UNIVERSIDAD
DE ANTIOQUIA

A complexity fingerprint for the localization of the Epileptogenic Zone using machine learning and data-driven approaches

Yorguin José Mantilla Ramos

This thesis is submitted for the degree of
Electronics Engineer

Advisors
Richard Leahy, Ph.D.
Karim Jerbi, Ph.D.
Hernán Felipe García Arias, Ph.D.

Universidad de Antioquia
Facultad de Ingeniería
Departamento de Ingeniería Electrónica y de Telecomunicaciones
Ingeniería Electrónica
Medellín, Colombia
2023

Biomedical Imaging Group, University of Southern California.

Cognitive and Computational Neuroscience Laboratory, Université de Montréal.

Grupo Neuropsicología y Conducta, Universidad de Antioquia.

Centro de Documentación de Ingeniería (CENDOI)

I would like to dedicate this thesis to . . .

learning from others,
and others learning from me.

family, and friends
mentors, and mentees
beings. . . living and non-living alike
the places that received me.

No matter where I'd go,

Venezuela,
Colombia,
Australia,
Canada,
the United States. . .

I would find **you**
and **your enchantment**
and **enraptured**
you would transform me.

Te encontraría a **ti**
y a **tu magia**
y **encantado**
me transformarías.

Acknowledgements

The following acknowledgements are beyond this project, as I consider this thesis a culmination of my undergraduate development as a whole; a journey that took me 10 years, and almost five countries. I will acknowledge all of it, and I hope everyone is here. For you my dear, I will start from the beginning... I want to thank and acknowledge...

My friends, teachers and groups from **Colegio Arturo Michelena**, and **Instituto Madre Matilde**; among you **I enjoyed** my childhood and teen years.

My friends from **Gaitas Intercolegiales, Toto Music, Conservatorio Juan José Landaeta, Crea Musica, Grupo Instrumental y Voces USB**; from you I learned the way of music which **accompanies me forever and wherever**.

My professors and friends from **Universidad Simón Bolívar**: **Electrónica, Física, Matemática, Mecánica, Computación, Teleco..., Club de Audio USB, Laboratorio de Acústica y Comunicaciones**; from you I learned the way of Audio and Acoustics, the way of Math, Physics and their Philosophy, and **to dream big and endure, even if your country is on fire.**

My professors and friends from **Universidad de Antioquia**, and in particular from **Electrónica y Bioingeniería**. From you I learned the way of the engineer, and **to be a friend no matter where you are from, and how weird you are.**

Biohacking Colombia and **Grupo Genética y Bioquímica de Microorganismos**. From you I learned the way of Biology, but more importantly, **how Science has something for all of us**: for children, for parents, for grand-parents, for the victims, for the victimizers, for the lucky and the unlucky, for the professional and for the curious bypasser, for you and for me. I only came to work and to climb as high as I could... **Who knew I would find this much and much else...**

Grupo Neuropsicología y Conducta, its **Línea de Enseñanza y Aprendizaje de la Investigación**, **Semillero Neurociencias Computacionales NeuroCo** and my friends from **Medicina**. From you I learned the way of the Scientist in Neuroscience, EEG, Medicine, Social Science, and more importantly, **what it means to be a mentor and a mentee**, and to have una **alma semillerista**. It is all in the spirit of multidisciplinary curiousness and **I'm forever sorry that I had to go.**

My mentors and friends from **CoCo Lab at Université de Montréal**, from the neuroscience community I found in **Queensland and Swinburne, Australia**, from the **BIG Lab at the University of Southern California**, the **Cleveland Clinic** and from the **Open Source Neuroscience Software Community and Hackathons**. From you I learned the way of Neuro-AI and Neuroinformatics, how science is connected through the world and its cultures, and how it all stems from collaboration. **And so, we shall collaborate away and even when away collaborate.**

My family and friends from all over the world: **Baruta, Rosales, Gavilán, Caracas, Valencia, Mariara, Maracaibo, Río Caribe, Táchira, Coro, Medellín, Barranquilla, Bucaramanga, Sincelejo, Arauca, Miami, Orlando, Jacksonville, Homestead, Montréal, Austin, San Francisco, Mayo Clinic, Londombia, Australia, España, Grecia, Francia, Italia, Inglaterra, Noruega, Alemania, Nigeria, Rumania, Serbia, Argentina, Chile, México, Brásil, Los Angeles, the drum circle at Venice Beach, the dance community from the 3rd Street Promenade, West LA and Moorpark**, and who knows where else... I keep discovering you. You accompanied me through all of these stages and those to come, from you I learned the most important: **to love.**

My parents **Ingrid Soledad Ramos Blanco** and **Yorguin Mantilla Colmenares**. My sisters **Gabriela** and **Daniela**. My close friends, and dogs, and cats, and opossums, and parrots, and turtles, the mountains, the beaches ... **From you, I ...** And there are no words expressive enough that would do justice... **Let us just enjoy the journey.**

It all began, is, and will be...

because of you...

Abstract

In this research, we sought to delineate the epileptogenic zone using a dataset from the Cleveland Clinic, encompassing 28 patients who successfully underwent resective surgery and had prior SEEG recordings from both ictal and interictal periods. From time-windowed segments of these recordings, we derived complexity features and characterized them using their mean and standard deviation. Our analysis incorporated features such as Lempel-Ziv complexity, various entropies, fractal dimensions, and the 1/f slope of the brain activity spectrum, among others. We trained three distinct Logistic Regression Models: one using only ictal data, another using only interictal data, and a hybrid model leveraging both periods. Additionally, we trained a model using the Bern-Barcelona dataset, a known benchmark in interictal prediction. Our findings underscored that while the interictal period might be less informative in isolation, it enhances the insights drawn from the ictal phase when combined. A pivotal aspect of our research was discerning a distinctive epileptogenic zone fingerprint. Feature importance analysis pinpointed the Mean Lempel-Ziv Complexity, the standard deviation of the 1/f Slope, and the standard deviation of specific fractal dimensions as the most significant characteristics differentiating resected locations. These results not only contribute to understanding the epileptogenic zone but also foster discussions about complexity in the brain, particularly in the context of the brain criticality hypothesis.

Contents

List of Figures

List of Tables

Notation

Abbreviations

Concept	abbrv.	Concept	abbrv.
Epileptogenic Zone	EZ	Stereotactic electroencephalography	SEEG
Seizure Onset Zone	SOZ	Subdural electrodes electroencephalography	SDE
Potential Seizure Onset Zone	PSOZ	Power Spectral Density	PSD
Intracranial electroencephalography	iEEG	High-frequency oscillations	HFOs
Interictal Epileptiform Discharges	IEDs	1 over f	1/f
Self-Organized Criticality	SOC	Excitation/Inhibition	E/I
Approximate Entropy	ApEn	Sample Entropy	SampEn
Permutation Entropy	PE	Spectral Entropy	SE
Singular Value Decomposition Entropy	SVDEn	Multi-Scale Entropy	MSE
Lempel-Ziv Complexity	LZC	Normalized Lempel-Ziv Complexity	LZCn
Number of Zero Crossings	NZC	Zero Crossing Rate	ZCR
Hjorth Complexity	HComp	Hjorth Mobility	HMob
Katz fractal dimension	KFD	Petrosian fractal dimension	PFD

Concept	abbrv.	Concept	abbrv.
Higuchi Fractal Dimension	HFD	Generalized Linear Model	GLM
True Positive Rate (Sensitivity)	TPR	True Negative Rate (Specificity)	TNR
False Negative Rate	FNR	False Positive Rate	FPR
Positive Predictive Value (Precision)	PPV	False Discovery Rate	FDR
Accuracy	ACC	Receiver Operating Characteristic	ROC
Area Under the Curve	AUC	Minimum Redundancy Maximum Relevance	mRMR
Predicted Epileptogenic Inside Resection	EIR	Predicted Non-Epileptogenic Inside Resection	NEIR
Predicted Epileptogenic Outside Resection	EOR	Predicted Non-Epileptogenic Outside Resection	NEOR
Epileptogenic Predictive Value	EPV	False Epileptogenic Rate	FER
Resected Miss Rate	RMR	Standard Deviation	SD

Chapter 1

Introduction

1.1 Context

Epilepsy is a common neurological disease that affects 50 million people worldwide [6, 61] and has an estimated point prevalence of 6.38 per 1000 persons for active epilepsy[20]. The first line of treatment consists of anti-epileptic drugs. Unfortunately, around one third of the cases turn out to be pharmacoresistant [42, 82]. In these cases, one common course of action is to evaluate the viability of a surgical treatment, as this has the potential to eliminate the seizures. In some cases, other alternatives such as neural stimulation or the ketogenic diet are also considered [19]. If a surgery is successful, the patient becomes seizure-free, and in turn his/her quality of life greatly improves. Indeed, epilepsy is associated with psychological and social problems including conditions such as depression, anxiety, and cognitive dysfunction [15, 54].

A prerequisite for a surgical intervention is the identification of the Epileptogenic Zone (EZ). Currently, the standard procedure for this is through the visual inspection of intracranial electroencephalography (iEEG), a process which is challenging, time-consuming and prone to subjectivity [26, 35, 63]. Therefore, finding an objective set of characteristics to localize the EZ is of utmost importance. With advances in quantitative electroencephalography and data-driven approaches, researchers have made efforts to obtain a set of biomarkers of the EZ. So far, the most common clinical approach has relied on pathological high-frequency oscillations (HFOs) as the main characteristic [3, 35]. Nonetheless, the precise localization of the EZ is still an unsolved and challenging problem [26, 35].

In 2017, Grinenko et al. [26] proposed the use of time-frequency characteristics for the identification of the EZ, which they conceptualized as a "fingerprint". Such a spec-

trotemporal fingerprint expanded the set of high-level features by including not only fast activity but also preictal spikes and low-frequency suppression.

Interestingly, beyond oscillation-related features, electrophysiological brain signals contain a wealth of other properties that have remained largely underexplored in the context of epilepsy surgery; these include complexity-related metrics such as power laws appearing in the form of $1/f$ aperiodic component of the electrophysiological spectrum [16, 29, 56], and brain criticality measures [60]. Thus, the idea of deriving complexity fingerprints of the epileptogenic zone may be a promising approach that complements and extends previous work. If found to be efficient, novel complexity fingerprints may pave the way for more efficient approaches to localize the EZ and thereby improve the outcome of surgical resection in drug-resistant epilepsy. Furthermore, these fingerprints may improve our understanding of the etiology and pathophysiology of pharmacoresistant focal epilepsies.

1.2   Aims

1.2.1   General aim

To identify a complexity fingerprint for the localization of the Epileptogenic Zone using intracranial electroencephalography, machine learning and exploratory data-driven approaches with the goal of supporting the process of resective surgery and further improve the understanding of epilepsy.

1.2.2   Specific aims

1. To characterize and contrast through exploratory data analysis different complexity metrics with the goal of establishing a basis for the identification of the fingerprint.

2. To develop machine learning models through the iterative testing of different combinations of parameters, hyperparameters and feature sets.

3. To evaluate the complexity fingerprint and the developed models through performance metrics adapted to the clinical context of resective epilepsy surgery.

1.3   Research Contribution

In this study, a complexity-based approach to understanding the epileptogenic zone is presented. The research offers deeper insights into both ictal and interictal periods through the evaluation of different machine learning models leveraging each of them. The results showcase how information from seemingly less informative periods can bolster overall insights. A notable contribution lies in the identification of a distinctive fingerprint for the epileptogenic zone, emphasizing the importance of Lempel-Ziv Complexity, Multi-Scale Entropy, the 1/f Slope and specific fractal dimensions. Additionally, benchmarking against the recognized Bern-Barcelona dataset strengthens the validity of some of the main results. This work not only deepens the comprehension of the epileptogenic zone but also catalyzes discussions on the role of complexity in the brain, echoing implications relating to the broader brain criticality hypothesis.

1.4   Outline of the Thesis

The thesis is structured as follows:

- Chapter 1 Provides general context to the problem and presents the aims and contributions of the project, setting the stage for the subsequent chapters.

- Chapter 2 delves into the theoretical background, introducing concepts from Epilepsy, Complexity, and Machine Learning.

- Chapter 3 details the methodology employed, discussing the process of data acquisition, feature extraction, and the design and training of the Logistic Regression Models.

- Chapter 4 presents the results of the models, comparing the insights drawn from the ictal, interictal, and hybrid approaches. Benchmark comparisons with the Bern-Barcelona dataset are also shown.

- Chapter 5 provides interpretations of the results obtained for the exploratory analysis of the dataset, the machine learning performances, and their feature importance.

- Chapter 6 summarizes the key contributions of the thesis and discusses potential avenues for future work.

Chapter 2

Theoretical Framework

2.1 Epilepsy

Epilepsy can be thought of as a family of disorders with a common factor: an enduring predisposition for seizures once one has occurred [21]. Epilepsy syndromes comprise a diverse set of etiologies (e.g. structural, genetic, among others); nevertheless they can be classified into two broad categories depending on whether the seizure is generalized or focal in nature [87]. A generalized epilepsy will have seizures with an onset occurring on both hemispheres while a focal epilepsy will be characterized by seizures originating in one hemisphere from one or multiple foci [10]. In the current work, the epilepsies best treated by surgical resection are explored; these correspond to the focal pharmacoresistant variants [43].

2.1.1 The Epileptogenic Zone

The surgical procedure rests on the assumption of a region of the cortex that upon resection produces seizure freedom, and moreover, such area must be the one necessary and sufficient for their generation. Lüders et al. [49] coined the term "Epileptogenic Zone" (EZ) to conceptualize this region. More than a decade later the authors revised the concept as the minimum cortex region that when resected causes seizure freedom [50]. Neither of these concepts is directly measurable as they are not tied to electrophysiological discriminative features. Nevertheless, it is postulated that this region must include the Seizure Onset Zone (SOZ, also known as Ictal Onset Zone), which is defined as the area of the cortex that initiates seizures [48, 72]. This zone could be measured as it may showcase a signal signature such as fast synchronizing discharge [48]. To further

complicate the issue, the SOZ is only a subset of the EZ, as the later also encompass a "Potential Seizure Onset Zone" (PSOZ), which accounts for multiple SOZs in the same epileptogenic zone, but some of them being evident only after the resection of the others [72]. Thus the Epileptogenic Zone can be thought of as including an "Actual" SOZ (the conventional one) and a "Potential" SOZ. As illustrated in Figure 2.1, failing to resect both the actual and the potential SOZs would lead to a lack of seizure freedom. There is no agreement over what characteristics could the PSOZ have or if it can be measured at all [50]; some viewpoints conceptualize it as part of the areas of early seizure spread, others as it being related to another conceptual region called the Irritative Zone, which is defined as the cortex area that generates interictal spikes [36]. As it may be apparent from the previous recount of concepts, the question of "what to cut?" is complex. Moreover, it is extremely relevant as the resection has the potential of causing postoperative neurological deficits, particularly if it is near the eloquent cortex [85].

