

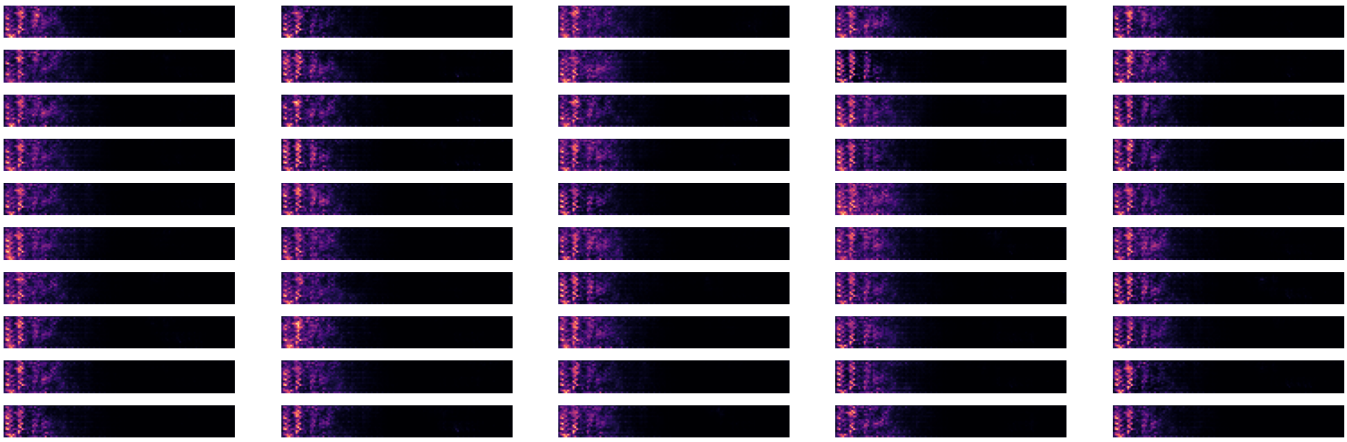


Figure 10.11: 50 spectrograms generated as the output of the 2D-GAN architecture

What we aim to retrieve from this output now is the corresponding 1D-signal. This proves to be quiet challenging up to this day because as we specified earlier, the mode ***"magnitude"*** used for spectrogram generation gives the absolute value of the amplitude of the signal through the frequency and time variations. Therefore, the phase information of the signal is discarded which translates in the practical reality that two signals not exactly identical as a 1D-signal in the temporal domain can have the exact same spectrogram with a phase difference.

We first present the algorithm which allows us to retrieve the invert the magnitude spectrogram and retrieve the 1D-signal 2:

Algorithm 2 Spectrogram Inversion Algorithm

Input S_{xx}, t, f, f_s
Output x(t)

$i \leftarrow 0$
$j \leftarrow 0$
$tic \leftarrow 0$
$phases \leftarrow random.uniform(0, 2\pi, length(f))$
$length \leftarrow length(t) \times f_s$
while $i \leq length(t)$ **do**
  $duration \leftarrow t[i] - t[i-1]$
  $start \leftarrow t[i-1] \times f_s$
  $tics \leftarrow duration \times f_s$
  while $tic \in tics$ **do**
    $index \leftarrow start + tic$
    while $j \leq length(f)$ **do**
      $magnitude \leftarrow S_{xx}[i][j]$
      $x[index] \leftarrow x[index] + magnitude \times sin(\frac{2\pi}{f_s} f[j] \times index + phases[j])$
      $j \leftarrow j + 1$

    $tic \leftarrow tics[]$
  $i \leftarrow i + 1$

Therefore, this could be considered relatively questionable. Yet, we believe that a large part of the time-frequency information being preserved and this, added to the "random" and stochastic nature of epileptic-seizure data, this approach is still a viable option and in order to defend this statement, we conclude on conducting a comparison study between epileptic seizure original and generated data segments in both time-frequency and temporal domains, but also in term of the features extracted from both data.

10.3.6 Comparison between the distribution of original and generated spectrograms

We compute the average distribution of the original epileptic data with the objective of comparing it with the average distribution of the resulting generated data. We observe the following results in terms of spectrograms:

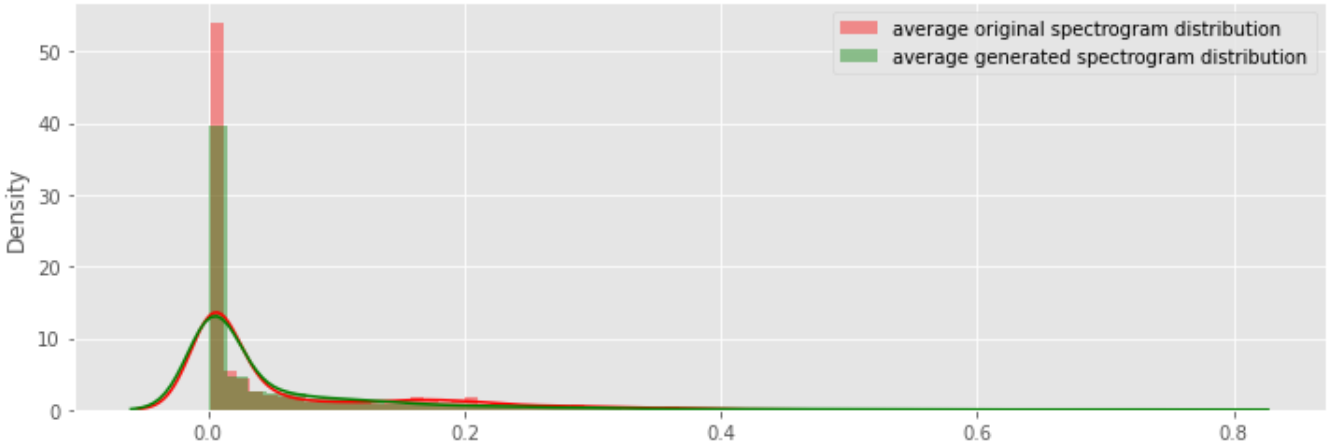


Figure 10.12: Spectrogram data distribution

We can clearly deduct the correspondence between the two distributions that result in practical matching of the distribution.

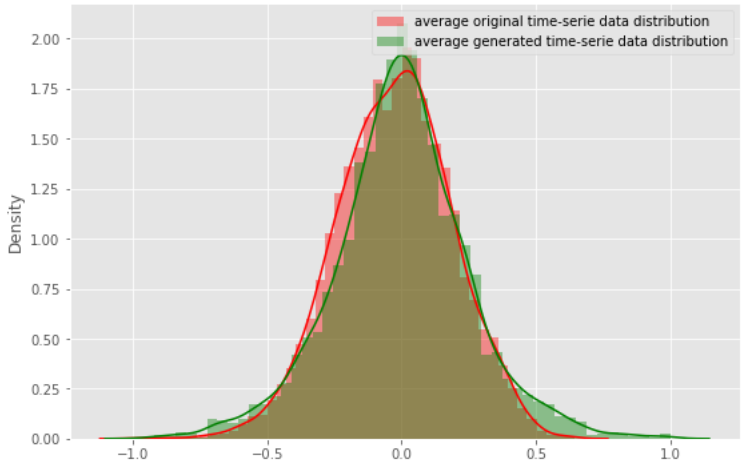


Figure 10.13: Time-series data distribution

Moreover, after the inversion of the spectrograms, the comparison of the 1D-signal time-series distribution produces results which allow us more confidence and reliabitlity in our generated data as we can observe in Fig. 10.13 10.13.