Seizure-Free Resection

Non Seizure-Free Resection missing part of the potential seizure onset zone

Non Seizure-Free Resection missing part of the actual seizure onset zone
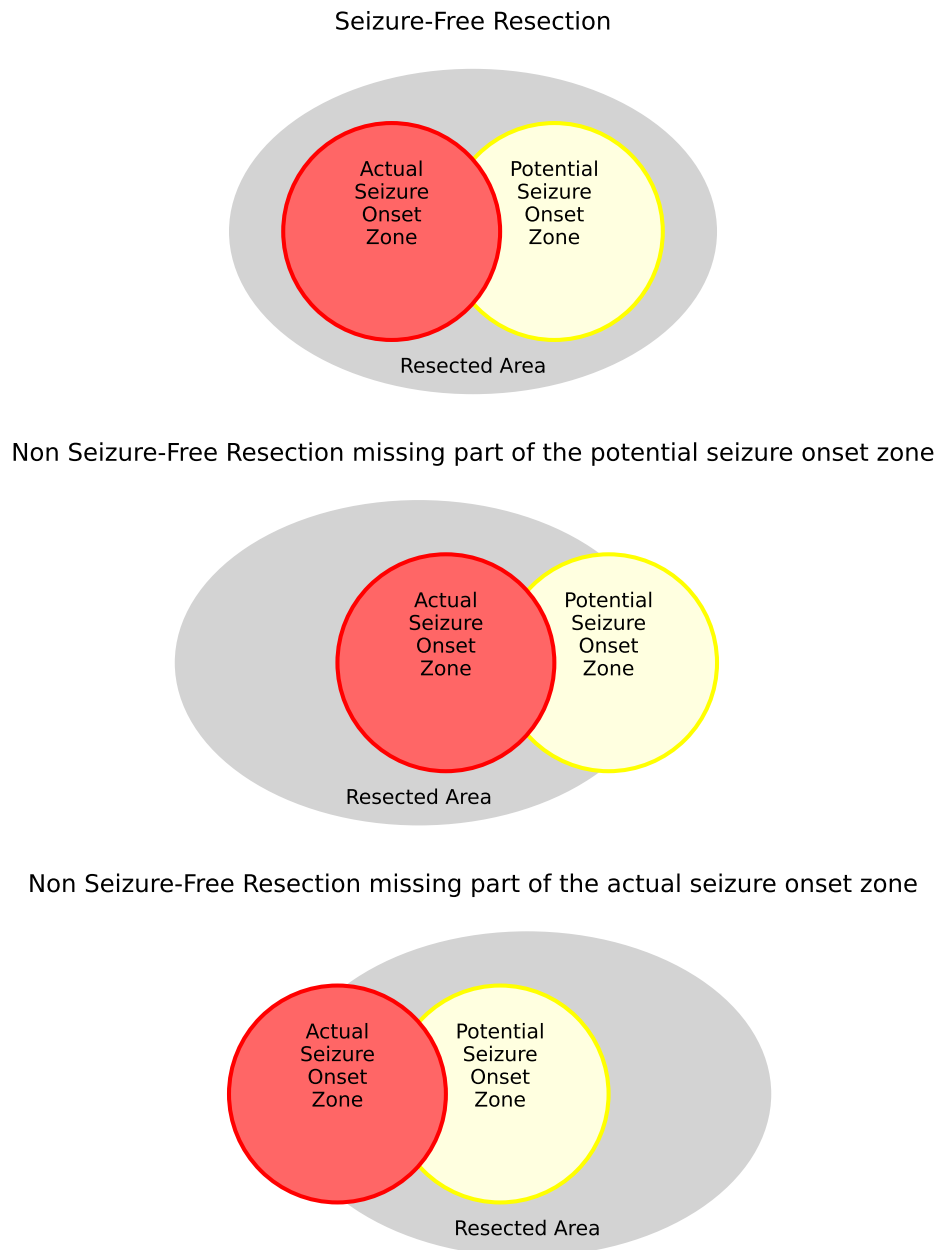
Figure 2.1: Illustration of the different overlaps between the Epileptogenic Zone and the surgical resection. The Epileptogenic Zone is theorized to encompass both an actual and a potential seizure onset zone. The two must be resected to achieve seizure freedom. Adapted and redrawn from Lüders et al. [50].

2.1.2 Ictal and Interictal Activity

Different periods naturally arise in the study of epilepsy, each defined in relation to the occurrence and development of a seizure, known as the ictal phase. The periods immediately before and after a seizure are defined as the preictal and postictal periods, respectively. Importantly, the interictal stage refers to the periods of time when no seizure is occurring, and the brain is not in a state particularly close to one. This is the stage where most people with epilepsy spend the majority of their lives. Thus, the ability to extract useful information from this stage is especially relevant. Naturally, these two periods have different behaviors within the Epileptogenic Zone. These can be seen on the time-frequency maps of Figure 2.2. Notably, Grinenko et al. [26] proposed a spectro-temporal fingerprint of the Epileptogenic Zone, which comprised three main characteristics: 1) Spikes, preceding (2) Multiband Fast Activity, coinciding with (3) Low-Frequency Suppression, also shown on Figure 2.2.
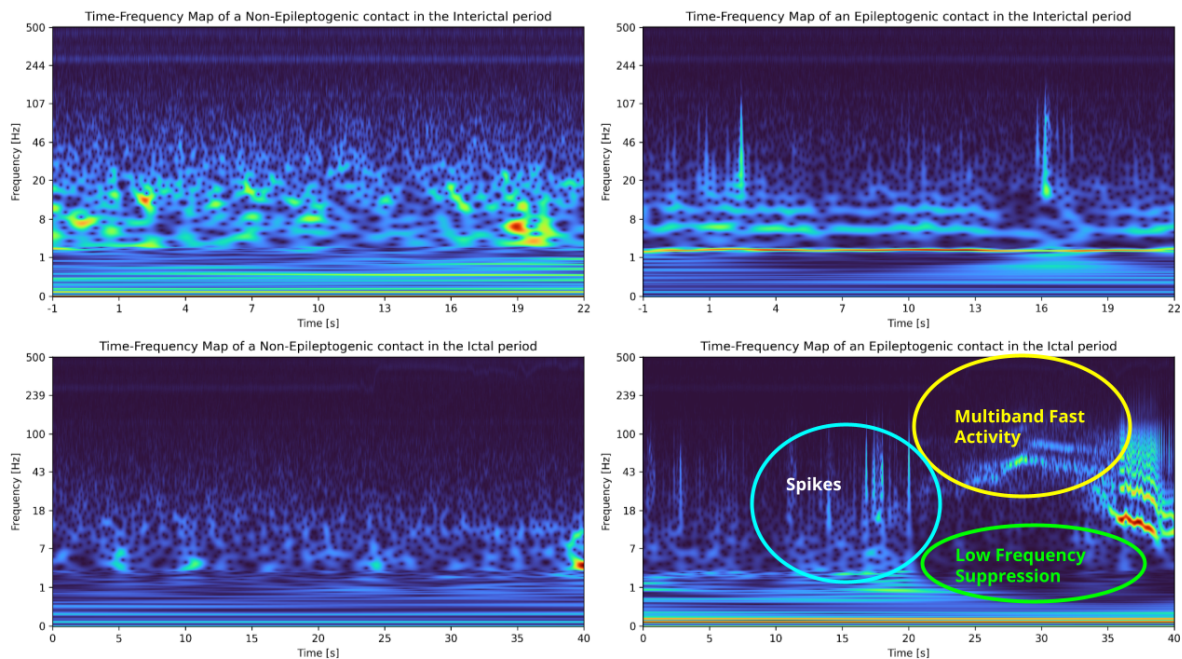
Figure 2.2: Ictal and Interictal stages on Epileptogenic and Non-Epileptogenic contacts. Grinenko et al. [26] hypothesizes that the Epileptogenic Zone has a particular spectrotemporal behavior during the ictal stage: Spikes preceding Multiband Fast Activity, which coincides with Low-Frequency Suppression.

2.2 Intracranial Electroencephalography

To diagnose, study and localize focal epilepsy, the first line of techniques consists of non-invasive methods like video-electroencephalography and magnetic resonance imaging; in some cases, these tools may be sufficient to select a resection area [5, 59]. Nevertheless, the gold standard remains intracranial electroencephalography (iEEG), used when non-invasive methods are insufficient. iEEG provides a higher signal-to-noise ratio, less artifacts, and naturally, a more close view of the phenomenon [58]. Two common forms of iEEG are used for epilepsy: stereotactic electroencephalography (SEEG) and subdural electrodes electroencephalography (SDE) [38]. In this work, SEEG recordings from pharmacoresistant epilepsy patients are used. Stereoelectroencephalography consists of depth electrodes with contacts spaced 2 to 10 mm apart and with less than 1 mm of diameter [33]. In comparison to SDE, SEEG has a sparse sampling on the cortical surface but offers access into the depth of the brain, such as sulcal areas and deep brain structures along with fewer surgical risks [24, 33].

2.3 Complexity Features

Here "complexity" is used in the sense of complexity science, which is an umbrella term encompassing the study of complex systems, across many domains, under the same view: the appearance of emergent behavior from the interconnection and dynamics of inter-dependent components [44]. In this framework, tools such as those derived from chaos theory in mathematics and dynamical systems in physics are used to get insights about the behavior of complex systems. In the last decade, the study of brain dysfunction through a complexity point of view has become increasingly relevant [44, 52, 93]. Different categories of complexity features extracted from electrophysiological brain signals were explored in this project. Motivation for each one of these stems from the literature as will become apparent in the following paragraphs.

2.3.1 Brain Criticality

Criticality is a concept stemming from the study of dynamical systems. It consists of the idea of "critical states" near phase transitions where many-bodied systems exhibit a special kind of collective behavior. Correspondingly, Brain Criticality consists of the hypothesis of the brain self-organizing into critical-states to achieve optimal information processing capabilities [60, 97]. One of the core concepts of criticality is the existence of a

control parameter (or many), which upon its crossing through a critical point, makes the collective behavior of the system different. For example, one may think of temperature as the control parameter of water, as it is possible to change its large-scale behavior (e.g. solid, liquid, gas states) through the variation of temperature.

Some natural questions that arise when discussing Brain Criticality include: What kinds of states are near the phase transition? What is the control parameter in the context of brain dynamics?. In the case of water, the large-scale behavior is radically different when it freezes or boils. This is an example of a discontinuous phase transition. In the case of the brain, its dynamics don't suddenly jump, but rather a more gradual or continuous transition is exhibited. In particular, the brain seems to maintain itself at -or near- a phase transition point. To achieve this some degree of tuning is expected to be happening , and even more, some systems exhibit what is known as Self-Organized Criticality (SOC), which consists of a system being able to self-tune to be around these critical points (e.g. sandpiles). Indeed, some authors support that SOC may be happening in the brain [30]. On the other hand, one of the possible candidates of a control parameter in the brain is the Excitation/Inhibition (E/I) balance, which reflects the interplay between excitatory and inhibitory neural signals, and its role on brain functioning. It is important to note that **Criticality is not a single feature or metric** but rather a property of a system that can be measured through many different ones, such as the E/I balance. Which metric to use is also subject to the theoretical underpinnings of the study.

Despite the growing interest in criticality interpretations of brain dynamics, these approaches have very rarely been explored for the specific task of EZ delineation. However, it has been used to characterize the epileptic brain. Witton et al. [91] argues that the Hurst Exponent may capture the deviation of the epileptic brain from a "healthy criticality" state, and that it could be used to localize the EZ. Similarly, Meisel et al. [55] suggest that during epileptic seizures the brain departs from self-organized criticality. In apparent contradiction, Cranstoun et al. [13] shows that the human epileptic hippocampus demonstrates SOC, and the work by Worrell et al. [92] agrees by mentioning that the epileptic focus may exhibit SOC. Moreover, Hagemann et al. [27] points out that in spite of epilepsy, the cortex works in a stable, slightly subcritical regime, disproving their initial hypothesis of focal areas operating at instability.

2.3.2 The 1/f Aperiodic Component

The $1/f$ distribution of the power spectral density of brain signals can be related to scale-free network dynamics which follow power laws [29, 56]. Specifically what this means is that the brain shows decaying power as frequency increases, following a power law with a negative exponent. This type of behavior has been associated to networks that show self-similar properties when inspected at different scales, hence "scale-free". In the brain this is represented by similar patterns occurring across many spatiotemporal scales, e.g. when comparing the spectrum of electroencephalograpic signals to functional magnetic resonance imaging signals (which have a higher spatial resolution but a lower temporal one in respect to electroencephalography) [28]. One of the defining properties of the power laws representing this scale-free behavior is the exponent, which correspond to the slope when the data is plotted on log-log space. A neurophysiological interpretation of this exponent of the $1/f$ distribution as a power law has been made by Gao et al. [22], relating the log-log space $1/f$ slope to the excitation-inhibition balance in the brain. Some research has suggested that this balance may be relevant to the pathophysiolgical underpinnings of epilepsy [18, 47, 79]. Recently, Jiang et al. [37] noted that the $1/f$ slope was steeper inside the Seizure Onset Zone, and that it was insufficient when used as a feature in isolation to discriminate the EZ from normal tissue. In particular, the slope will be negative, and one is interested in its magnitude: the steeper (higher magnitude) suggests more inhibition, and correspondingly, more flatness represents more excitation.

Electroencephalographic brain signals exhibit both a periodic (e.g. oscillatory) component and an aperiodic one. The periodic components have been studied thoroughly since the birth of electroencephalography, being usually associated with oscillations in frequency bands such as alpha, beta, and gamma. On the other part, the aperiodic component, as the name suggest, is related to arrhythmic activity and it is actually the part of the spectrum corresponding to the 1/f dynamic. Because of this, one way to measure the exponent (or slope) of the 1/f power law is to first remove the periodic component and then model the remaining spectrum as a decaying power law (or line in log-log space) to parameterize the 1/f. Figure 2.3 (A) illustrates the original raw spectrum typical of an electroencephalographic signal; notice the decaying power behaviour and the oscillatory peaks. Figure 2.3 (B) shows the decomposition of the spectrum into the periodic and aperiodic components, where the aperiodic would be subsequently parameterized to obtain the exponent of the decaying power law.
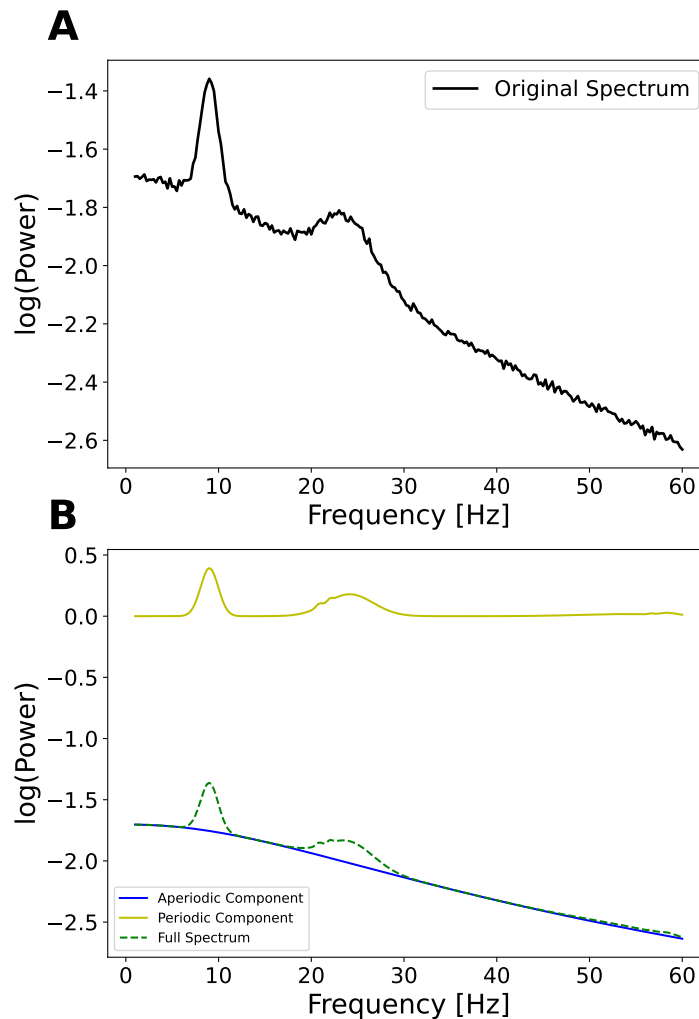
Figure 2.3: Illustration of the typical 1/f distribution of electrophysiological brain signals. (A) shows an example of an electroencephalography spectrum with two oscillatory components centered around 8 and 23 Hz. (B) shows the decomposition of this spectrum into its periodic and aperiodic components, where the sum should sum to the full spectrum. The exponent (or slope) can be extracted from the aperiodic component which serves as a characteristic parameter of the 1/f behavior.

2.3.3 Entropies

In complexity science, the concept of "entropy" is used in the sense explored in information theory, that is, it can be thought of as a measure of the amount of uncertainty (or surprise) inside of a signal, which in turn can arguably be interpreted as the amount of information it contains. In the theory of dynamical systems, entropy can be interpreted as a measure of the amount of disorder (or conversely, regularity), in a dynamical

system [57]. For example, one could measure the amount of repetition in a time series stemming from a biological system. Nevertheless, one may ask what is the connection between complexity and entropy. What becomes apparent is that there is a close relation between complexity, information and the order-disorder gradient. Grassberger [25] emphasizes that complexity is mediated by perception, as it is related to something having meaning, which requires an observer. To simplify the discussion, one can think of complexity as being somewhere in between order and disorder, as illustrated in Figure 2.4. There we can see that meaningful information is not equated to randomness.

In the context of epilepsy, a wide range of entropy approaches for the identification of the EZ were reported in a recent survey done by Islam et al. [35]. Another example is the work by Akter et al. [1] where a set of 8 entropy metrics, computed along the high-frequency components of interictal iEEG, were used for the problem of EZ localization. In these works, many different variants of entropy are employed. Here we will make use of the following measures of entropy.
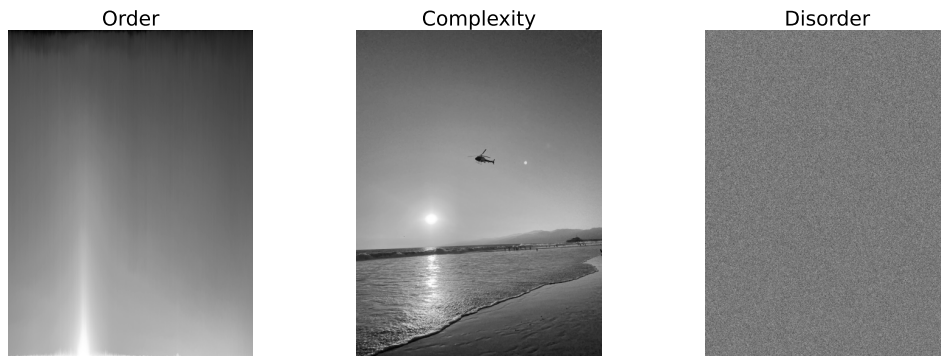
Figure 2.4: Illustration of complexity arising between order and disorder. The image at the center is a photo of a beach, where meaningful objects can be perceived. The image to the left is the same photo sorted by intensity on each column, representing an ordered set. The image to the right is the same image but randomly ordered, where no meaningful information is distinguished by us. Adapted and redrawn from [35].

Shannon's Entropy

The first introduction of entropy to the field of information theory was done by Claude Shannon [76]. For a discrete random variable X taking values in the alphabet $\mathcal{X}$ with probability mass function $p(x)$ Shannon's Entropy can be defined as:

$$H(X) = \sum_{x \in \mathcal{X}} p(x) \log_2 \left(\frac{1}{p(x)} \right) = - \sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \tag{2.1}$$

Intuitively, Shannon's entropy can be thought of as the average amount of uncertainty of a sequence. It involves the sum of the probabilities of the occurring elements of a sequence weighted by the amount of surprise associated to them. Surprise has an inverse relation to probability (e.g. $1/p$), but to avoid obtaining 1 when there is complete certainty (1/p=1/1), the logarithm is applied ($log(1/1) = 0$). Shannon's formula includes the negative as a result of simplifying the logarithm of a fraction. The logarithm base used is flexible as one can convert between bases easily.

Approximate Entropy

Approximate Entropy (ApEn) is a metric used to gauge the consistency, stability, or predictability within a data series introduced by Pincus et al. [68]. Low ApEn scores indicate a highly stable, repetitive, and easily forecastable system with recurring patterns throughout the data set. In contrast, high ApEn values suggest a lack of correlation among the data points, fewer repeating sequences, and greater randomness.

Approximate Entropy quantifies a data series' predictability by measuring the frequency of similar sub-sequences for varying lengths. Consider a time-series $\{u(t)\}_{t=1}^{N} = \{u_1, u_2, \dots, u_N\}$, parameters m (window length) and r (tolerance), and a choice of a distance metric $d(a, b)$. We can define Approximate Entropy procedurally as follows:

- **Vector Formation**. Construct $N - m + 1$ vectors which are subsequences of length m from the original vector:

$$X(i) = [u(i), u(i+1), ..., u(i+m-1)] \tag{2.2}$$

- **Distance Computation**. Calculate the distance of each subsequence to all others. In essence, this produces a 2D distance matrix between all pairs of sequences.

$$D(i, j) = d(X(i), X(j)) \tag{2.3}$$

- **Calculate Counts**. For each subsequence (the current one being called the template), we obtain the number of subsequences whose distance are below a tolerance r. This means that we consider those vectors as similar to the current template. During this we accept the occurrence of $i = j$, that is the template matches itself. The result is normalized by the number of subsequences. More formally, this quantity represents a conditional probability.

$$C_r^m(i) = \frac{\text{Number of } X(j) \text{ similar to } X(i) \text{ given the distance metric and r}}{N - m + 1} \quad (2.4)$$

- **Average the Counts** after using the natural logarithm to transform the data. This quantity is define as ϕ.

$$\phi_r^m = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln(C_r^m(i)) \quad (2.5)$$

- **Compare to an increase of window length**. Approximate Entropy quantifies the log-based probability that sequences similar over m data points continue to be similar in subsequent, incremental comparisons. As such, the metric is defined as:

$$\text{ApEn}(m, r, N)(u) = \phi_r^m - \phi_r^{m+1} \quad (2.6)$$

Typically, the distance metric used corresponds to the Chebyshev distance, which is defined as the largest disparity between two vectors in any single coordinate dimension. Moreover, the tolerance is usually chosen between 0.1 and 0.25 times the standard deviation of the signal under study. For a more comprehensive explanation of Approximate Entropy, many works exists such as the ones by Delgado-Bonal and Marshak [14], and Pincus and Goldberger [67].

Sample Entropy

Sample Entropy (SampEn) is a modification of Approximate Entropy proposed by Richman and Moorman [71]. This adjustment makes the metric less dependent on the length of the data N in comparison to ApEn. The argument is that, since ApEn counts similarities of subsequences to themselves, this generates a bias that makes the signal appear more regular than it actually is [71]. Specifically, when the data length is large, the frequency of self-matches increases, augmenting this bias. On the contrary, Sample Entropy approaches the same question and procedure as ApEn but without including self-similarities. Moreover, the logarithm in SampEn is included in a slightly different manner to avoid indeterminations.

Consider a time-series $\{u(t)\}_{t=1}^N = \{u_1, u_2, \ldots, u_N\}$, parameters m (window length) and r (tolerance), and a choice of a distance metric $d(a, b)$. Similar to ApEn, the pro-

cedure involves the formation of subsequences and distance calculations. The difference between the two occurs at the counting stage:

- **Calculate Counts**. For every specific subsequence, referred to as the template, the count of other subsequences that fall within a predefined distance r is determined. In this context, these counted subsequences are viewed as similar to the template in question. The template's self-match, where $i = j$, is deliberately not included in the count. Finally, this count is normalized by dividing it by the total number of possible subsequences (excluding the self-match).

$$B_r^m(i) = \frac{\text{Num. of } X_m(j) \text{ similar to } X_m(i) \text{ given d(a,b) and r, excluding self-matches}}{N - m - 1}$$

(2.7)

Similarly this is defined for the incremental increase of the points:

$$A_r^m(i) = \frac{\text{Num. of } X_{m+1}(j) \text{ similar to } X_{m+1}(i) \text{ given d(a,b) and r, excluding self-matches}}{N - m - 1}$$

(2.8)

- **Average the Counts**. Contrary to ApEn, the natural logarithm won't be applied prior to this step.

$$B_r^m = \frac{\sum_{i=1}^{N-m} B_r^m(i)}{N - m}$$

(2.9)

$$A_r^m = \frac{\sum_{i=1}^{N-m} A_r^m(i)}{N - m}$$

(2.10)

- **Compare**. Find the ratio of A_r^m to B_r^m and apply the natural logarithm.

$$\text{SampEn}(m, r, N) = -ln\left(\frac{A_r^m}{B_r^m}\right)$$

(2.11)

Permutation Entropy

Permutation Entropy (PE) is a non-parametric complexity metric that uses the order of data points to analyze the patterns and relationships over time in either linear or non-linear sequences. It was defined by Bandt and Pompe [4]. It has the advantage of being

robust against dynamical or observational noise. Consider a time-series $\{u(t)\}_{t=1}^N = \{u_1, u_2, \ldots, u_N\}$, parameters m (segment length, or embedding dimension) and τ (delay). Permutation Entropy is computed as follows:

- **Produce a time-delay embedding** of the signal through a series of $N - (m-1)\tau$ vectors defined as:

$$X_i = (u_i, u_{i+\tau}, u_{i+2\tau}, \ldots, u_{i+(m-1)\tau}) \quad \text{for} \quad i = 1, \ldots, N - (m-1)\tau \qquad (2.12)$$

From this, the embedding space is constructed as:

$$X = [X_1, X_2, \ldots, X_{N-(m-1)\tau}]^\top \qquad (2.13)$$

The resulting embedding will be of shape $[N - (m-1)\tau, m]$, which represents $N - (m-1)\tau$ segments produced by taking samples of the signal each τ indexes, until you have m samples (so the segments are of length m). This is done separately for consecutive starting points until the last one possible (the $N - [m-1]\tau$ sample). The use of such embedding is supported by Taken's theorem, which states that it is possible to reconstruct the phase space of a dynamical system through its observed time-series by constructing a time-delay embedding [80]. In particular, the phase space reconstructed is m dimensional. An exploration of the different choices of embedding can be found in [81].

- **Find sorted permutations** for each embedding vector. Specifically, compute the indexes that order each vector in ascending order after performing a ranking of the values, which is equivalent to one of its possible permutations. Each permutation is a list of integers, as follows for embedding vector X_i:

$$\pi_i = (\pi_{i1}, \pi_{i2}, \ldots, \pi_{ij}, \ldots, \pi_{im}) \qquad (2.14)$$

Where π_{ij} represents the position of the j-th smallest element in the embedded vector X_i. Thus there will be $N - (m-1)\tau$ total permutations found, which are not necessarily unique.

- **Calculate the frequency of each permutation**. Each embedding vector will have an associated permutation, but some vectors will have the same permutation that sorts them. Given this, count the number of times a given permutation

appears and divide that count by the total number of permutations $(N-[m-1]\tau)$, effectively obtaining the frequency of each. This can also be interpreted as the probability of permutation π.

$$p(\pi) = \frac{\text{Count of } \pi}{N - (m-1)\tau} \tag{2.15}$$

- **Compute Shannon's Entropy** over the possible permutations Π to obtain the desired permutation entropy:

$$PE = -\sum_{\pi \in \Pi} p(\pi) \log_2(p(\pi)) \tag{2.16}$$

Spectral Entropy

Inouye et al. [34] proposed to calculate Shannon's Entropy over the Power Spectral Density (PSD) of a signal as a way to measure the "relative peakedness or flatness" of the power spectrum. Highly oscillatory signals will have pronounced peaks in the spectrum, whereas irregular signals will show a flatter spectrum. For example, a pure sine will have the smallest amount of entropy, while white noise produces the greatest entropy. Thus, the degree of signal irregularity can be measured by applying the concept of entropy to the spectrum.

To calculate Spectral Entropy (SE), one merely has to extract the PSD, normalize it to obtain relative power, and then apply Shannon's formula. Supposing f is the frequency variable, $P(f)$ the relative power spectral density, and f_s the sampling frequency, we obtain spectral entropy as:

$$SE = -\sum_{f=0}^{\frac{f_s}{2}} P(f) \log_2(P(f)) \tag{2.17}$$

Singular Value Decomposition Entropy

Similar to Spectral Entropy, Singular Value Decomposition Entropy (SVDEn) transforms the data before computing Shannon's Entropy. Singular Value Decomposition (SVD) is a classical matrix factorization method developed in the late 18th century by a number of mathematicians. The idea to apply an entropy calculation to it was first introduced by Alter et al. [2] in the context of genome-wide expression data. SVD factorizes a matrix A into three matrices:

$$A = U\Sigma V^*$$ (2.18)

Matrices U and V^* represent the the left and right singular vectors in column and row space respectively. These vectors constitute orthonormal bases. Matrix Σ is a non-negative diagonal matrix containing the singular values associated to the singular vectors of the other two matrices. It is possible to think of each singular vector as a mode or component of the data. The vectors in U capture patterns in the column dimension whereas vectors in V^* capture them in the row dimension of the original data. Thus, a specific singular value of Σ is linked to two vectors, one in row-space and the other in column-space, serving as a coupling between these two dimensions. Moreover, the singular values can also be interpreted as the variance, energy, or even importance of the modes linked to it.

To compute SVDEn the following steps are carried out:

- **Produce a time-delay embedding** as done in Equation 2.12. Thus, from a single time-series we obtain a matrix where the row and column dimensions correspond to the $N - (m-1)\tau$ delayed time series samples and the m embedding dimensions respectively.

- **Apply the Singular Value Decomposition** to the time-delay embedding.

- **Normalize the singular values**. Assuming K singular values calculated, divide each by the total sum:

$$\overline{\sigma_i} = \frac{\sigma_i}{\sum_{i=1}^{K} \sigma_i}$$ (2.19)

- **Compute Shannon's Entropy** over the normalized singular values to obtain the desired SVD Entropy:

$$SVDEn = -\sum_{i=1}^{K} \overline{\sigma_i} \log_2(\overline{\sigma_i})$$ (2.20)

SVDEn captures the amount to which each mode found by the SVD reflects an equal amount of information [78]. Notice that it incorporates a similar idea as Spectral Entropy, but instead of acting over a power spectrum across frequencies, it acts over a "Singular Value Spectrum" across the different modes. Some sources use a Euclidean normalization: $\overline{\sigma_i} = \frac{\sigma_i^2}{\sum_{i=1}^{K} \sigma_i^2}$.

Multiscale Entropy

Costa et al. [11] introduced Multi-Scale Entropy (MSE) to address the issue of conventional entropy measures assigning high complexity values to processes with random outputs, in contrast to those exhibiting long-range correlations, such as the ones stemming from biological systems. The authors' argument rests on the fact that by not considering multiple time scales, the rich complexity of long-range correlations is lost. Thus, MSE proposes a method to account for multiple time-scales [12]. The main idea of the procedure is to calculate an Entropy value for each time scale, and from them make a profile curve. This curve is then summarized into an index by finding the area under the curve. More formally, starting from a one-dimensional discrete time-series $\{x_1, \cdots, x_i, \cdots, x_N\}$, the embedding dimension m, and tolerance r:

- **Choose a range of consecutive scales to explore**. One choice is to use a range from 1 to $\frac{N}{(m+10)}$ or to $\frac{N}{2}$.