10.3.7 Features extracted for both original and generated data segments

With the same set of hyperparameters and conditions, we compute the features defined earlier for both set of segments and compare the distribution obtained for each feature between both time-series group.

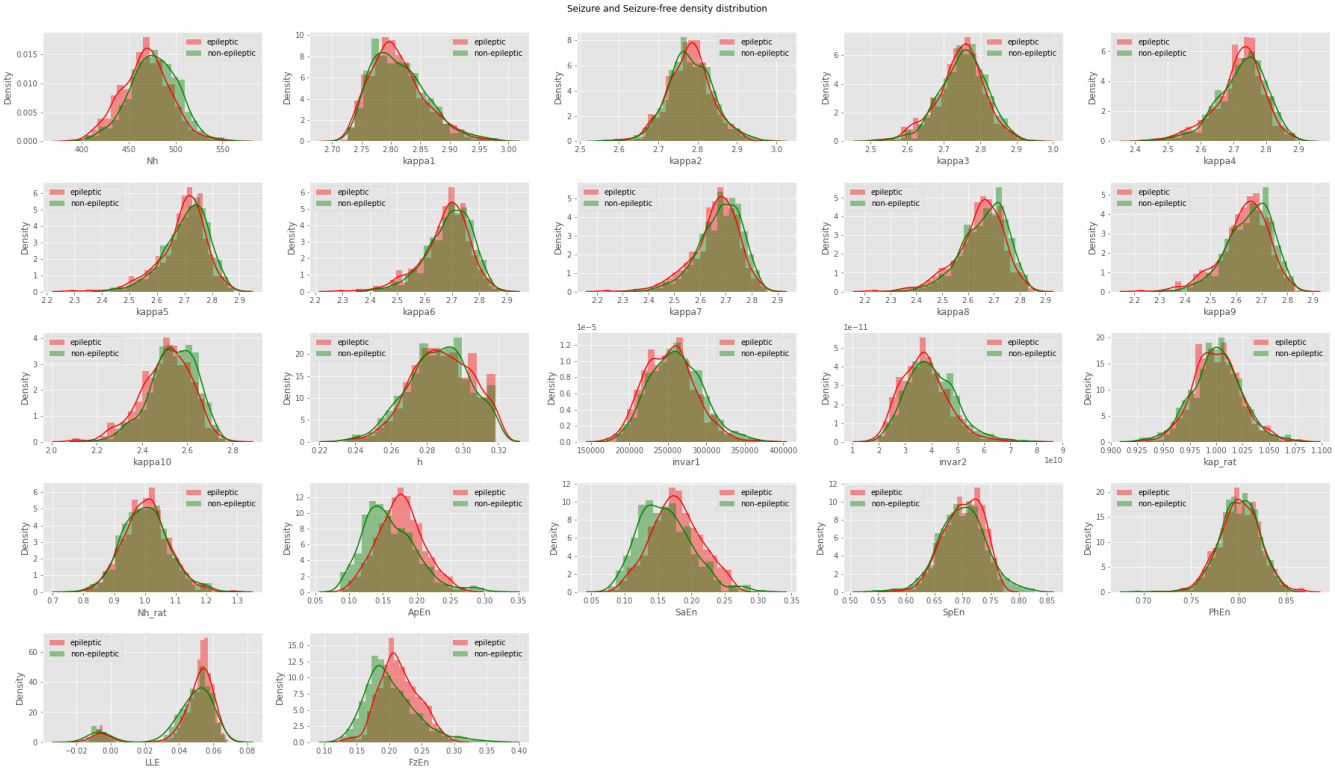


Figure 10.14: Distribution of the features for original and generated epileptic segments

From Fig. 10.14 we can asses how identical the distributions of the features for both data are, and therefore, we can confidently make the point of the correspondence and validity of the proposed approach for data augmentation of epileptic seizure time-series.

Conclusion

From this chapter, and more specifically the comparison section, we can confidently carry on using the generated EEG segments of epileptic-seizure data in the rest of our workflow. Given we consider 400 segments of seizure-free EEG data, we generated 300 additional epileptic-seizure segments. We can now rely on this balanced dataset for the upcoming and last phase: classification.

Part IV

Classification

Chapter 11

Machine Learning for Epileptic Seizure Detection

Definitions we established in Chapter 2 enables us to introduce the problem definition, where the seizure detection problem can be treated as a **binary classification** problem.

Definition 17 *Classification is the process of predicting the target associated with each sample in the data. Classification task is done usually using Machine Learning (ML) algorithms through a supervised learning approach.*

Now, the step to take in the designed framework for our study case is the training and evaluation of multiple classification algorithms based on the generated, ranked and selected features along the previous phases.

11.1 Presentation of ML classifiers

In this section, we ought to present briefly the theory behind each of the standard and very well-known classification algorithms we plan on using for the discrimination performance evaluation of our framework.

One of the most interesting things with out choice of programming language and libraries is that each and any classifier can be built and ready to work in a total of four (4) steps:

1. Import the model we are planning to use, knowing the overwhelming majority of classifiers are implemented as Python classes and ready to use by importing the **sklearn** library.
2. Create an instance of the Classifier model.
3. Train the model on the training data and store the information learned from the training dataset.
4. Predict the labels of the test dataset.

We plan on relying on very standard implementations of well-established classifiers in the field of binary classification: **Logistic Regression, Gaussian Naive Bayes, Decision Tree, Random Forest, Quadratic Discriminant Analysis**.

11.1.1 Logistic Regression

Logistic regression is a supervised learning algorithm designed for classification and given it returns a continuous value, a necessary step to set up a **threshold** [73]. This parameter decides on the conversion of the predicted probability into a class label. A value above that threshold indicates one class while the one below indicates the other.

Additionally, LR is also highly sensitive to imbalanced data as it tries to fit the line by minimizing the error (distance between the line and the actual value).

The cost function is also referred to as loss and is given by the equation:

$$cost(h_\theta(x), y(actual)) = \begin{cases} -\log(h_\theta(x)) & if \quad y = 1 \\ -\log(1 - h_\theta(x)) & if \quad y = 0 \end{cases} \tag{11.1}$$

11.1.2 Gaussian Naive Bayes

Naive Bayes is a classification algorithm for binary problems. It is called Naive or Idiot Bayes because the calculation of the probabilities are simplified to make their calculation tractable. Rather than attempting to calculate the values of each attribute value $P(d_1, d_2, d_3|h)$, they are assumed to be conditionally independent given the target value $P(d_1|h) \times P(d_2|h)$ are based on the **Bayes'** theorem 11.2:

$$P(h \mid d) = \frac{P(d \mid h) \times P(h)}{P(d)} \tag{11.2}$$

Where:

- $P(h/d)$: the posterior probability: probability of hypothesis h given the data d; ***the probability we are interested in computing***
- $P(d/h)$: the probability of data d given that the hypothesis h was true.
- $P(h)$: the prior probability of h: the probability of hypothesis h being true (regardless of the data).
- $P(d)$: the probability of the data (regardless of the hypothesis).