- **Make the coarse-grained signal for each scale**. Mathematically we construct the coarse-grained signal $\{y^{(\tau)}\}$ for scale factor τ following the equation:

$$y_j^{(\tau)} = \frac{1}{\tau} \sum_{i=(j-1)\tau+1}^{j\tau} x_i \ , \text{ for } j = 1, 2, \cdots, \frac{N}{\tau} \tag{2.21}$$

- **Calculate entropies** for each time scale explored. The choice of entropy measure S_E is open. But in general it will depend on the embedding dimension m and tolerance r. For example, one could use SampEn applied to each $\{y^{(\tau)}\}$:

$$S_E(\tau) = SampEn\left(m, r, \frac{N}{\tau}\right) \tag{2.22}$$

- **Produce the profile** of S_E against τ, in other words, the $S_E(\tau)$ curve.

- **Reduce the profile to a single index** if needed. For example, one may use the area under the curve:

$$MSE = \text{Area Under the Curve of } S_E(\tau) \tag{2.23}$$

2.3.4 Lempel Ziv Complexity

First introduced by Lempel and Ziv [45], this feature measures the number of different substrings in a sequence when it is parsed from start to finish. Lempel-Ziv Complexity (LZC) has been associated to "Edge of Chaos" criticality [84], which reflects critical points that are between a stable phase and a chaotic one. As it may be apparent, LZC is related to the lossless compressibility of a sequence using a dictionary. It is thus, a way to apply the concept of Solomonoff-Kolmogorov-Chaitin Complexity to signals [9, 41, 77]. One of the interpretations is that it reflects the complexity, randomness, or repetitiveness of a finite sequence. Procedurally, LZC is calculated as follows:

- **Initialize** an empty dictionary and an auxiliar empty substring (AUXSTR).

- **Parse the sequence** from left to right, continuously adding new elements to AUXSTR. Whenever you encounter an unseen substring, add it to the dictionary, clear AUXSTR, and advance to the end of the current substring to continue the procedure.

- **Count** the number of substrings in the dictionary once you have reached the end of the string. This is the number of unique substrings encountered during the parsing.

Naturally, this variant of LZC is highly dependent on the sequence length as longer sequences will tend to have higher counts. To deal with this, Zhang et al. [96] proposed a normalized Lempel-Ziv Complexity (LZCn):

$$LZCn = \frac{LZC}{(n/log_b n)} \tag{2.24}$$

where LZC is the raw count of unique substrings as defined by the procedure, n is the sequence length, and b is the size of the alphabet (number of unique characters in the sequence).

Notice that Lempel Ziv Complexity is more appropriate to sequences with a discrete alphabet. Because of this, LZC is usually applied to a binarized version of an EEG time-series, e.g. through median clipping as shown in the following function applied to each element of the sequence **x**:

$$f(x_i) = \begin{cases} 1 & \text{if } x_i > \text{median}(\mathbf{x}) \\ 0 & \text{otherwise} \end{cases} \tag{2.25}$$

One may consider LZC as a form of entropy, as it also reflects the amount of information inside a sequence. In particular, LZC does not rely on the calculations of probabilities as many of the entropies do.

2.3.5   Number of Zero Crossings

Given a signal, count the number of times it crosses zero, or alternatively, the number of sign changes. Despite its simplicity, the Number of Zero Crossings (NZC) is widely used in seizure detection tasks [17, 70, 94, 95]. Naturally, to make the feature less dependent on signal length, one may normalize it. One of the common options for this is to divide the count by the number of samples in the signal. In this context, the NZC can be interpreted as the rate of zero crossings per sample (Zero Crossing Rate, ZCR).

2.3.6   Hjorth Parameters

Proposed by Hjorth [32] in the 70s, these parameters characterize the signal through statistical means, retaining information about its trace. These features were originally conceived in the context of EEG analysis, and correspondingly, have been widely-used in neuroscience, including seizure detection and lateralization tasks [8, 40, 62, 83]. Two of the parameters, **mobility** and **complexity**, were used in this work, given that the **activity** is essentially equivalent to the variance of the signal.

- **Hjorth Mobility** (HMob) reflects the mean rate of change of the signal, which is related to its frequency content. It is usually thought of as the "mean frequency" of the signal. Correspondingly, a high mobility suggests a rough signal whereas a low mobility indicates a smoother one. Given the signal $y(t)$, Hjorth Mobility is calculated as follows:

$$Mobility = \sqrt{\frac{var\left(\frac{dy(t)}{dt}\right)}{var(y(t))}} \tag{2.26}$$

- **Hjorth Complexity** (HComp) can be considered as a way to estimate the frequency bandwidth of the signal. More irregular and unpredictable signals will have a higher complexity, and viceversa. In particular, the lowest value of Hjorth complexity is 1, indicating that the signal is a pure sinusoid. Given the signal $y(t)$, Hjorth Complexity is defined as:

$$Complexity = \frac{\text{Mobility}\left(\frac{dy(t)}{dt}\right)}{\text{Mobility}(y(t))} \qquad (2.27)$$

2.3.7 Fractal Dimensions

Mandelbrot [53] coined the term fractal dimension (FD) motivated by the current inability to characterize intricate structures appearing in nature, such as coastlines, mountain landscapes, lighting bolts, etc. These structures would usually be described as rough in contrast to more regular euclidean shapes. A dimension is usually defined as the number of coordinates to specify a point in a space. Nevertheless, in the context of rough shapes a dimension is defined in terms of what happens to a measure of size when you change the scale of the unit of measure. To illustrate this, imagine a square, then suppose the side of the square doubles. How many of the original squares fit on the larger square? In this case is 4. The fractal dimension is extracted from this procedure following $N = \epsilon^D$, where N is the number of the original squares inside the scaled one (here 4), ϵ the scale applied (in this case 2), and D the fractal dimension. Naturally, $4 = 2^D$ and solving for D we get that the fractal dimension of a square is 2, consistent with what is expected of the euclidean definition. This same procedure is applied to rough shapes, allowing for D to be a fraction, hence the term "fractal dimension". For example, the Koch snowflake, one of the earliest fractals described, has a fractal dimension of approximately 1.262. The value itself can be interpreted as a measure of complex detail in a pattern, in particular, higher values indicate higher complexity, irregularity, more intricate details, more space-filling capacity, less predictability of the pattern, etc. It is possible to apply these concepts to the analysis of time series, as has been done for physiological and financial signals [69, 74]. Three measures of fractal dimension will be explored:

Katz Fractal Dimension

Katz [39] devised a procedure in the late 80s for the analysis and comparison of complex waveforms (such as EEGs). For a planar curve, the fractal dimension is generally the ratio:

$$D = \frac{log_{10}(L)}{log_{10}(d)} \qquad (2.28)$$

where d is the planar extent of the curve and L the total length. Here L is simply the sum of euclidean distances between successive points. The planar extent is the farthest

distance between the first point and any other.

Fractal characterizations assume that the pattern is formed by similarly-sized discrete units. For Katz's procedure, the relative unit of measure will be the "average step" a, i.e. the average distance between successive points. Considering that the total number of steps is $n = L/a$, the full formula of the Katz fractal dimension (KFD) becomes:

$$KFD = \frac{log_{10}(L/a)}{log_{10}(d/a)} = \frac{log_{10}(n)}{log_{10}(n) + log_{10}(d/L)} \tag{2.29}$$

Petrosian Fractal Dimension

Petrosian [66] defined this measure for the characterization of epileptic EEG recordings, in particular the interictal, ictal and postictal stages. It leverages the idea of binarizing the sequence as done in Lempel-Ziv Complexity. Different ways of binarization were explored by the author. However, in this context, we use binarization through the "differential zero crossing" method. This means the sequence from which the fractal dimension is derived represents sign changes in the signal's derivative. Mathematically, the Petrosian fractal dimension (PFD) formula is as follows:

$$PFD = \frac{log_{10}(N)}{log_{10}(N) + log_{10}\left(\frac{N}{N+0.4N_\delta}\right)} \tag{2.30}$$

where N is the length of the sequence and N_δ is the number of sign changes in the signal's derivative.

Higuchi Fractal Dimension

Higuchi [31] introduced his non-parametric method in the late 80s as a way to present a more stable and precise metric. In particular, the author notes that for natural phenomena the fractal dimension may not be constant over all time scales, but it can be defined over a range in which the self-similarity property holds. Thus, Higuchi's method contemplates finding the fractal dimension D by evaluating it over different time-scales explored until a final scale k_{max}. Consider a time-series $X(1), X(2), \ldots, X(N)$. Higuchi's method is defined procedurally as:

- **Generate new time-series** exploring different scales (or interval times) at different starting points. That is, explore interval times $k = 1, 2, \ldots, k_{max}$ from the starting points $m = 1, 2, 3, \ldots, k$:

$$X_m^k : X(m), X(m+k), X(m+2k), \ldots, X\left(m + \lfloor \frac{N-m}{k} \rfloor \cdot k\right) \tag{2.31}$$

where $\lfloor \ \rfloor$ denotes the floor function. Notice that the greater the scale, the more starting points are explored. Thus, each k time-scale will have its own k starting points and time-series, constituting a sub-sampling of the original time series by taking each k-point of it.

- **Calculate the length of the curve** for each X_m^k. This value will be:

$$L_m(k) = \left(\text{Sum of distances between consecutive points of } X_m^k\right) \cdot \frac{\eta}{k}$$

$$L_m(k) = \left(\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} |X(m+ik) - X(m+(i-1)k)|\right) \cdot \frac{\eta}{k} \tag{2.32}$$

where $\eta = \frac{N-1}{\lfloor \frac{N-m}{k} \rfloor k}$ is a normalization factor for the curve length.

- **Average the lengths for each scale**. For each time-scale (or interval time) k calculate the average across the different starting points explored:

$$\langle L(k) \rangle = \frac{\sum_{m=1}^{k} L_m(k)}{k} \tag{2.33}$$

- **Find the slope of the log-log plot**. Specifically, the fractal dimension D exists if $\langle L(k) \rangle \propto k^{-D}$. Make a log-log plot of $\langle L(k) \rangle$ against k and apply a linear regression to obtain the slope. Its absolute value represents the Higuchi Fractal Dimension (HFD).

2.4   Machine Learning

2.4.1   Binary Logistic Regression

Binary Logistic Regression is a type of Generalized Linear Model (GLM); these are characterized by three elements, the dependent and independent variables (Y, X), a linear predictor (e.g $\eta = X\beta$), and a link function that relates the distribution of the response variable to the input the variables (e.g $E(Y|X) = f(X\beta)$). The particularities of the Binary Logistic Regression are as follows:

- The dependent variable follows a Binomial distribution, usually taking values of either 0 or 1.

- As others GLMs, the predictor is a linear combination of the independent variables:

$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n$$

- The link function corresponds to the logistic function:

$$\sigma(\eta) = \frac{e^\eta}{1 + e^\eta} = \frac{1}{1 + e^{-\eta}}$$

The logistic function is the inverse of the logit, also known as the log-odds:

$$\text{logit}(p) = ln\left(\frac{p}{1 - p}\right)$$

The reason it is known as the log-odds is because the function outputs the natural logarithm of the *odds*, which is the ratio between the probability of an event occurring to the probability of that event not occurring, provided that input is between 0 and 1.

The output of the model corresponds to the probability of the response variable Y equating a success case rather than a failure case. In the context of machine learning this corresponds to the probability of belonging to the positive class. Naturally, the output is between 0 and 1, and the final class prediction can be obtained by thresholding this value. The full input-output formula is:

$$p(X) = \sigma(\eta(X)) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n)}} \tag{2.34}$$

2.4.2 Training

Training the Logistic Regression model is formally the process of optimizing the following objective function:

$$\min_w C \sum_{i=1}^n \left[-y_i log(p(X_i)) - (1 - y_i)log(1 - p(X_i))\right] + r(w) \tag{2.35}$$

where $r(w)$ is the regularization term, and C is the inverse of regularization strength. X and y correspond to the feature input vector and the binary class variable respectively.

2.4.3 Subject-wise Cross-validation

Cross-validation is a statistical technique used in machine learning to gauge the predictive performance of models on unseen data. Typically, the dataset is divided into a training set, where the model learns patterns, and a test set, which evaluates its predictive capabilities.

Expanding on this, k-fold cross-validation systematically divides the dataset into k equally sized folds or subsets. Here, the model undergoes k training iterations, with each iteration using $k-1$ folds for training and a unique fold for testing. The model's final performance is the average of its scores across the k test folds. This rigorous approach ensures every data point's involvement in both training and testing, providing a robust performance assessment.

In clinical scenarios, where data characteristics can differ dramatically between subjects, the approach of subject-wise cross-validation is especially relevant. Sometimes referred to as leave-p-subjects-out validation, this method trains the model on data from all subjects minus p, subsequently testing its predictions on those p omitted subjects. This strategy ensures the model's capacity to generalize over diverse individuals, upholding its dependability in real-world clinical situations where patient variability is a given.

2.4.4 Hyperparameter Optimization

In machine learning, hyperparameter optimization or tuning refers to the systematic process of finding the best combination of model settings (hyperparameters) that maximizes performance. Unlike model parameters, which are learned directly from the training data (e.g., coefficients in a linear regression), hyperparameters are external configurations to the model and cannot be estimated from the data. Notably, hyperparameters can significantly influence model outcomes. To ensure the optimal configuration is robust and avoids overfitting to the training data, it's imperative to use a separate validation set for this purpose rather than the training or test sets. By doing so, the selected hyperparameters are more likely to generalize well to new, unseen data. For logistic regression, the following hyperparameters can be tuned:

- Penalty: In logistic regression, the penalty parameter determines the type of regularization introduced to the model. Regularization introduces a penalty to the optimization process to avoid overfitting, aiding model generalization. L1 regularization applies a penalty equivalent to the absolute magnitude of coefficients. L2

regularization uses a penalty proportional to the squared magnitude of coefficients, ensuring they remain small. One could use a hybrid approach, incorporating both L1 and L2 penalties or no regularization at all.

- Formulation: Determines whether to approach logistic regression from a primal or dual perspective. The primal form directly optimizes the model's weights by minimizing the logistic loss, typically with a regularization term. The dual form, stemming from the Lagrangian duality principle, offers an alternative viewpoint by optimizing over Lagrange multipliers associated with the problem's constraints. While the primal form is more commonly employed, the dual can be advantageous in specific contexts, especially when the number of features surpasses the number of samples.

- Tolerance: Specifies the threshold for the stopping criteria during the optimization process in logistic regression. When the improvement in the model's objective function between iterations is less than this set tolerance, the algorithm deems the optimization as converged and halts further iterations. A smaller tolerance value demands a more precise solution, potentially requiring more iterations, while a larger value might result in a quicker but potentially less accurate convergence.

- C: Is the inverse of regularization strength. It plays a crucial role in determining the extent of regularization applied to the model. A smaller value of C indicates stronger regularization, leading to more significant constraint on the model's weights, which can help in preventing overfitting. Conversely, a larger value implies weaker regularization, allowing the model to fit more closely to the training data.

- Decision Function Bias: indicates whether to include an intercept (or bias) in the logistic regression model. If included the model adds an intercept to the decision function, enabling more flexible fits.

- Class Weights: allows for the adjustment of importance assigned to different classes during model training. By manipulating these weights, one can counteract the potential biases arising from imbalanced class distributions in the dataset. Even if a dataset is balanced, there might be situations where certain classes are deemed more important or critical than others based on the problem's context.

- Solver: designates the optimization algorithm used. The choice of solver can influence the convergence speed, memory usage, and compatibility with certain penalty terms or problem types. Experimenting with different solvers allows for fine-tuning the model's efficiency and effectiveness based on the dataset's characteristics and the problem's requirements.

- Maximum Solver Iterations: sets the maximum number of iterations the optimization algorithm will run before halting. The purpose of this parameter is to prevent endless running in cases where the solver struggles to converge.