After calculating the posterior probability for a number of different hypotheses, we select the hypothesis with the highest probability. This is the maximum probable hypothesis and is formally called the ***maximum a posteriori (MAP) hypothesis*** [26].

Naive Bayes can be extended by assuming a Gaussian distribution and is called **Gaussian Naive Bayes** where we only need to estimate the mean and the standard deviation from the training data, for each class to summarize the distribution.

11.1.3 Decision Tree

The Decision tree algorithm is a supervised learning algorithm wherein the data points are continuously split according to certain parameters and the problem the algorithm is trying to solve [72]. Every decision tree includes a **root node, some branches, and leaf nodes**.

Decision trees follow a top-down approach meaning the root node is always at the top of the structure while the outcomes are represented by the tree leaves. Decision trees are built using a

heuristic, recursive partitioning. Each node following the root node is split into several nodes.

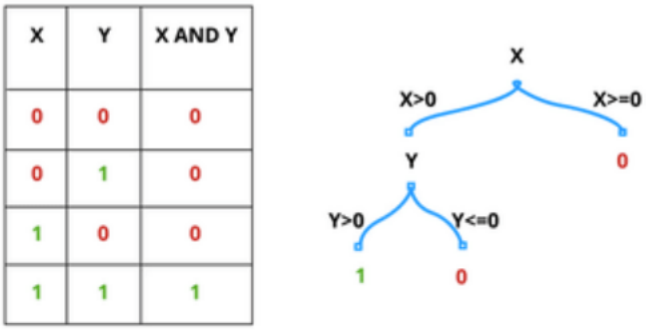


Figure 11.1: Decision Tree splitting [72]

The idea is to use a tree to partition the data space until the data is sufficiently homogeneous.

An important factor in this algorithm is **Entropy**. It measures the uncertainty of the dataset and describes the degree of randomness of a particular node. The higher the entropy, the higher the randomness in the dataset. While building a decision tree, a lower entropy is preferred and its formula is:

$$\text{Entropy} \ = \sum_{i=1}^{C} -p_i * \log_2 (p_i) \tag{11.3}$$

There are multiple parameters to tune to avoid overfitting which can be managed by deciding an optimal value for the **maximum depth** parameter that specifies the number of splits. Other parameters can be used to control the splitting: **the maximum number of features considered, the minimal number of samples by split and the minimum number of leaves**.

11.1.4 Random Forest

Random forest is a Supervised Machine Learning algorithm and builds decision trees on different samples and takes their majority vote for classification and average in case of regression [71]. We can enumerate the steps involved in running the random forest algorithm and illustrate them in Fig. 11.2

Step 1: n, a number of random records are taken from the dataset having k number of records.

Step 2: Individual decision trees are constructed for each sample.

Step 3: Each decision tree will generate an output.

Step 4: Final output is considered based on Majority Voting for Classification and (averaging for regression).

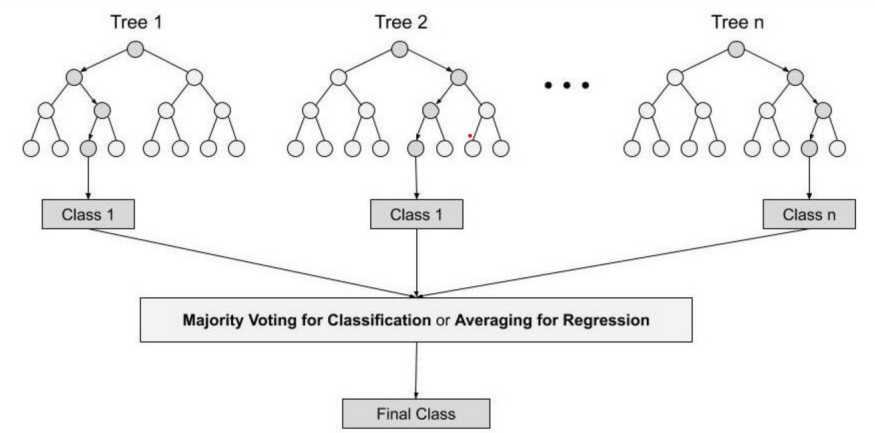


Figure 11.2: Random Forest algorithm synoptic scheme [71]

11.1.5 Quadratic Discriminant Analysis

Discriminant analysis belongs to the branch of classification methods called generative modeling, where we try to estimate the within-class density of X given the class label. Combined with the prior probability (unconditioned probability) of classes, the posterior probability of Y can be obtained by the Bayes formula.

11.2 Classification

To evaluate and estimate the rigor and robustness of the methodology designed throughout this work, and in order to access to the finality of the framework, we train and test the classifiers presented above. Yet, there are more considerations to take into account and the most important and well-known of them is undoubtedly ***over-fitting***.

When evaluating different hyper-parameters for estimators, there is still a risk of over-fitting on the test set because the parameters can be tweaked until the estimator performs optimally. This way, knowledge about the test set can "leak" into the model and evaluation metrics no longer report on generalization performance. To solve this problem, yet another part of the dataset can be held out as a "validation set". However, this partitioning of the data into three sets drastically reduces the number of samples used for learning the model.

A solution to this problem is a procedure called **cross-validation** [60]. A test set should still be held out for final evaluation, but the validation set is no longer needed when performing cross-validation. In the basic approach, called k**-fold CV**, the training set is split into k smaller sets. The following procedure is followed for each of the k "folds" and is summarized and illustrated in Fig. 11.3:

1. A model is trained using $k-1$ of the folds as training data;
2. the resulting model is validated on the remaining part of the data (*i.e.*, it is used as a test set to compute a performance measure such as accuracy).

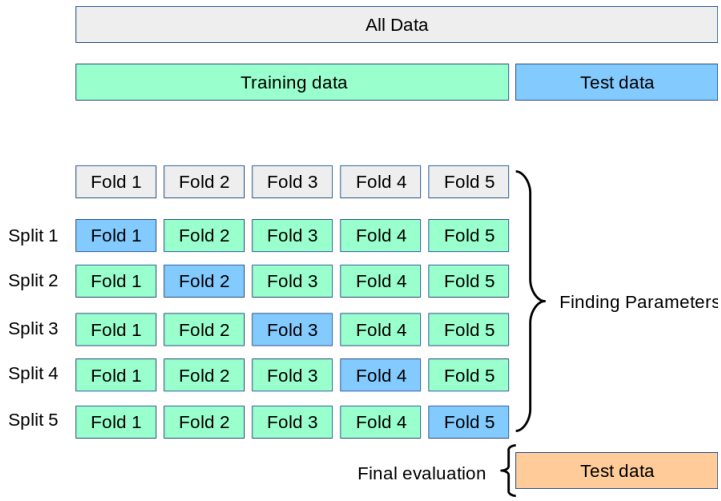


Figure 11.3: Cross-validation for $k = 5$ folds [60]

11.3 Hyperparameters' tuning for classifiers

In this section, we summarize in Tab. 11.1, the values we considered when we create an instance of each classifier. It's also important to mention that these parameters were hard-coded and not optimized. Though, we definitely can operate an optimization approach using a ***genetic algorithm***.