2.4.5   Confusion Matrix and Performance Metrics

The confusion matrix offers a comprehensive view of a classifier's performance, going beyond simple accuracy to provide insights into types of errors made. It is especially useful when classes are imbalanced or when the costs of different types of errors vary. Considering a binary classification problem, the matrix is structured as shown in Table 2.1.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Table 2.1: A typical binary confusion matrix.

From the confusion matrix, several insightful metrics can be derived:

- **True Positive Rate (TPR)** or **Sensitivity**: Measures the proportion of actual positives correctly identified. Useful to know the model's capability to detect the positive class.

$$TPR = \frac{TP}{TP + FN} \tag{2.36}$$

- **True Negative Rate (TNR)** or **Specificity**: Indicates the proportion of actual negatives correctly identified.

$$TNR = \frac{TN}{TN + FP} \tag{2.37}$$

- **False Negative Rate (FNR)**: Measures the proportion of actual positives that got predicted as negative. A lower FNR is desired when the cost of missing a positive instance is high.

$$FNR = \frac{FN}{FN + TP} \tag{2.38}$$

- **False Positive Rate (FPR)**: Indicates the proportion of actual negatives that got predicted as positive. Essential when the cost of a false positive is high.

$$FPR = \frac{FP}{FP + TN} \tag{2.39}$$

- **Positive Predictive Value (PPV)** or **Precision**: Represents the proportion of positive identifications that were actually correct. Important when the consequences of falsely claiming a positive are severe.

$$PPV = \frac{TP}{TP + FP} \tag{2.40}$$

- **False Discovery Rate (FDR)**: Denotes the proportion of positive identifications that were false. A counter to Precision, it's critical when ensuring the purity of positive identifications.

$$FDR = \frac{FP}{FP + TP} \tag{2.41}$$

- **Accuracy (ACC)**: Reflects the proportion of total predictions that were correct. Though commonly used, it might be misleading on imbalanced datasets.

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{2.42}$$

When one normalizes the confusion matrix, most of the aforementioned metrics appear naturally on it. Confusion matrices can be normalized in two primary ways: by row or by column. Row normalization involves dividing each value in a row by the sum of that row's values, effectively giving the proportion of correctly (or incorrectly) predicted instances per predicted class. Column normalization, on the other hand, divides each value in a column by the sum of that column's values, presenting the proportion of correctly (or incorrectly) predicted instances per "actual" class. Normalizing the confusion matrix provides clearer insights into classification performance, especially in imbalanced datasets. By presenting values as proportions or percentages, researchers can more

readily discern patterns, compare classifier performances across different datasets, and identify specific areas where the model may require improvement.

Beyond the individual metrics, there exist complementary relationships between several of the metrics derived from the binary confusion matrix. These relationships are succinctly described by the following equations:

- **Specificity and False Positive Rate**: These metrics offer complementary insights into the classifier's performance concerning actual negative instances. Their relationship is defined as:

$$TNR = 1 - FPR \tag{2.43}$$

- **Positive Predictive Value and False Discovery Rate**: These two metrics provide complementary perspectives on the positive predictions made by the classifier:

$$FDR = 1 - PPV \tag{2.44}$$

- **Sensitivity and False Negative Rate**: These are direct complements, detailing the classifier's performance on actual positive instances:

$$FNR = 1 - TPR \tag{2.45}$$

2.4.6 Receiver Operating Characteristic

The ROC (Receiver Operating Characteristic) curve is a graphical representation that showcases the performance of a binary classification model across all possible threshold settings. It is obtained by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR) as the decision threshold for classification is varied. The area under the ROC curve, known as the AUC (Area Under the Curve), provides a single value metric that summarizes the overall performance of the classifier, with a value of 1.0 representing perfect classification and 0.5 indicating performance no better than random chance. By analyzing the ROC curve and its AUC, one can select an optimal threshold that balances sensitivity and specificity based on the problem's objectives, ensuring that the model performs well in differentiating between the classes. This makes the ROC curve an essential tool in situations where the classification threshold needs fine-tuning.

2.4.7   Minimum Redundancy Maximum Relevance Algorithm

The Minimum Redundancy Maximum Relevance algorithm (mRMR, [65]) is a feature selection technique that aims to identify an optimal subset of features that maximizes their mutual information with the target while minimizing redundancy among the features themselves. A distinct advantage of mRMR is its ability to determine the best number of uncorrelated features without the user specifying the desired number of features.

To achieve this, mRMR is implemented by using the SULOV method and a Recursive XGBoost stage. SULOV, which stands for "Searching for Uncorrelated List of Variables," operates by identifying pairs of highly correlated variables surpassing a specific threshold (e.g., absolute correlation of 0.7). For each pair, the variable with a lower Mutual Information Score (MIS) with the target is removed, ensuring that the remaining features have high information scores and low inter-feature correlations.

Following SULOV's feature pruning, Recursive XGBoost is employed to refine the feature selection further. By dividing the data into training and validation sets, XGBoost iteratively identifies the top features based on their significance to model performance. This process is repeated multiple times, with each iteration focusing on the remaining set of features post-SULOV. The culmination of these steps results in a feature set that embodies both maximum relevance and minimum redundancy.

Chapter 3

Materials and Methods

3.1 Datasets

3.1.1 Cleveland Clinic Dataset

Working under an Institutional Review Board approved protocol at the Cleveland Clinic, 41 pharmacoresistant focal epilepsy patients underwent presurgical evaluation, stereotactic-EEG (SEEG) implantation and its subsequent inspection. SEEG signals were recorded on a Nihon Kohden EEG system with a sampling rate of 500Hz (until 2012, 7 patients) or 1000Hz (2012 and later, 34 patients). The patients had surgery consisting of resection or laser ablation of the identified epileptogenic zone. 28 of the patients were seizure-free after the procedure (Table 3.1). Anatomical locations of the electrode leads and their relation to the resected/ablated region are available through the alignment of a post-implantation CT image to a preoperative and postoperative MR image respectively. Table A.1 on Appendix A shows the clinical profiles of these patients.

Each patient in the dataset has the outcome of the resection (Seizure-Free or Non Seizure-Free), and one label per electrode contact that indicates whether it was resected or not. Moreover, each patient has a series of seizures where both ictal and interictal periods of activity were recorded. These periods are of 40 seconds each, except for some cases where more data (60 seconds) or less (20 seconds) is available. Interictal segments were captured at least 2 minutes before the seizure onset, and each is paired to a corresponding ictal record. Each SEEG record corresponds to one of these periods, and it will contain the time-series for all of the SEEG contacts. The total number of contacts as well as their exact locations vary from subject to subject as the stereotactic implantation differs for each. Table 3.1 shows the data available per subject and its

properties regarding the contacts, resection and outcome. Each SEEG records is a 2D matrix of contacs × samples (time), as can be seen in Figure 3.1.

Table 3.1: Data available per subject with resection details.

Subject	Num. Interictal Records	Num. Ictal Records	Outcome	Num. Contacts	% of Resected Contacts
S001	3	3	Seizure-Free	88	12.0
S002	3	3	Seizure-Free	111	54.0
S003	3	3	Seizure-Free	79	28.0
S004	3	3	Seizure-Free	39	8.0
S005	3	3	Seizure-Free	30	17.0
S006	3	3	Seizure-Free	55	36.0
S007	3	3	Seizure-Free	42	45.0
S008	3	3	Seizure-Free	61	18.0
S009	3	3	Seizure-Free	57	46.0
S010	3	3	Seizure-Free	67	46.0
S011	3	3	Seizure-Free	82	16.0
S012	3	3	Seizure-Free	60	55.0
S013	3	3	Seizure-Free	41	63.0
S014	3	3	Seizure-Free	70	14.0
S015	3	3	Seizure-Free	100	32.0
S016	3	3	Seizure-Free	139	17.0
S017	3	3	Seizure-Free	92	38.0
S101	2	2	Seizure-Free	92	43.5
S102	1	1	Seizure-Free	87	36.0
S103	3	3	Seizure-Free	115	29.0
S106	11	11	Seizure-Free	116	16.0
S108	6	6	Seizure-Free	65	23.3
S111	11	11	Seizure-Free	121	37.1
S112	5	5	Seizure-Free	83	12.0
S113	2	2	Seizure-Free	147	19.5
S116	8	8	Seizure-Free	139	35.0
S118	2	2	Seizure-Free	91	19.0

Table 3.1: Data available per subject with resection details.

Subject	Num. Interictal Records	Num. Ictal Records	Outcome	Num. Contacts	% of Resected Contacts
S140	2	2	Seizure-Free	96	21.5
S215	2	2	Non Seizure-Free	109	4.0
S219	11	11	Non Seizure-Free	88	10.0
S220	9	9	Non Seizure-Free	128	2.0
S221	7	7	Non Seizure-Free	122	10.0
S222	7	7	Non Seizure-Free	117	8.9
S223	3	3	Non Seizure-Free	156	21.0
S226	5	5	Non Seizure-Free	159	15.2
S228	4	4	Non Seizure-Free	75	33.0
S231	3	3	Non Seizure-Free	155	5.0
S232	2	2	Non Seizure-Free	121	25.0
S233	4	4	Non Seizure-Free	157	36.0
S237	3	3	Non Seizure-Free	148	7.0
S238	6	6	Non Seizure-Free	86	9.8

An additional dataset is provided below for minor experiments. Within the context of this project, the Cleveland Clinic Dataset is the primary dataset. Hence, any reference to "the dataset" will pertain specifically to the Cleveland Clinic Dataset.

3.1.2   Bern-Barcelona Dataset

This dataset has acted as a benchmark in the literature for the task of epileptic focus detection based on interictal data. It includes 7500 samples of focal and non-focal signals, each of 20 seconds and sampled at 512 Hz. Focal signals were defined as those that exhibited the first ictal signal changes judged visually by at least two expert neurologists. The signals were recorded from both intracranial strip and depth electrodes. The dataset was collected from five long-standing pharmacoresistant temporal lobe epilepsy patients prior to their resective surgery, all having good surgical outcomes. This dataset is balanced in terms of the classes but has some important drawbacks such as a lack of information about electrode locations and patient provenance. Given that this dataset

only includes interictal data, it was only used for a few experiments to obtain performance metrics comparable to those in the literature. Any mention of this dataset will be done explicitly as the "Bern-Barcelona" dataset.
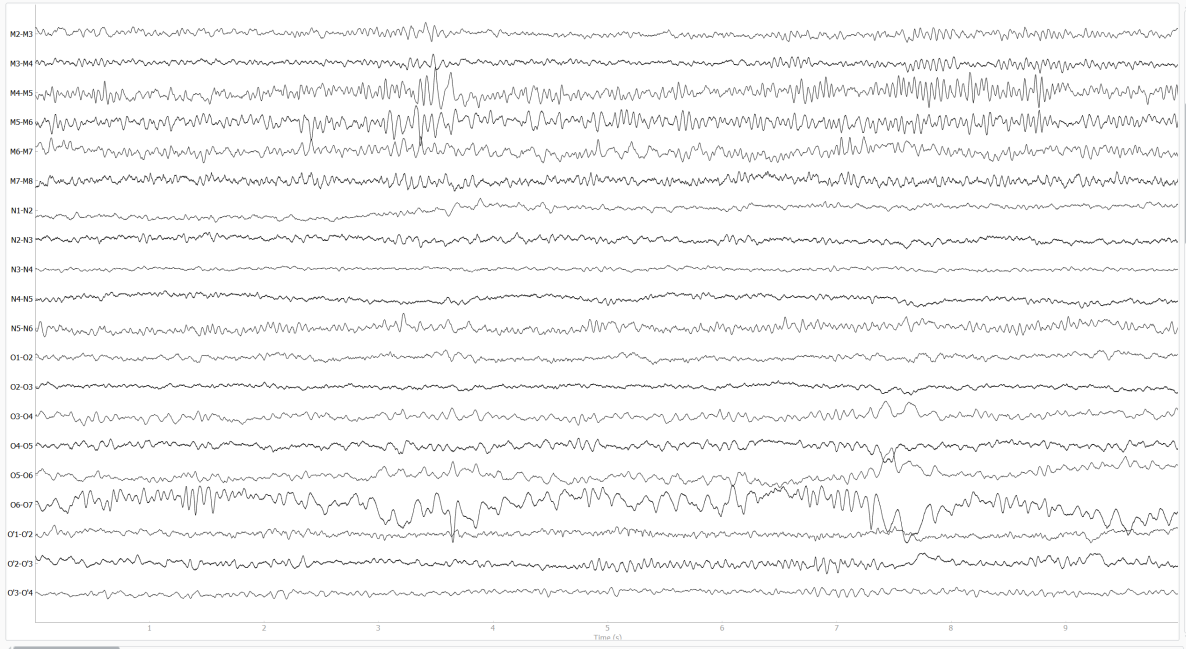
Figure 3.1: Illustration of the content of a SEEG record as visualized on the MNE-Python software.

3.2   Preprocessing

Data was imported and manipulated within Python, in particular using the MNE-Python software [23]. Each SEEG record was band-passed between 0.5Hz to 150Hz, and then resampled to 256Hz (with an antialiasing lowpass filter of 128Hz). Subsequently, the data was normalized by applying a z-score to the amplitudes of each contact. This allowed a degree of robustness towards record-wise and contact-wise amplitude differences in the signals which could affect the process of feature computation. Afterwards, the data was segmented into 1 second epochs with 0.5 seconds of overlap.

3.3   Feature Extraction and Normalization

The previous epoch segmentation was done to extract multiple instances of a feature per SEEG channel rather than a single one, which allows to extract statistics (e.g. mean

and standard deviation) from each SEEG channel. The idea was to decrease the bias in the measurement of the features, albeit involving an increase in variance. Another advantage of this windowed approach is that the features will be defined even if some of the calculations fail for a particular window (unless all of the windows fail which is highly improbable).

A set of 14 complexity features were computed using the MNE, Antropy, NeuroKit and FOOOF (specparam) python packages [16, 23, 51, 86]. These features are listed along the parameters used for their computation in Table 3.2. Note that for each of these features both the mean and the standard deviation (SD) were computed from the multitude of windows as explained in the previous paragraph, thus one could consider this to be actually a set of 28 features. A second level of normalization was applied given the variance of the ranges natural to each of them (e.g. some of the features won't necessarily be on the $[0, 1]$ range by default). This step consisted of applying a z-score across the contacts (i.e. the channels) of the corresponding SEEG record. That is, the output values signify the position (number of standard deviations) of a given contact in respect to the mean value of the feature across the channels. More importantly, this second level of normalization alleviated variations of the feature ranges that were relative to each subject.

3.4   Exploratory Analysis

3.4.1   Statistical Characterization

Ideally one would want to study the relationship between epileptogenicity and the complexity features. Nevertheless, this is not possible as the relevant information available is in regards to the resection area of the subjects that became seizure-free afterwards. Given this, we characterized the relationship between the resection area and the features across these Seizure-Free subjects. One possible way to do this would be to apply statistical tests on the means of each feature distributions split by resected and non-resected locations with data from all of the subjects. Nonetheless, it is also relevant to see these patterns on a subject-wise level. For example, it could happen that some subjects become over-represented when pooling data from all of them, as some subjects contain more contacts than others. Importantly the subject-wise level of detail shows whether a pattern that is observed at the overall level is indeed repeated across most of the subjects. To accomplish both the overall and subject-wise level of detail in a summarized

Feature	Abbreviation	Parameters	Package
Slope of the aperiodic component	1/f Slope	aperiodic_mode : knee freq_range : [5,100] with a previous Welch's spectrum calculation using: n_fft : 51 n_overlap : 26 n_per_seg : 51 window : hamming	FOOOF MNE
Approximate Entropy	ApEn	order : 2 metric : chebyshev	antropy
Sample Entropy	SampEn	order : 2 metric : chebyshev	antropy
Permutation Entropy	PE	order: 3 delay: 1 normalize: True	antropy
Spectral Entropy	SE	method : welch nperseg : 51 normalize : True	antropy
Singular Value Decomposition Entropy	SVDEn	order: 3 delay: 1 normalize: True	antropy
Multiscale Entropy	MSE	scale : default dimension : 3 tolerance : sd method : MSEn	NeuroKit
Lempel-Ziv Complexity	LZC	normalize : True with median clipping before	antropy
Zero Crossings Rate	ZCR	normalize : True	antropy
Hjorth Mobility	HMob	None	antropy
Hjorth Complexity	HComp	None	antropy
Katz Fractal Dimension	KFD	None	antropy
Petrosian Fractal Dimension	PFD	None	antropy
Higuchi Fractal Dimension	HFD	kmax : 10	antropy

Table 3.2: Features explored to build the complexity fingerprint. The parameters used for each as well as the packages used for the computation are detailed.

manner we characterized the relationship between the resection area and the features by obtaining t and p values on both levels. This was done using an unpaired Welch's t-test with unequal variances and unequal sample sizes [90]. As a matter of fact, the Resected and Non-Resected groups of contacts have different number of samples as usually the number of resected locations will be lower than its non-resected counterpart (as shown in Table 3.1).