Logistic Regression	Gaussian Naive Bayes	Decision Tree
penalty=$'$ $l1'$ solver=$'$ $liblinear'$ max_iter= 2 tol= 0.9	$var\_smoothing =$ 3×10^{-21}	criterion='entropy' max_depth= 1

Quadratic Discriminant Analysis	Random Forest
One hyperparameter for QDA: *tolerance*, and through our experiments, we observed very little sensitivity of our performance to that parameter.	n_estimators= 1 criterion=$'entropy'$ max_depth= 2

Table 11.1: Summary of the classifiers' hyperparameters values used

This done, We also mention that we opt for a $10 - fold$ cross-validation and a final test size of 30% of the dataset as explicited in Fig. 11.3.

11.4 Classification results and discussion

The framework carried out and completed through Feature Selection and their hyper-parameters optimization, and the classifiers' established, built and trained, we illustrate the results of the previously established workflow for each of the classifiers presented.

We recall the definition of the accuracy of a model for the sake of precision of terms.

Definition 18 *Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right [44]. Formally, accuracy has the following definition:*

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11.4}$$

We first run, train, and perform an evaluation for each of the 10 folds of cross-validation. We obtain Fig. 11.4

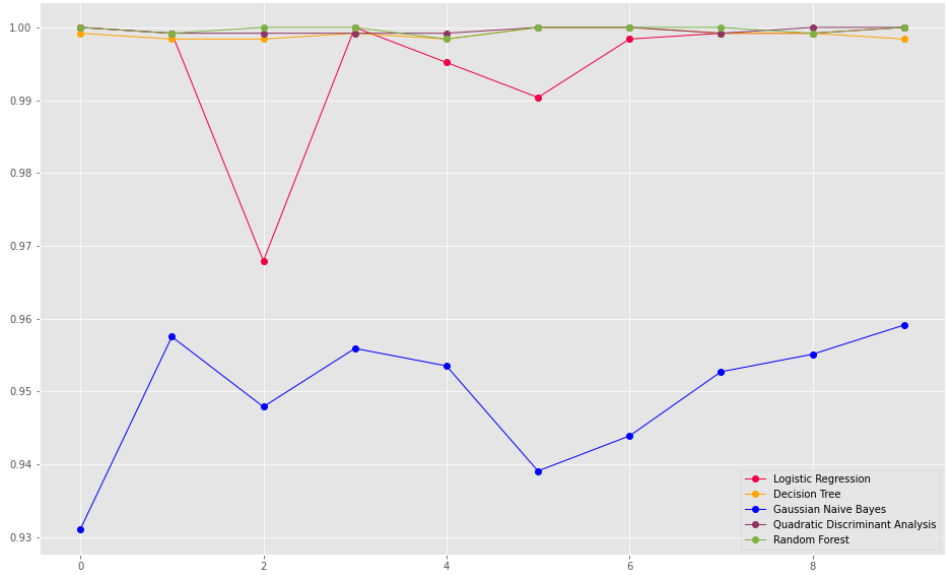


Figure 11.4: Performance of each of the 7 classifiers on each of the 10 folds

We first observe a very satisfying and consistent over the 10 folds for each of our algorithms. The majority of the performances, with an assurance of avoiding over-fitting by cross-validating. The accuracy obtained on the test set approximates the 1.0, which reflects close to $zero - error$. The Gaussian Naive Bayes performs just as well with an average accuracy above 0.9.

An additional step we considered in order to evaluate the contribution of each type of features is to include an accuracy evaluation once feeding the models with ***nonlinear dynamical features***. Right after, we consider exclusively the selected ***SCSA features*** and both results are set to be compared with the ones obtained when combining both features subsets. The results are illustrated in the heatmap in Fig. 11.5.

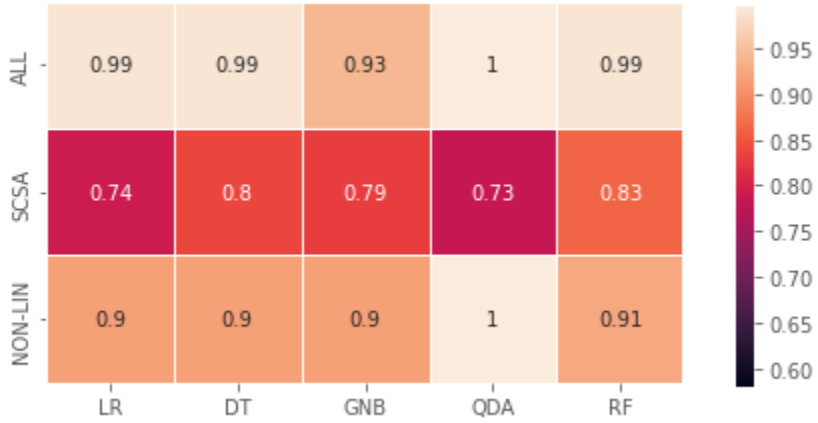


Figure 11.5: Average classification performance

From this summarizing heatmap, we first kick off our interpretations by focusing on the performance of the ***SCSA features*** which are the innovative addition and newest approach to characterize epileptic signals. Of the SCSA performance for classification, we denote a performance that approaches the $0.7 \sim 0.8$ performance. Hence, although an average of 78% of accuracy is not enough and translates an error rate that could be too high to consider directly; this accuracy rate reflects first and foremost a certain and clear ability of **Semi-Classical Signal Analysis** to characterize and capture the specificity within epileptic EEG signals. Therefore, the motives behind our introduction of SCSA for the characterization of epileptic seizures, are justified and supported by these results.

Moreover, the performance of the nonlinear dynamical features is, as expected and backed by previous research papers who affirmed the relevance of these features, quite high and reflect undeniable efficiency. Yet, there are two important points to note and discuss. The first concerns the hyper-parameters optimization we perform earlier in the workflow contributes to enhancing drastically their discrimination power between seizure and seizure-free samples by estimating in the right and rigorous way the dynamics encompassed within the neural signals processed.
Indeed, the overwhelming majority of the research cases who introduced and considered these features in the same context went with the standard values suggested by the already implemented algorithms or hard-coded values obtained by scrambling around numerically to reach a better classifier performance for these hyper-parameters [31] [87] [4].

The second point we can raise now is the indisputable boost SCSA features are capable of bringing to the classifiers' performance, allowing it most times a raise in accuracy close to the 10%. Therefore, the contribution of SCSA to totally autonomous frameworks designed to distinguish and detect epileptic seizures is noticeable and an axis to consider further and dig deeper in.

An additional step in our discussion is the observation of two scores and metrics considered by biomedical engineering and medical experts as most relevant in the estimation of the performance of detection in biomedical devices: **specificity and sensitivity**.