Specifically, this t-test was applied to the Resected and Non-Resected groups of contacts from all of the subjects for all of the features. The statistic applied was $\frac{Mean(Resected) - Mean(Non\ Resected)}{Standard\ Error}$. Thus, the t-value will be positive when the mean of the resected contacts is greater than the mean of the Non-Resected contacts, and vice-versa. Similarly, we did the same procedure on a subject-by-subject basis, and then calculated the number of subjects per feature with positive and negative t values, and within each, the number of significant subjects (i.e. $p < 0.05$). In addition, the average t-value and p-value across subjects was computed. Note that these statistical quantities are only for exploration purposes, so they shouldn't be interpreted as showcasing rigorous statistical results. Indeed, to do so appropriately one would need to at least apply a multiple comparisons correction.

3.4.2   Feature Sets and Feature Selection

One of the interests of this project is to explore ictal and interictal features both jointly and in isolation. In this regard, only using interictal data poses the harshest condition, but also the most usable and promising one. To study the impact of using these different feature sets, three different machine learning models were developed:

1. Using only ictal features ("The Ictal Model").

2. Using only interictal features ("The Interictal Model")

3. Using both ictal and interictal periods ("The Hybrid Model").

To prevent feature redundancy a feature selection procedure was carried out. This was done through automatic means so that subjective judgements didn't affect the salient features of epileptogenicity found, which is the main result of this project. Feature selection was done through the Minimum Redundancy Maximum Relevance (mRMR) algorithm [65], as implemented in the Featurewiz python package [75]. This was done in isolation for ictal and interictal feature sets, using a value of 0.7 as the correlation limit

between features. The hybrid feature set is formed by taking the union of the selected feature sets for ictal and interictal conditions, each determined separately by the mRMR procedure.

3.4.3   Correlations to Spiking Activity

In the efforts for Epileptogenic Zone delineation, some biomarkers have become ubiquituos, such as High Frequency Oscillations and Interictal Epileptiform Discharges (IEDs) [35]. This last biomarker is seen as spikes in the signal. We are interested in finding biomarkers beyond these, yet the complexity features explored can be particularly sensible to IEDs. Following this, a simulation experiment was carried out to understand more deeply the relation between complexity features and spikes.

To this purpose, an interictal recording from one of the subjects was chosen based on its lack of spiking activity. Then, spikes were added to the recording at random positions. The spikes were modeled with a Gaussian window with a defined width (which would represent the abruptness of the spike) scaled by the maximum value of the signal. In addition, the standard deviation of the Gaussian function was defined as $\max(\text{width}//6, 1)$. The spikes are added to the signal by simply summing the resulting Gaussian window to the signal at appropriate samples. From this scheme, two dimensions of spiking activity were explored in regards to how they would impact each of the features:

- The width of the Gaussian window, explored from 3 to 41 samples with only odd numbers.

- The numbers of spikes in the recording, from 3 to 21 with only odd numbers.

Figure 3.2 illustrates the simulated signal spectrogram on the four border cases. After this procedure, the features are computed using the same epoch segmentation mentioned in section 3.2. Then, for each combination of width and number of spikes, we compute the features through each window of time and subsequently obtain the mean and standard deviation. Afterwards, the correlation between each feature and each spike parameter (width and number of spikes) is computed. Results are summarized in figures showcasing the strength of these correlations. As a complementary analysis, heatmaps for each feature were produced exploring the two spike parameters and from these the average gradient across each dimension was computed. Before calculating the gradients the data was normalized through a z-score to avoid seeing differences originating from the

natural range of each feature. These complementary analysis can be seen on Appendix B.
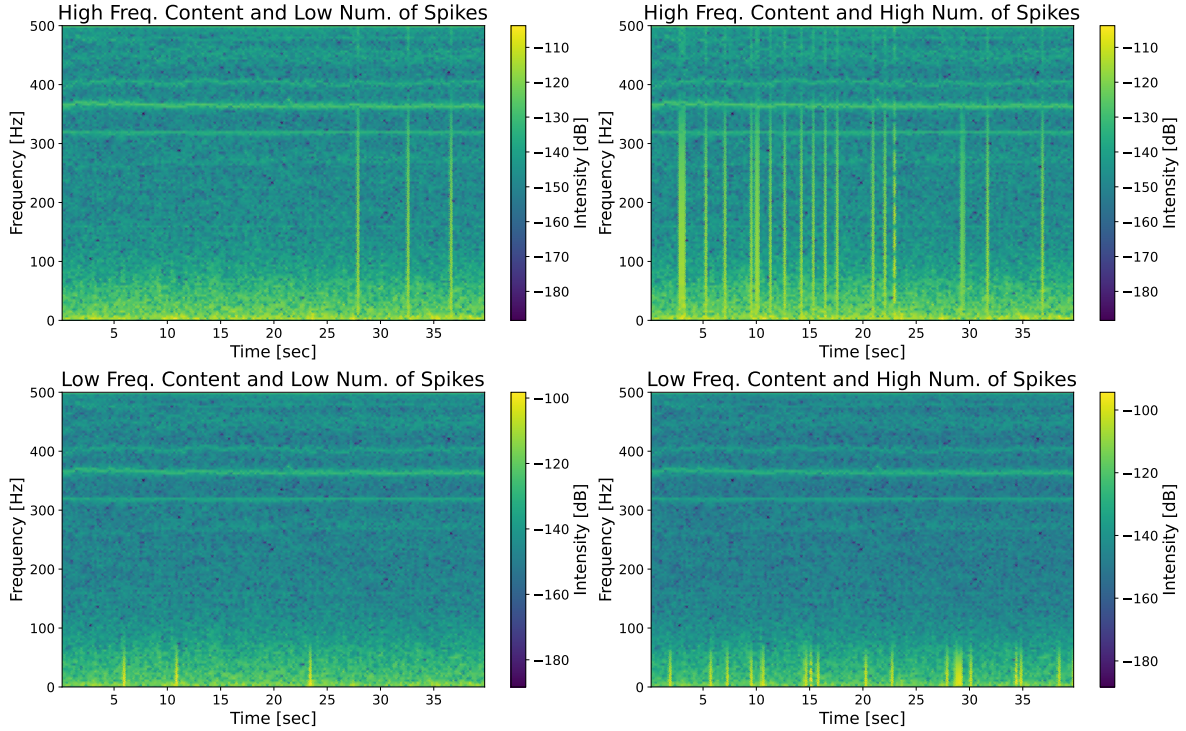
Figure 3.2: Spectrograms of the signals containing the simulated spikes on four border cases of the *Frequency Content* and *Number of Spikes* parameters.

3.5 Machine Learning

3.5.1 Task

The main goal of the machine learning component was the prediction of the epileptogenic electrodes, namely, classifying the electrodes into those inside and outside of the Epileptogenic Zone. On a naive look at the dataset this appears easily accomplishable through a supervised learning scheme. Nevertheless, this dataset only contains information in regards to the resection of the electrodes contacts (whether a particular contact was resected or not). In respect to epileptogenicity, what is known is that in those patients that are seizure-free the Epileptogenic Zone is a subset of the resected area. This means that some of the resected contacts are actually epileptogenic whereas some others are not, even though they were resected. On the other hand, we can be sure that the non-resected contacts are not part of the Epileptogenic Zone in the seizure-free patients.

In the case of non seizure-free subjects, even less information is known as some of the non-resected electrodes may be epileptogenic. The aforementioned conditions could be summarized as a "partially uncertain labelled data" problem [26]. For seizure-free subjects, the uncertainty only exists on the resected class. This situation complicates not only the training but also the evaluation of the model.

One possible way to approach this problem is to apply an unsupervised clustering within the uncertain class. If there is some intuition regarding the relation of the features to the uncertain class, then one could change the label or omit those instances that correspond to the cluster that is less prone to be the one reflecting the real underlying class. This was the approach taken by Grinenko et al. [26], where an intuition was available as the features were directly related to hypothesized biomarkers of the Epileptogenic Zone. In our case, there is a lack of such intuition and moreover the goal is to explore and find features that may then be interpreted as biomarkers, rather than from the clinical biomarkers design related features. In our work, a more simple approach is taken. The task proceeds as considering the uncertain class as ground truth, that is, all labels are initially thought to be correct and that we are predicting Epileptogenicity rather than the resection. The model is expected to naturally find meaningful patterns among the "real" epileptogenic electrodes and that it will predict as non-epileptogenic those "fake" ones that are inside the resection. The underlying assumption is that "real" Epileptogenic contacts are separable from the fake ones in the complexity-based feature space that is explored. And moreover, that the characteristics of the "fake" resected contacts are closer to those of the non-resected group than to the "real" -and unknown- epileptogenic set. The uncertainty is dealt with not at the training phase but rather in the interpretation and evaluation of the predicted labels. This is detailed on section 3.5.5. In summary, the model will be trained to predict the resection label using a supervised learning scheme, but its predictions will be interpreted in terms of epileptogenicity.

3.5.2   Model Architecture

A logistic regression model implemented in scikit-learn [64] was used. One of the advantages of such a model is its simplicity which allows for a straightforward interpretation of the feature importance and of its decisions. These characteristics are ideal in clinical contexts, specially for high-stakes decisions [73]. The input of the model are the features of a single contact. Different combinations of features were explored, stemming from the usage of both periods (ictal and interictal) or either one in isolation. When both periods are used, the features entering the model are the concatenation of the features from

each of the periods. Moreover, a probability aggregation mechanism is implemented to manage cases where more than a single EEG record is available per subject. The different records represent different samples of information originating from the same instance we want to classify (a single subject's contact). The probability aggregation mechanism consists of obtaining the probability assigned to the positive class in a single classified instance (a contact of a subject in a particular iEEG record). Then the probabilities from different instances (each classified separately by the model) are averaged and thresholded. If the average probability surpass the threshold (which is defined on the hyperparameter optimization stage), that particular contact is marked as Positive (meaning epileptogenic). Otherwise it is marked negative. An illustration of this probability aggregation mechanism is shown on Figure 3.3.
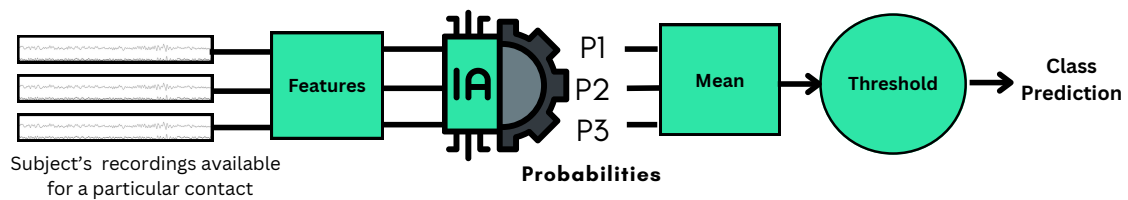
Figure 3.3: Probability Aggregation Mechanism. Put simply, each subject's contact has different recordings, but the final prediction is done per contact instead of per recording. To achieve this, the probabilities coming from individual recordings as they pass through the Logistic Regression model are averaged and subsequently thresholded to get a class prediction.

3.5.3   Data Balancing and Split

Data from non seizure-free subjects were discarded (S215 to S238, 13 subjects) as the resected electrodes don't fully contain the epileptogenic zone, thus preventing the ex-

traction of useful information from them. The other 28 subjects can be separated into two groups:

- Set A: 17 subjects (S001 to S017) correspond to the data used in [26].

- Set B: 11 subjects (S101 to S140) correspond to the data used in [46].

Set A was used for training and test using a subject-wise cross-validation strategy. Prior to the training of the model, Set B was used as a validation set, namely, the hyperparameters of the model were optimized using Set B.

It is important to note that for both Set A and Set B the data is imbalanced favoring the Negative (Not-Resected) class, as can be seen in Table 3.1. The degree of imbalance varies from subject to subject (min: 2%, max: 63%). Moreover, there is an imbalance regarding the number of contacts per subject (min: 30, max: 159) and the number of available records per subject (min:1, max: 11) which is shown in the same table. Thus, there is imbalance at the class and subject level. One way to approach this problem would be to randomly oversample the minority class on each subject. Nevertheless, the minority class is usually the Positive (Resected contacts) one, which is precisely the one with uncertainty. In this sense, randomly grabbing samples from the uncertain class could bias the dataset in unexpected ways in regards to the actual ground truth, that is, choosing inadvertely more contacts outside the real Epileptogenic Zone or viceversa. To keep the unknown but inherent distribution of "real and fake" Epileptogenic contacts within the Resected class we choose not to randomly oversample but rather copy the whole minority class samples within a subject as many times times as needed to make the number of samples of both classes as close as possible.

After this procedure, the number of samples per subject has changed in respect to the original dataset, so new subject-wise imbalances are encountered. To deal with this, a similar procedure is done but at the subject level. We identify the subject with the most samples and then oversample the whole data of the other subjects as much as it is needed to make their numbers as close to the ones of the most represented subject. Both balancing procedures are done before splitting the whole seizure-free dataset into the sets A and B.

3.5.4   Hyperparameter Optimization

The training of the logistic regression model depends on certain hyperparameters, which are listed in Table 3.3 along with the corresponding ranges that were explored for each

of them. Some hyperparameters were not varied, instead being defined with a specific value. This decision was made to save time during the training phase of the project. These hyperparameters have a single value within their defined range. For the sake of completeness, the unexplored range is also displayed in Table 3.3.

As mentioned in section 3.5.3, set B was used for the optimization. To explore the hyperparameter space, a simple grid search procedure was carried out, albeit incorporating a subject-wise cross-validation strategy. Concretely, each combination of hyperparameters is used to train the model across 11 folds, each leaving one of the subjects out. Afterwards, the mean performance metric associated to that particular combination of hyperparameters is computed. The chosen metric for this procedure was the Epileptogenic Predictive Value (equivalent to PPV, as explained in 3.5.5). The result of this is a table of hyperparameters combinations each linked to a EPV value, and from this, the combination with the greatest EPV is identified and used to train the model.

Although not strictly considered a hyperparameter, the threshold used in the final model was chosen at this stage. Basically, after selecting a set of hyperparameters, the model is trained using the same subject-wise cross-validation strategy on the validations subjects (11 folds). **Thus the threshold is not selected based on the performance of the training set**. The predicted probabilities on each of the test sets of the validation folds were collected and used to generate a ROC curve. On this ROC curve two points were identified:

- Point with the highest TPR constrained to a FPR below a certain percentage (chosen to be 5%). This point is of interest given that the priority is that the contacts marked epileptogenic reside inside the resection zone.

- Points nearest to the top-left corner (TPR=1, FPR=0). Which is the usual operating point considered on binary classification models. Nevertheless, here is included only for completeness' sake.

3.5.5 Training & Evaluation

As mentioned before, training was done using a subject-wise cross-validation approach. These folds leave one subject out at a time. Thus for training on Set A, 17 folds are used. Before training, the data is shuffled using a random number generator which was initialized with a seed (0). After fitting the model, predictions are obtained for each subject's contact by aggregating them using the probability aggregation mechanism

Hypeparameter	Explored Range	Unexplored Range
Norm of the Penalty	[l1,l2]	None, elasticnet
Formulation	Primal	Dual
Tolerance for stopping criteria	[1e-3,1e-2]	...
Inverse of regularization strength	[0.001,0.01, 0.1, 1]	. . .
Decision Function Bias	with	without
Class weight	Combinations of : Negative Class=[1] Positive Class=[0.25,0.5,0.75,1]	. . .
Solver	saga	[lbfgs,liblinear,newton-cg,newton-cholesky,sag]
Maximum number of iterations to wait for solver convergence	100	. . .

Table 3.3: Hyperparameters explored for the training of the linear regression model

detailed before. Naturally, these predictions correspond to the ones where each subject was the test set. From these aggregated predictions, confusion matrices are computed per subject. Afterwards, an overall confusion matrix is constructed by combining the matrices from all of the subjects. Both levels of detail are shown in the results.

As generally done, we use quantities derived from the confusion matrices to evaluate the model. Given the uncertainty inherent to the classes, the quantities inside of the confusion matrix were re-interpreted in the context of Epileptogenic Zone delineation. Indeed, if we were to interpret them in their usual sense, the fact that one of the classes is uncertain is loss. The re-interpretation can be seen on Table 3.4.

	Predicted Epileptogenic	Predicted Non-Epileptogenic
Inside Resection	EIR (TP)	NEIR (FN)
Outside Resection	EOR (FP)	NEOR (TN)

Table 3.4: Reinterpretation of the quantities inside the confusion matrix in the context of Epileptogenic Zone Delineation when only resection labels are available. EIR: (Predicted) Epileptogenic Inside Resection. NEIR: (Predicted) Non-Epileptogenic Inside Resection. EOR: (Predicted) Epileptogenic Outside Resection. NEOR: (Predicted) Non-Epileptogenic Outside Resection.

The ideal performance of the classifier would result in EOR to be empty, while NEOR would contain every non-resected contact. NEIR contacts would correspond to locations resected but that were not part of the (real) Epiletogenic Zone, and EIR would be interpreted as the (real) epileptogenic contacts. Nevertheless, neither EIR nor NEIR have a ground truth. The result we can be most confident about is EOR being as low as possible without recurring to label everything as Non-Epileptogenic.

From the confusion matrices, three performance metrics were chosen to carry out the evaluation of the models based -as before- on their interpretation given the uncertainty of the positive class. These metrics were calculated from the subject-by-subject confusion matrices as well as on the overall one. The chosen metrics were:

- Epileptogenic Predictive Value (EPV) which is a re-interpretation of the Positive Predictive Value. This quantity reflects our confidence on the fact that the contacts that were marked epileptogenic are at least inside the resection zone. Naturally, the desire is to maximize this quantity.

$$
\begin{aligned}
EPV &= \frac{TP}{TP + FP} = \frac{EIR}{EIR + EOR} \\
&= \frac{\text{Marked Epileptogenic Inside Resection}}{\text{All Marked Epileptogenic (Inside or Outside)}}
\end{aligned}
\tag{3.1}
$$

Note this corresponds to the complement of the False Discovery Rate (PPV=1-FDR). To discourage the model from marking none of the contacts as epileptogenic,

EPV is defined as 0 when this happens.

- False Epileptogenic Rate (FER), which is a re-interpretation of the False Positive Rate. This quantity reflects the percentage of contacts marked Epileptogenic which we know can't actually be so (as they were non-resected and the patients became seizure-free). Ideally we want to minimize this quantity.

$$FER = \frac{FP}{FP + TN} = \frac{EOR}{EOR + NEOR}$$
$$= \frac{\text{Marked Epileptogenic Outside Resection}}{\text{All Outside Resection}} \qquad (3.2)$$

Note this corresponds to the the complement of specificity (TNR=1-FPR).

- Resected Miss Rate (RMR), which is a re-interpretation of the False Negative Rate. This quantity reflects the amount of contacts resected that were marked Non-Epileptogenic. In a classical task were labels are completely certain, the ideal would be to minimize this quantity. Nevertheless, it is suspected that some resected contacts were Non-Epileptogenic. Thus, having a null FNR in this context would actually be both unexpected and undesirable.

$$RMR = \frac{FN}{FN + TP} = \frac{NEIR}{EIR + NEIR}$$
$$= \frac{\text{Marked Non-Epileptogenic Inside Resection}}{\text{All Inside Resection}} \qquad (3.3)$$

Note this corresponds to the complement of sensitivity (TPR=1-FNR).

Moreover, ROC curves are obtained for each model trained based on the unseen data results and the averaged probability of the positive class obtained through the probability aggregation mechanism.

3.5.6   Feature Importances

Apart from the performance evaluation, feature importance is characterized for each model. As we used a cross-validation strategy, multiple sets of models are obtained (and thus different weights for the same feature exist given the different folds). Because of this, boxplots are made from the distributions of the weights per feature.

3.5.7   Benchmark Experiment

The Bern-Barcelona dataset was used to have performance metrics comparable to other results in the literature as explained in section 3.1.2. Specifically, the interictal feature set was computed on this dataset and a single machine learning model was trained with a 5-fold cross-validation strategy. Notice that since no subject labels are given in this dataset it is not possible to do a subject-wise cross-validation procedure. The hyperparameters used for training were likewise obtained from Set B as mentioned in section 3.5.3. An overall confusion matrix is built from the unseen data in each fold along with the previously described performance metrics. The model resulting from this experiment will be referred to as "The Benchmark Model".

3.6   Methods Overview

To summarize, Figure 3.4 illustrates the complete methodology:
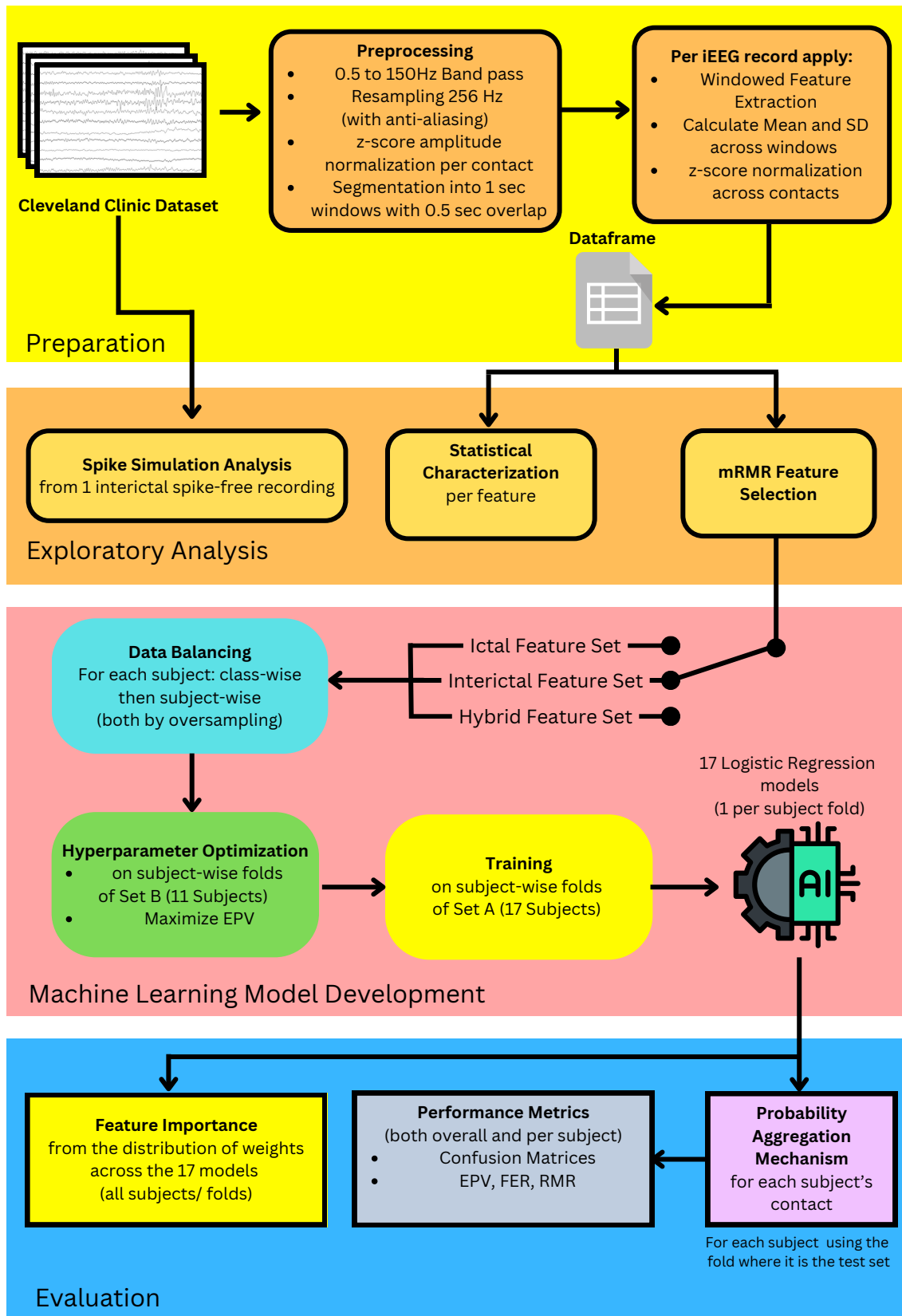
Figure 3.4: Overview of the Methodology. The switch of feature sets explores the three possibilities. For the Benchmark experiment, Set A is replaced by the data from the Bern-Barcelona Dataset and is only applicable to the interictal feature. Notice that Set B of 11 subjects (from the Cleveland Clinic Dataset) is still used in this case for hyperparameter optimization. Moreover, all folds from the Benchmark experiment grab data from its 5 subjects randomly, as no subject label is available for that dataset.

Chapter 4

Results

4.1 Results from the Exploratory Analysis

4.1.1 Statistical Characterization

The analysis of the data presented in Table 4.1, which is sorted by the magnitude of the average t-value across subjects, reveals a diverse mix of both positive and negative t-values, signifying the variability in how features differentiate the Resected from the Non-Resected groups during the ictal period. A closer examination shows that 12 mean-based features have negative t-values, indicating a lower metric in the Resected group relative to the Non-Resected group. Conversely, two mean-based features have positive t-values, suggesting that the majority of mean-based features tend to have more negative values within the resected area. In stark contrast, all 14 standard deviation features exhibit positive t-values, pointing towards greater variability in the Resected group.

Specific features, such as Approximate Entropy SD, stand out with higher t-values within the resection area, hinting at their potential discriminative power. However, others like ApEn Mean display lower values, suggesting a different trend. The t-values' absolute magnitude spans an expansive range, from 4 to 30, and similarly, the p-values range from 1×10^{-186} to 1×10^{-5}. This broad variability implies that while some features, like the mean Hjorth Mobility and the mean 1/f slope, may not be highly discriminative, others, particularly the standard deviation of Approximate Entropy and the Higuchi Fractal Dimension, markedly stand out. An intriguing observation is the prominence of standard deviations in the upper part of the table, which implies that variability within certain features might be pivotal in distinguishing the two groups during ictal states. Interestingly, the mean Multiscale Entropy secures a top position in the table.

Diving deeper into a subject-wise analysis, the data unveils notable variability across features. For instance, although the mean Multiscale Entropy predominantly has a negative t-value for most subjects, 4 out of the 28 subjects deviate from this pattern. The 1/f Slope's standard deviation emerges as a standout feature, consistently presenting a similar pattern across all subjects, underlining its significance. Yet, as we navigate towards the table's lower end, we encounter increasing ambivalence concerning the percentages of subjects with positive and negative trends, epitomized by the HFD Mean, which is split evenly at 50/50.

The analysis of the interictal features, as detailed in Table 4.2, reveals a few distinct patterns. When sorted by the magnitude of the average t-value across subjects, the table displays a clear dominance of features reflecting a mean-based value. This is in stark contrast to the ictal table, where standard deviations took precedence. Unlike the mixed directions of the mean-based t-values seen in the ictal table, the interictal table presents a more consistent direction, with the majority of t-values being negative, delineating a trend. Specifically, 13 out of 14 mean-based features show negative t-values, signifying a lower metric in the Resected group in comparison to the Non-Resected group, while only one mean-based feature leans towards a positive t-value. On the other hand, for the interictal period, only 9 out of the 14 standard deviation features exhibited positive t-values, suggesting enhanced variability in the Resected group. This lack of directional consistency contrasts with the ictal table's standard deviations, which all displayed positive t-values. Notably, features such as SampEn and LZC Mean emerged with particularly high t-values, indicating their potential discriminative power during the interictal phase. The KFD and ZCE means displayed an impressively consistent pattern, with 24 out of the 28 subjects mirroring the same trend. Interestingly, when juxtaposed with the ictal period, the interictal table manifests more features with t-values proximal to zero, suggesting a potentially reduced discriminative capability during interictal states.

Feature	t value (All Subs.)	p value (All Subs.)	% Subs. with +t values	% Sig. Subs. with +t values	% Subs. with -t values	% Sig. Subs. with -t values	Average t value across subs.	Average p value across subs.
MSE Mean	-30.32	9.9e-186	14% (4)	7% (2)	86% (24)	79% (22)	-6.29	7.1e-02
ApEn SD	28.36	4.5e-163	82% (23)	79% (22)	18% (5)	7% (2)	5.68	1.1e-01
HFD SD	26.32	7.0e-141	89% (25)	71% (20)	11% (3)	7% (2)	5.54	7.2e-02
ApEn Mean	-26.3	1.8e-143	25% (7)	11% (3)	75% (21)	57% (16)	-5.32	1.2e-01
SVDEn SD	26.2	5.2e-140	86% (24)	68% (19)	14% (4)	7% (2)	5.08	7.1e-02
1/f Slope SD	26.63	5.3e-143	100% (28)	89% (25)	0% (0)	0% (0)	5.07	1.5e-02
MSE SD	24.95	2.6e-128	93% (26)	75% (21)	7% (2)	7% (2)	4.92	4.6e-02
SampEn SD	20.83	1.7e-91	86% (24)	71% (20)	14% (4)	7% (2)	4.65	1.1e-01
SE SD	22.54	1.8e-106	82% (23)	57% (16)	18% (5)	14% (4)	4.62	1.1e-01
PFD SD	25.29	1.2e-131	89% (25)	71% (20)	11% (3)	7% (2)	4.4	1.0e-01
PE SD	24.33	5.0e-123	82% (23)	68% (19)	18% (5)	11% (3)	4.23	9.8e-02
HComp SD	23.47	4.4e-115	89% (25)	64% (18)	11% (3)	0% (0)	4.19	1.7e-01
HMob SD	19.81	3.5e-83	79% (22)	64% (18)	21% (6)	4% (1)	4.14	1.1e-01
LZC SD	18.93	1.5e-76	82% (23)	64% (18)	18% (5)	11% (3)	4.13	1.0e-01
SampEn Mean	-20.0	7.2e-86	39% (11)	18% (5)	61% (17)	46% (13)	-4.11	1.1e-01
KFD SD	19.83	2.4e-83	79% (22)	68% (19)	21% (6)	7% (2)	4.01	1.1e-01
LZC Mean	-15.2	4.9e-51	36% (10)	29% (8)	64% (18)	43% (12)	-3.21	1.3e-01
ZCR SD	14.43	5.5e-46	75% (21)	61% (17)	25% (7)	18% (5)	3.02	1.0e-01
HComp Mean	13.89	5.6e-43	57% (16)	54% (15)	43% (12)	29% (8)	2.33	8.7e-02
SVDEn Mean	-11.91	3.1e-32	43% (12)	32% (9)	57% (16)	36% (10)	-2.16	1.7e-01
SE Mean	-10.63	4.1e-26	39% (11)	32% (9)	61% (17)	39% (11)	-2.11	1.1e-01
PE Mean	-13.88	4.9e-43	36% (10)	21% (6)	64% (18)	43% (12)	-2.05	1.8e-01
ZCR Mean	-9.19	5.6e-20	43% (12)	32% (9)	57% (16)	39% (11)	-1.94	1.4e-01
KFD Mean	-9.0	3.1e-19	46% (13)	32% (9)	54% (15)	39% (11)	-1.75	1.6e-01
PFD Mean	-9.55	1.9e-21	39% (11)	25% (7)	61% (17)	39% (11)	-1.47	1.4e-01
1/f Slope Mean	8.1	7.3e-16	64% (18)	32% (9)	36% (10)	11% (3)	1.29	2.8e-01
HFD Mean	-5.49	4.2e-08	50% (14)	39% (11)	50% (14)	39% (11)	-0.74	4.1e-02
HMob Mean	-4.02	5.9e-05	54% (15)	36% (10)	46% (13)	36% (10)	-0.69	1.4e-01

Table 4.1: t-test statistics at the overall and subject-wise levels for ictal complexity features. The compared groups are the Resected and Non-Resected contacts. The "All Subs." columns show the t and p values at the overall level (the contrasted groups have the corresponding contacts from all of the subjects). The columns starting with "% Subs. with ± t values" showcase the percentage of subjects that had positive or negative t values respectively. After the percentage, the real number of subjects is shown in parentheses. Similarly, columns starting with "% Sig. Subs. with ± t values" present the percentage of subjects that had positive or negative t values respectively and that were also significant (i.e. $p < 0.05$). All percentages are with respect to the total number of seizure-free subjects (28). No multiple-comparisons correction was applied.