Definition 19 *Sensitivity measures the proportion of actual positive cases that got predicted as positive (or true positive). Sensitivity is also termed as Recall. This implies that there will be another proportion of actual positive cases, which is predicted incorrectly as negative (or false negative). In other words, the person who is unhealthy actually got predicted as unhealthy [53].*

Mathematically, sensitivity can be calculated as the following:

$$Sensitivity = \frac{TP}{TP + FN} \tag{11.5}$$

Definition 20 *Specificity is defined as the proportion of actual negatives, which got predicted as the negative (or true negative). This implies that the existence of another proportion of actual negative, which got predicted as positive and could be termed as false positives. In other words, the person who is healthy actually got predicted as healthy is specificity [53].*
Mathematically, specificity can be calculated as the following:

$$Specificity = \frac{TN}{TN + FP} \tag{11.6}$$

According to theses definitions, we obtain the following heatmap:

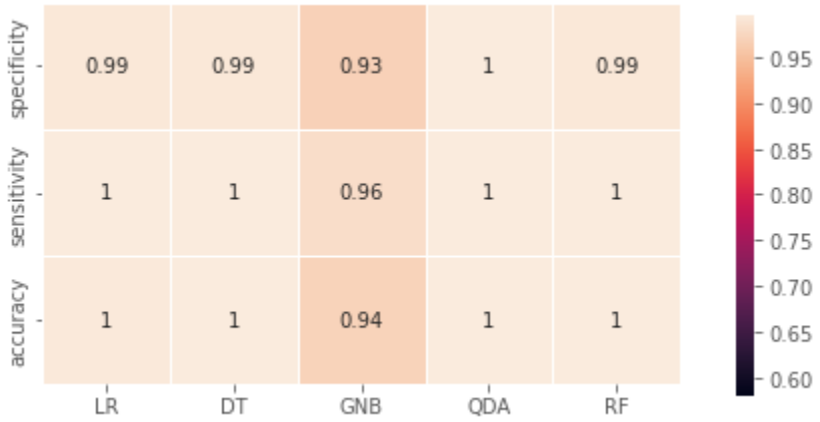


Figure 11.6: Heatmap of accuracy, sensitivity and specificity

This heatmap is satisfying and all the more backs the designed approach given how these two metrics are not as objective as Accuracy. Rather, they are a rigorous way to quantify performance in biomedical thematic given how each is interpreted and has very much its place in medical interpretation in terms of False and Missed alarms.

Going back to the past performance summary collected and presented in Chapter 4, we confidently assess the superior performance our framework exhibit in most cases and in a very much consistent way and that, in terms of accuracy, sensitivity and specificity.

We also assess the contribution of tackling the imbalanced nature of epileptic seizure data and therefore our ability to use over 600 samples for testing exclusively which definitely gives us more generalization power over past work.

Part V

General Conclusions and Future Directions

Chapter 12

General Conclusions and Future Directions

This contribution to the field of autonomous and automatic epileptic seizure detection aims to provide a brick in the wall of biomedical and neural engineering research that strives to help medical experts, improve the quality of life of patients and contribute, step by step, to the making of revolutionary devices for neuro-stimulation and neuro-feedback for rehabilitation and assistance. These devices help in medical context but also show to be of great help to neuroscientific in identifying biomarkers and neurometrics for neurological disorders. And with the advance of civilization and a rise in quality of life, interest in assisting and advancing in our contribution rates to marginalized populations have gained a few places in the list of priorities in major research powers. This axis of research has therefore seemed essential and natural to catch our interest and deserve our attention.

The dissertation before us have succeeded in encompassing a very interesting number of disciplines within engineering and each has proved its relevance and imposed its place within the workflow by its personal contribution to the ensemble. First and foremost, we ought to be satisfied with our contribution, that can be first and foremost formulated in:

> ***"Thoroughly striving to understand the effect and contribution of each and all features considered, their correlation and selection, and that, instead of tuning the hyper-parameters of the Machine Learning Classifiers we employ."***

The concrete contributions we made can be explicited and summarized in the following points:

- The introduction of the Semi-Classical Signal Analysis (SCSA) and its characterization of EEG data for epileptic seizures.

Indeed, we introduce, for the very first time in the research axis of Epileptic Seizure Detection, the Semi-Classical Signal Analysis that previously proved its efficiency for physiological signals. It has once more demonstrated its ability to characterize and extract specificity from physiological signals -in this case, epileptic seizure EEG data- and therefore it legitimizes our introduction of this novel signal representation in the context of epileptic seizure detection.

- Bounding explicitly the definition of Epileptic Seizures as an Engineering problem, to the nonlinear dynamical features.

In fact, the reasoning consists of capturing signal behavior as a nonlinear, chaotic system it behaves as for epileptic patients.

- The optimization of the time-embedding hyper-parameters is an additional aspect to the nonlinear feature engineering.

We established a very clear methodology to optimize these said hyper-parameters, according to the processed data we were exploiting, rather than utilize values directly from literature. This is an undeniable plus to the ensemble of the framework and manifested its power in the midst of nonlinear, dynamical feature extraction and generation.

- As for Data augmentation, we consider our contribution to be considerable and consequent to biomedical engineering and physiological signal processing. Our solution consists of using instead spectrogram matrices as input of a two-dimensional image generator Generative Adversarial Networks.

Our satisfactory results; and that, despite questions raised on the inversion of spectrograms to retrieve 1D-signals, can be summarized in:

– An avoidance of over-fitting

– A similarity, in terms of spectrogram, signal and features distributions demonstrated the integrity of the results of our 2D-GANs and the peripheral contributions necessary to its implementation.

While Generative Adversarial Networks have been disputed and questioned in their utility and utilization in the biomedical field, and the concern of the temporal coherence and consistency of segments of EEG information have been raised even during our own exploration of solutions for Imbalanced Learning.

- The aim of our approach to Classification is specifically to be able to tackle the issue of over-learning from raw data. In other words, it is putting Artificial Intelligence to the service of the extracted information and not the other way around.

We choose to rather rise in abstraction to extract the essence of the information held in EEG signals. In fact, our approach to classification as a final evaluation of the end-to-end framework using standard, far from complex classification algorithms, well-known and well-acquainted within the Machine Learning literature was a conscious choice. Additionally, the tuning of the parameters was done with keeping values in general close to the default ones to minimize its part of responsibility in the good, or bad performance of the framework.

- We were glad to be able to attain results that topped and compete with the best performances on the record to this day in terms of accuracy, sensitivity and specificity.

As for the future directions we consider at the epilogue of this research project, the most interesting aspect is inexorably the multiple doors it opens in return to contribute further, to do better. It is a confirmation that we were capable of designing an approached that proved its relevance but moreover, it opens questions, and more horizons to explore.

One crucial point we mentioned earlier concerning data, is on the subject of validating our framework on a dataset as interesting and rich as the CHB-MIT database. As explained, we will rely on the Ibex resources provided by King Abdullah University of Science and Technology's Supercomputing Lab. This database offers the possibility to deal with data from one same, identified

subject with coherence between inter-ictal and ictal data in one same segment, and spatial diversity through the multiple available channels. This peripheral work can open the door to more complex, better-defined seizures within the medical field as we explained early on in this report and allow a classification of types of seizures.

A second axis we are interested in exploring is exploiting spectrograms and extracting time-frequency features directly at their level. This suggestion stems from the interest time-frequency domain features are generating. That, also given we are in parallel developing further a **2D-SCSA** that is under formalization and exploitation.

As we mentioned SCSA, we also agree that there is still considerable work and exploration to do at the level of **1D-SCSA**. Although it has largely proved its legitimacy in our context, optimization at the level of the semi-classical parameter h is also within out scope of work using a gradient-descent optimization we are actively shaping or the use of the Curvature-SCSA as an optimization coupled to denoising.

In the optic of leaving a footprint in the field, considering the multiple recording and monitoring devices we presented in this report are also a key factor given how each of them has a specific axis of action, a population it touches more, an information it retrieves better.

And last but not least, is the consideration of other neurodegenrative disorders within our scope of research as Parkinson and Alzheimer and contribute to the common scientific effort to help the most disadvantaged of us.

Bibliography

[1] World Health Organization (WHO). *Epilepsy,* URL: https://www.who.int/health-topics/epilepsy#tab=tab_1.

[2] A. Aarabi, F. Wallois, and R. Grebe. "Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis". In: *Clinical Neurophysiology* 117.2 (2006), pp. 328–340. ISSN: 1388-2457. DOI: https://doi.org/10.1016/j.clinph.2005.10.006. URL: https://www.sciencedirect.com/science/article/pii/S1388245705004165.

[3] U Rajendra Acharya et al. "Characterization of focal EEG signals: a review". In: *Future Generation Computer Systems* 91 (2019), pp. 290–299.

[4] U. Rajendra Acharya et al. "Application of entropies for automated diagnosis of epilepsy using EEG signals: A review". In: *Knowledge-Based Systems* 88 (2015), pp. 85–96. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2015.08.004. URL: https://www.sciencedirect.com/science/article/pii/S0950705115003081.

[5] U. Rajendra Acharya et al. "Automated diagnosis of epileptic EEG using entropies". In: *Biomedical Signal Processing and Control* 7.4 (2012), pp. 401–408. ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2011.07.007. URL: https://www.sciencedirect.com/science/article/pii/S1746809411000838.

[6] U. Rajendra Acharya et al. "Automated EEG analysis of epilepsy: A review". In: *Knowledge-Based Systems* 45 (2013), pp. 147–165. ISSN: 0950-7051. DOI: https://doi.org/10.1016/j.knosys.2013.02.014. URL: https://www.sciencedirect.com/science/article/pii/S0950705113000798.

[7] MZ Ahmad et al. "Seizure detection using EEG: A survey of different techniques". In: *2016 International Conference on Emerging Technologies (ICET)*. IEEE. 2016, pp. 1–6.

[8] Qingsong Ai et al. "Chapter 2 - State-of-the-Art". In: *Advanced Rehabilitative Technology*. Ed. by Qingsong Ai et al. Academic Press, 2018, pp. 11–32. ISBN: 978-0-12-814597-5. DOI: https://doi.org/10.1016/B978-0-12-814597-5.00002-3. URL: https://www.sciencedirect.com/science/article/pii/B9780128145975000023.

[9] Emina Alickovic, Jasmin Kevric, and Abdulhamit Subasi. "Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction". In: *Biomedical Signal Processing and Control* 39 (2018), pp. 94–102. ISSN: 1746-8094. DOI: https://doi.org/10.1016/j.bspc.2017.07.022. URL: https://www.sciencedirect.com/science/article/pii/S1746809417301544.

[10] Sabrina Ammar and Mohamed Senouci. "Seizure detection with single-channel EEG using Extreme Learning Machine". In: *2016 17th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*. 2016, pp. 776–779. DOI: 10.1109/STA.2016.7952088.

[11] Ralph G Andrzejak et al. "Improved spatial characterization of the epileptic brain by focusing on nonlinearity". In: *Epilepsy research* 69.1 (2006), pp. 30–44.

[12] Ralph G. Andrzejak, Kaspar Schindler, and Christian Rummel. "Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients". In: *Phys. Rev. E* ().

[13] Ralph G. Andrzejak et al. "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state". In: *Phys. Rev. E* 64 (6 Nov. 2001), p. 061907. DOI: 10.1103/PhysRevE.64.061907. URL: https://link.aps.org/doi/10.1103/PhysRevE.64.061907.

[14] Elie Bou Assi et al. "Towards accurate prediction of epileptic seizures: A review". In: *Biomedical Signal Processing and Control* 34 (2017), pp. 144–157.

[15] Jan Awrejcewicz et al. "Quantifying Chaos by Various Computational Methods. Part 1: Simple Systems". In: *Entropy* 20.3 (2018). ISSN: 1099-4300. DOI: 10.3390/e20030175. URL: https://www.mdpi.com/1099-4300/20/3/175.

[16] Muhittin Bayram. "Assessment of EEG Signals Using Chaos Analysis Methods". In: 2015.

[17] Richard Bellman and E Lee. "History and development of dynamic programming". In: *IEEE Control Systems Magazine* 4.4 (1984), pp. 24–28.

[18] Wassila Benhamed. *Journal Elmoudjahid, Journée mondiale de l'épilepsie : 400.000 cas en algérie*. URL: http://www.elmoudjahid.com/%20fr/actualites/90338..

[19] Gabriel B Benigno, Ravi S Menon, and Hacene Serrai. "Schrodinger filtering: a precise EEG despiking technique for EEG-fMRI gradient artifact". In: *NeuroImage* 226 (2021), p. 117525.

[20] Sourav Bhaduri et al. "Matlab Tool for Residual Water Suppression and Denoising of MRS Signal using the Schrodinger operator". In: (2019).

[21] Boualem Boashash and Samir Ouelha. "Designing high-resolution time–frequency and time–scale distributions for the analysis and classification of non-stationary signals: a tutorial review with a comparison of features performance". In: *Digital Signal Processing* 77 (2018). Digital Signal Processing SoftwareX - Joint Special Issue on Reproducible Research in Signal Processing, pp. 120–152. ISSN: 1051-2004. DOI: https://doi.org/10.1016/j.dsp.2017.07.015. URL: https://www.sciencedirect.com/science/article/pii/S1051200417301653.

[22] Poomipat Boonyakitanont et al. "A review of feature extraction and performance evaluation in epileptic seizure detection using EEG". In: *Biomedical Signal Processing and Control* 57 (Mar. 2020), p. 101702. DOI: 10.1016/j.bspc.2019.101702.

[23] Elena Boto et al. "Wearable neuroimaging: Combining and contrasting magnetoencephalography and electroencephalography". In: *NeuroImage* 201 (2019), p. 116099.

[24] Jason Brownlee. *How to Develop a 1D Generative Adversarial Network From Scratch in Keras*. URL: https://machinelearningmastery.com/how-to-develop-a-generative-adversarial-network-for-a-1-dimensional-function-from-scratch-in-keras/.

[25] Jason Brownlee. *How to Develop a GAN for Generating MNIST Handwritten Digits*. URL: https://machinelearningmastery.com/how-to-develop-a-generative-adversarial-network-for-an-mnist-handwritten-digits-from-scratch-in-keras/.

[26] Jason Brownlee. *Naive Bayes for Machine Learning*. URL: https://machinelearningmastery.com/naive-bayes-for-machine-learning/.

[27] Jason Brownlee. *Tour of Data Sampling Methods for Imbalanced Classification*. URL: https://machinelearningmastery.com/data-sampling-methods-for-imbalanced-classification/.

[28] Richard C Burgess. "MEG for Greater Sensitivity and More Precise Localization in Epilepsy". In: *Neuroimaging Clinics* 30.2 (2020), pp. 145–158.

[29] Abderrazak Chahid et al. "Feature generation and dimensionality reduction using the discrete spectrum of the schrodinger operator for epileptic spikes detection". In: *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2019, pp. 2373–2376.

[30] Abderrazak Chahid et al. "Residual water suppression using the squared eigenfunctions of the Schrodinger operator". In: *IEEE access* 7 (2019), pp. 69126–69137.

[31] Shanen Chen et al. "Automatic Diagnosis of Epileptic Seizure in Electroencephalography Signals Using Nonlinear Dynamics Features". In: *IEEE Access* PP (May 2019), pp. 1–1. DOI: 10.1109/ACCESS.2019.2915610.

[32] Waiting Chen et al. "Characterization of Surface EMG Signal Based on Fuzzy Entropy". In: *IEEE transactions on neural systems and rehabilitation engineering : a publication of the IEEE Engineering in Medicine and Biology Society* 15 (July 2007), pp. 266–72. DOI: 10.1109/TNSRE.2007.897025.

[33] Eugenia Conti et al. "Looking for "fNIRS Signature" in Autism Spectrum: A Systematic Review Starting From Preschoolers". In: *Frontiers in Neuroscience* 16 (2022).

[34] Song Cui et al. "Learning EEG synchronization patterns for epileptic seizure prediction using bag-of-wave features". In: *Journal of Ambient Intelligence and Humanized Computing* (Sept. 2018). DOI: 10.1007/s12652-018-1000-3.

[35] Alfonso Delgado-Bonal and Alexander Marshak. "Approximate Entropy and Sample Entropy: A Comprehensive Tutorial". In: *Entropy* 21.6 (2019). ISSN: 1099-4300. DOI: 10.3390/e21060541. URL: https://www.mdpi.com/1099-4300/21/6/541.

[36] Jisu Elsa Jacob et al. "Can Chaotic Analysis of Electroencephalogram Aid the Diagnosis of Encephalopathy?" In: *Neurology Research International* 2018 (May 2018), pp. 1–8. DOI: 10.1155/2018/8192820.

[37] Oliver Faust et al. "Deep learning for healthcare applications based on physiological signals: A review". In: *Computer Methods and Programs in Biomedicine* 161 (2018), pp. 1–13. ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2018.04.005. URL: https://www.sciencedirect.com/science/article/pii/S0169260718301226.

[38] Oliver Faust et al. "Wavelet-based EEG processing for computer-aided seizure detection and epilepsy diagnosis". In: *Seizure* 26 (2015), pp. 56–64. ISSN: 1059-1311. DOI: https://doi.org/10.1016/j.seizure.2015.01.012. URL: https://www.sciencedirect.com/science/article/pii/S1059131115000138.

[39] Alberto Fernández et al. "Algorithm-Level Approaches". In: Jan. 2018, pp. 123–146. ISBN: 978-3-319-98073-7. DOI: 10.1007/978-3-319-98074-4_6.

[40] Robert Fisher et al. "Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)". In: *Epilepsia* 46 (May 2005), pp. 470–2. DOI: 10.1111/j.0013-9580.2005.66104.x.

[41] Robert S Fisher et al. "Instruction manual for the ILAE 2017 operational classification of seizure types". In: *Epilepsia* 58.4 (2017), pp. 531–542.

[42] Tomer Goldfriend and Jorge Kurchan. "Quantum Kolmogorov-Sinai entropy and Pesin relation". In: *Physical Review Research* 3.2 (2021), p. 023234.

[43] Ian Goodfellow et al. "Generative Adversarial Networks". In: *Advances in Neural Information Processing Systems* 3 (June 2014). DOI: 10.1145/3422622.

[44] Machine Learning Crash Course Google. *Classification: Accuracy*. URL: https://developers.google.com/machine-learning/crash-course/classification/accuracy?hl=fr.

[45] Aditya Grover. *Generative adversarial networks*. URL: https://deepgenerativemodels.github.io/notes/gan/.

[46] Roberto Hornero et al. "Nonlinear analysis of electroencephalogram and magnetoencephalogram recordings in patients with Alzheimer's disease". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 367.1887 (2009), pp. 317–336.

[47] Leonidas D Iasemidis and J Chris Sackellares. " REVIEW: Chaos Theory and Epilepsy". In: *The Neuroscientist* 2.2 (1996), pp. 118–126.

[48] ITSUsync. *Different Types Of Brain Waves: Delta, Theta, Alpha, Beta, Gamma*. URL: https://tex.stackexchange.com/questions/30145/inserting-citation-to-a-http-link-in-bibliography.

[49] Abeg Kumar Jaiswal and Haider Banka. "Local pattern transformation based feature extraction techniques for classification of epileptic EEG signals". In: *Biomedical Signal Processing and Control* 34 (2017), pp. 81–92.

[50] N. Kannathal et al. "Entropies for detection of epilepsy in EEG". In: *Computer Methods and Programs in Biomedicine* 80.3 (2005), pp. 187–194. ISSN: 0169-2607. DOI: https://doi.org/10.1016/j.cmpb.2005.06.012. URL: https://www.sciencedirect.com/science/article/pii/S0169260705001525.

[51] Holger Kantz. "A robust method to estimate the maximal Lyapunov exponent of a time series". In: *Physics Letters A* 185.1 (1994), pp. 77–87. ISSN: 0375-9601. DOI: https://doi.org/10.1016/0375-9601(94)90991-1. URL: https://www.sciencedirect.com/science/article/pii/0375960194909911.

[52] Anna Krakovská, Kristína Mezeiová, and Hana Budáčová. "Use of False Nearest Neighbours for Selecting Variables and Embedding Parameters for State Space Reconstruction". In: *Journal of Complex Systems* 2015 (Mar. 2015), p. 12. DOI: 10.1155/2015/932750.

[53] Ajitesh Kumar. *ML Metrics: Sensitivity vs. Specificity*. URL: https://dzone.com/articles/ml-metrics-sensitivity-vs-specificity-difference.

[54] Patrick Kwan et al. *Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies*. 2010.

[55] Taous-Meriem Laleg-Kirati, Emmanuelle Crépeau, and Michel Sorine. "Semi-classical signal analysis". In: *Mathematics of Control, Signals, and Systems* 25.1 (2013), pp. 37–61.

[56] Peihao Li and Taous-Meriem Laleg-Kirati. "Central blood pressure estimation from distal PPG measurement using semiclassical signal analysis features". In: *IEEE Access* 9 (2021), pp. 44963–44973.

[57] Peihao Li and Taous-Meriem Laleg-Kirati. "Schrodinger Spectrum Based PPG Features for the Estimation of the Arterial Blood Pressure". In: *2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2020, pp. 2683–2686.

[58] Peihao Li and T.M. Laleg-Kirati. "Signal denoising based on the Schrödinger operator's eigenspectrum and a curvature constraint". In: *IET Signal Processing* 15 (May 2021). DOI: 10.1049/sil2.12023.

[59] Peihao Li et al. "Automatic Detection of Epileptiform EEG Discharges based on the Semi-Classical Signal Analysis (SCSA) method". In: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2021, pp. 928–931.

[60] sci-kit learn literature. *What is a multi-layered perceptron?* URL: https://scikit-learn.org/stable/modules/cross_validation.html.

[61] Ziming Liu et al. "A systematic review on hybrid EEG/fNIRS in brain-computer interface". In: *Biomedical Signal Processing and Control* 68 (2021), p. 102595.

[62] David López-Sanz, Noelia Serrano, and Fernando Maestú. "The Role of Magnetoencephalography in the Early Stages of Alzheimer's Disease". In: *Frontiers in Neuroscience* 12 (2018). ISSN: 1662-453X. DOI: 10.3389/fnins.2018.00572. URL: https://www.frontiersin.org/article/10.3389/fnins.2018.00572.

[63] Sheng Lu et al. "Automatic Selection of the Threshold Value r for Approximate Entropy". In: *IEEE Transactions on Biomedical Engineering* 55.8 (2008), pp. 1966–1972. DOI: 10.1109/TBME.2008.919870.

[64] Dominique Makowski. *Complexity Analysis of Physiological Signals*. URL: https://neurokit2.readthedocs.io/en/latest/examples/complexity.html.

[65] SubTain Malik. *Demystifying the GAN using a 1D Function (Keras)*. URL: https://medium.com/analytics-vidhya/demystifying-the-gan-using-a-1d-function-keras-bc7b861bb304.

[66] Kasra Manshaei. *Complexity Analysis of Physiological Signals*. URL: How%20to%20understand%20ANOVA-F%20for%20feature%20selection%20in%20Python.%20Sklearn%20SelectKBest%20with%20f_classif.

[67] Ahmad Mirzaei et al. "Spectral Entropy for Epileptic Seizures Detection". In: *2010 2nd International Conference on Computational Intelligence, Communication Systems and Networks*. 2010, pp. 301–307. DOI: 10.1109/CICSyN.2010.84.

[68] Ahmed Hossam Mohammed et al. "Dynamics of Electrical Activity in Epileptic Brain and Induced Changes Due to Interictal Epileptiform Discharges". In: *IEEE Access* 10 (2022), pp. 1276–1288. DOI: 10.1109/ACCESS.2021.3138385.

[69] Mavuto Mukaka. "Statistics Corner: A guide to appropriate use of Correlation coefficient in medical research". In: *Malawi medical journal : the journal of Medical Association of Malawi* 24 (Sept. 2012), pp. 69–71.

[70] S Sen Purkayastha, VK Jain, HK Sardana, et al. "Topical Review: A review of various techniques used for measuring brain activity in brain computer interfaces". In: *Advance in Electronic and Electric Engineering* 4 (2014), pp. 513–522.

[71] Sruthi E R. *Understanding Random Forest*. URL: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/.

[72] Ashwin Raj. *An Exhaustive Guide to Decision Tree Classification in Python 3.x*. URL: https://towardsdatascience.com/an-exhaustive-guide-to-classification-using-decision-trees-8d472e77223f.

[73] Ashwin Raj. *Perfect Recipe for Classification Using Logistic Regression*. URL: https://towardsdatascience.com/the-perfect-recipe-for-classification-using-logistic-regression-f8648e267592.

[74] Joshua S. Richman and J. Randall Moorman. "Physiological time-series analysis using approximate entropy and sample entropy". In: *American Journal of Physiology-Heart and Circulatory Physiology* 278.6 (2000). PMID: 10843903, H2039–H2049. DOI: 10.1152/ajpheart.2000.278.6.H2039. eprint: https://doi.org/10.1152/ajpheart.2000.278.6.H2039. URL: https://doi.org/10.1152/ajpheart.2000.278.6.H2039.

[75] Gabriel Riviere. "Entropy of semiclassical measures in dimension 2". In: *Duke Mathematical Journal* 155.2 (2010), pp. 271–335.

[76] Ashish Rohila and Ambalika Sharma. "Phase entropy: a new complexity measure for heart rate variability". In: *Physiological Measurement* 40 (Oct. 2019). DOI: 10.1088/1361-6579/ab499e.

[77] Michael T. Rosenstein, James J. Collins, and Carlo J. De Luca. "A practical method for calculating largest Lyapunov exponents from small data sets". In: *Physica D: Nonlinear Phenomena* 65.1 (1993), pp. 117–134. ISSN: 0167-2789. DOI: https://doi.org/10.1016/0167-2789(93)90009-P. URL: https://www.sciencedirect.com/science/article/pii/016727899390009P.

[78] Kaveh Samiee, Peter Kovacs, and Moncef Gabbouj. "Epileptic seizure classification of EEG time-series using rational discrete short-time Fourier transform". In: *IEEE transactions on Biomedical Engineering* 62.2 (2014), pp. 541–552.

[79] Mark L Scheuer et al. "Seizure detection: interreader agreement and detection algorithm assessments using a large dataset". In: *Journal of Clinical Neurophysiology* 38.5 (2021), p. 439.

[80] Ali Hossam Shoeb. "Application of machine learning to epileptic seizure onset detection and treatment". PhD thesis. Massachusetts Institute of Technology, 2009.

[81] Siuly Siuly, Yan Li, and Yanchun Zhang. *EEG Signal Analysis and Classification: Techniques and Applications*. Jan. 2016. ISBN: 978-3-319-47652-0. DOI: 10.1007/978-3-319-47653-7.

[82] Piyush Swami et al. "A novel robust diagnostic model to detect seizures in electroencephalography". In: *Expert Systems with Applications* 56 (2016), pp. 116–130.

[83] A. Temko et al. "EEG-based neonatal seizure detection with Support Vector Machines". In: *Clinical Neurophysiology* 122.3 (2011), pp. 464–473. ISSN: 1388-2457. DOI: https://doi.org/10.1016/j.clinph.2010.06.034. URL: https://www.sciencedirect.com/science/article/pii/S1388245710006036.

[84] Lorena Vega-Zelaya, Elena Martín Abad, and Jesús Pastor. "Quantified EEG for the Characterization of Epileptic Seizures versus Periodic Activity in Critically Ill Patients". In: *Brain Sciences* 10.3 (2020). ISSN: 2076-3425. DOI: 10.3390/brainsci10030158. URL: https://www.mdpi.com/2076-3425/10/3/158.

[85] Sebastian Wallot and Dan Mønster. "Calculation of Average Mutual Information (AMI) and False-Nearest Neighbors (FNN) for the Estimation of Embedding Parameters of Multidimensional Time Series in Matlab". In: *Frontiers in Psychology* 9 (2018). ISSN: 1664-1078. DOI: 10.3389/fpsyg.2018.01679. URL: https://www.frontiersin.org/article/10.3389/fpsyg.2018.01679.

[86] Lina Wang et al. "Automatic epileptic seizure detection in EEG signals using multi-domain feature extraction and nonlinear analysis". In: *Entropy* 19.6 (2017), p. 222.

[87] Hussain Waqar et al. "Towards Classifying Epileptic Seizures Using Entropy Variants". In: *2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)*. 2019, pp. 296–300. DOI: 10.1109/BigDataService.2019.00052.

[88] Alan Wolf et al. "Determining Lyapunov exponents from a time series". In: *Physica D: Nonlinear Phenomena* 16.3 (1985), pp. 285–317. ISSN: 0167-2789. DOI: https://doi.org/10.1016/0167-2789(85)90011-9. URL: https://www.sciencedirect.com/science/article/pii/0167278985900119.

[89] Orhan G. Yalçın. *Image Generation in 10 Minutes with Generative Adversarial Networks*. URL: https://towardsdatascience.com/image-generation-in-10-minutes-with-generative-adversarial-networks-c2afc56bfa3b.