Feature	t value (All Subs.)	p value (All Subs.)	% Subs. with +t values	% Sig. Subs. with +t values	% Subs. with -t values	% Sig. Subs. with -t values	Average t value across subs.	Average p value across subs.
SampEn Mean	-35.12	2.9e-244	18% (5)	4% (1)	82% (23)	82% (23)	-7.09	5.2e-02
LZC Mean	-33.05	1.5e-218	18% (5)	7% (2)	82% (23)	75% (21)	-6.97	6.9e-02
ApEn Mean	-33.67	9.7e-225	18% (5)	7% (2)	82% (23)	79% (22)	-6.79	4.8e-02
MSE Mean	-34.6	4.1e-234	18% (5)	7% (2)	82% (23)	75% (21)	-6.65	7.6e-02
KFD Mean	-30.37	1.1e-187	14% (4)	4% (1)	86% (24)	64% (18)	-6.24	1.0e-01
ZCR Mean	-29.62	1.4e-179	14% (4)	7% (2)	86% (24)	64% (18)	-6.12	9.4e-02
SVDEn Mean	-26.32	2.9e-143	25% (7)	14% (4)	75% (21)	64% (18)	-5.46	6.0e-02
HMob Mean	-25.19	1.0e-132	21% (6)	18% (5)	79% (22)	57% (16)	-5.38	5.3e-02
SE Mean	-23.59	7.7e-117	29% (8)	14% (4)	71% (20)	54% (15)	-5.19	7.7e-02
HComp Mean	22.76	1.2e-108	75% (21)	61% (17)	25% (7)	11% (3)	4.37	7.3e-02
HFD Mean	-19.37	9.1e-81	25% (7)	18% (5)	75% (21)	61% (17)	-3.77	6.2e-02
HComp SD	20.23	3.7e-87	82% (23)	64% (18)	18% (5)	4% (1)	3.64	1.2e-01
KFD SD	-15.94	6.1e-56	18% (5)	11% (3)	82% (23)	46% (13)	-3.03	1.5e-01
ZCR SD	-16.92	1.2e-62	25% (7)	14% (4)	75% (21)	54% (15)	-3.0	1.5e-01
ApEn SD	8.44	4.2e-17	75% (21)	57% (16)	25% (7)	11% (3)	1.83	1.2e-01
HMob SD	-9.53	2.4e-21	39% (11)	14% (4)	61% (17)	36% (10)	-1.79	2.1e-01
LZC SD	-9.0	3.1e-19	29% (8)	11% (3)	71% (20)	25% (7)	-1.37	1.6e-01
PFD Mean	-8.2	2.9e-16	39% (11)	21% (6)	61% (17)	39% (11)	-1.23	1.6e-01
PE Mean	-9.95	4.3e-23	43% (12)	21% (6)	57% (16)	39% (11)	-1.2	1.8e-01
SampEn SD	-8.55	1.6e-17	39% (11)	14% (4)	61% (17)	29% (8)	-1.05	1.7e-01
PE SD	9.1	1.3e-19	57% (16)	36% (10)	43% (12)	25% (7)	1.02	1.6e-01
1/f Slope SD	6.61	4.4e-11	61% (17)	39% (11)	39% (11)	11% (3)	0.91	1.9e-01
SE SD	6.51	8.5e-11	61% (17)	39% (11)	39% (11)	14% (4)	0.82	2.0e-01
SVDEn SD	3.79	1.6e-04	61% (17)	36% (10)	39% (11)	18% (5)	0.78	1.8e-01
1/f Slope Mean	-2.2	2.8e-02	46% (13)	18% (5)	54% (15)	32% (9)	-0.58	2.4e-01
PFD SD	5.96	2.6e-09	61% (17)	25% (7)	39% (11)	29% (8)	0.56	1.9e-01
MSE SD	1.48	1.4e-01	39% (11)	29% (8)	61% (17)	32% (9)	0.1	2.2e-01
HFD SD	0.53	5.9e-01	54% (15)	29% (8)	46% (13)	21% (6)	-0.08	2.0e-01

Table 4.2: t-test statistics at the overall and subject-wise levels for interictal complexity features. The compared groups are the Resected and Non-Resected contacts. The "All Subs." columns show the t and p values at the overall level (the contrasted groups have the corresponding contacts from all of the subjects). The columns starting with "% Subs. with ± t values" showcase the percentage of subjects that had positive or negative t values respectively. After the percentage, the real number of subjects is shown in parentheses. Similarly, columns starting with "% Sig. Subs. with ± t values" present the percentage of subjects that had positive or negative t values respectively and that were also significant (i.e. $p < 0.05$). All percentages are with respect to the total number of seizure-free subjects (28). No multiple-comparisons correction was applied.

4.1.2   Feature Selection through the mRMR algorithm

The ictal and interictal features selected through the mRMR algorithm can be seen in Table 4.3. The selected features are surprising given the patterns seen on the previous statistical characterization. Specifically, in the analysis of the ictal period, while the statistical results highlighted a pronounced dominance of standard deviation (SD) features, the mRMR feature selection incorporated a blend of both mean and SD features. Similarly, for the interictal period, despite the statistical emphasis on mean-based features, the mRMR selection introduced a mix of both mean and SD metrics. This suggests that while one type of feature (mean or SD) might show stronger statistical differences, the mRMR method recognizes the value in incorporating both types. The selection of a diverse set of features by the mRMR procedure may hint at potential redundancy when solely relying on the dominant feature type and underscores the importance of leveraging both mean and SD features to capture comprehensive information about the Resected and Non-Resected groups of contacts.

Rank	Ictal	Interictal
1	MSE Mean	LZC Mean
2	1/f Slope SD	1/f Slope SD
3	HFD Mean	PFD Mean
4	HFD SD	1/f Slope Mean
5	MSE SD	HComp SD
6	HComp SD	PFD SD
7	PFD SD	HFD SD
8	PE Mean	

Table 4.3: Ictal and Interictal features selected by the mRMR algorithm.

4.1.3   Correlations with Spiking Activity

Figure 4.1 showcases the correlations of various features with the parameters designated for spike emulation. The main parameters examined are the width of the Gaussian window—which is inversely related to the spike's frequency content—and the total number of spikes in the signal. Notably, when the correlations are sorted by their absolute magnitude, both parameters preserve the same order of features. The figure elucidates both the mean and standard deviation variants of the features; intriguingly, their positions

appear mixed, without a clear pattern. An established correlation threshold at 0.5 reveals that more features (17) exhibit a strong correlation with the frequency content of the spikes than with the number of spikes (10). To further illustrate the relation between the features and the spike parameters, Appendix B contains figures B.1 and B.2 showcasing 2D heatmaps of the spike parameters against each feature (mean and standard deviations respectively). In addition, Figure B.3 in Appendix B shows the average normalized gradients of each feature with respect to the spikes parameters, calculated from the 2D heatmaps.
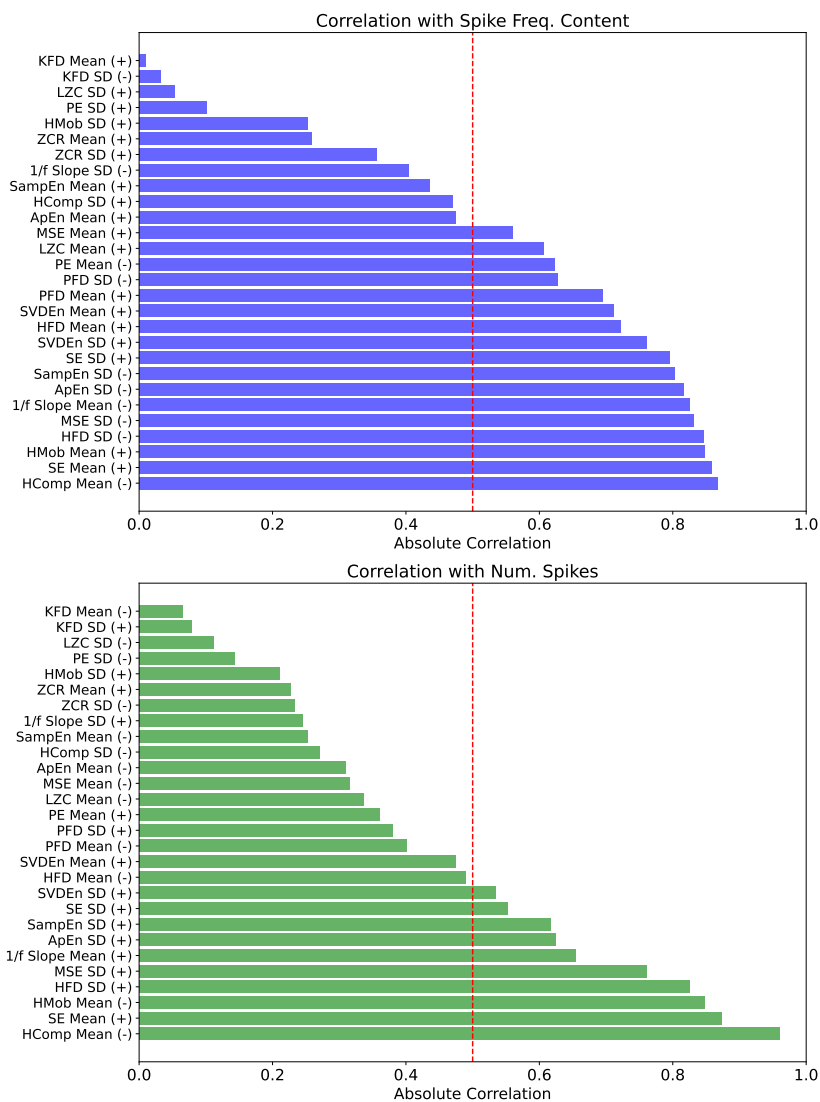

Figure 4.1: Absolute correlation of each feature with respect to the spike parameters. To the right of each feature the sign of the correlation is indicated ($\pm$).

4.2 Machine Learning Models' Performances

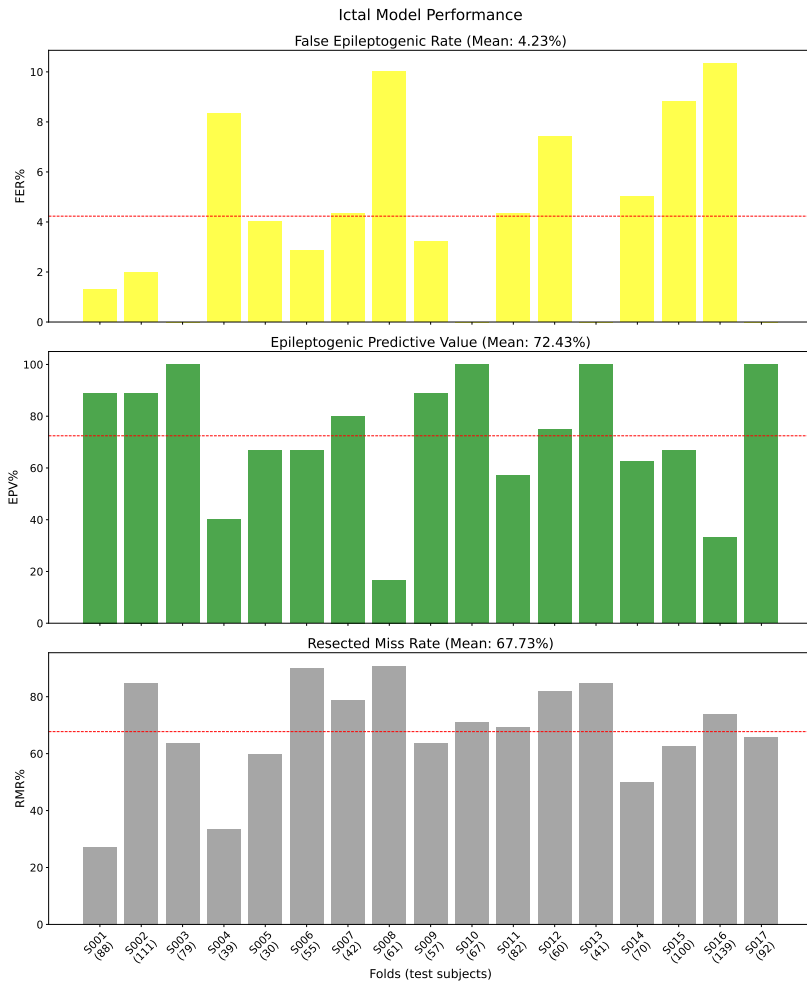## 4.2.1 Fold-wise performance

Ictal Model

Figure 4.2: Performance metrics of the Ictal model across the different subject-wise folds. The metrics are calculated on the test sets, which each represents a different subject. The mean value of each of the metrics is illustrated with a dashed red line and the exact value is shown in the title of each subplot. The number of contacts per subject is indicated in parentheses below the subject.

Figure 4.2 shows the performance of the Ictal Model across the 17 cross-validated subjects. This model used a threshold of 0.41, as explained in section 4.2.2. The model is conservative but every subject had at least one contact classified as epileptogenic. In particular, five subjects were problematic for the model: S004, S008, S012, S015 and

S016. These had low Epileptogenic Predictive Values and/or high False Epileptogenic Rates. Overall we can expect that around 72% of the contacts that were predicted epileptogenic to actually have been resected, implying a False Discovery Rate of 28%. Around 4% of the contacts outside the resection will be marked epileptogenic (96% of specificity) and 68% of the resected contacts won't be marked epileptogenic (32% of sensitivity).

Interictal Model

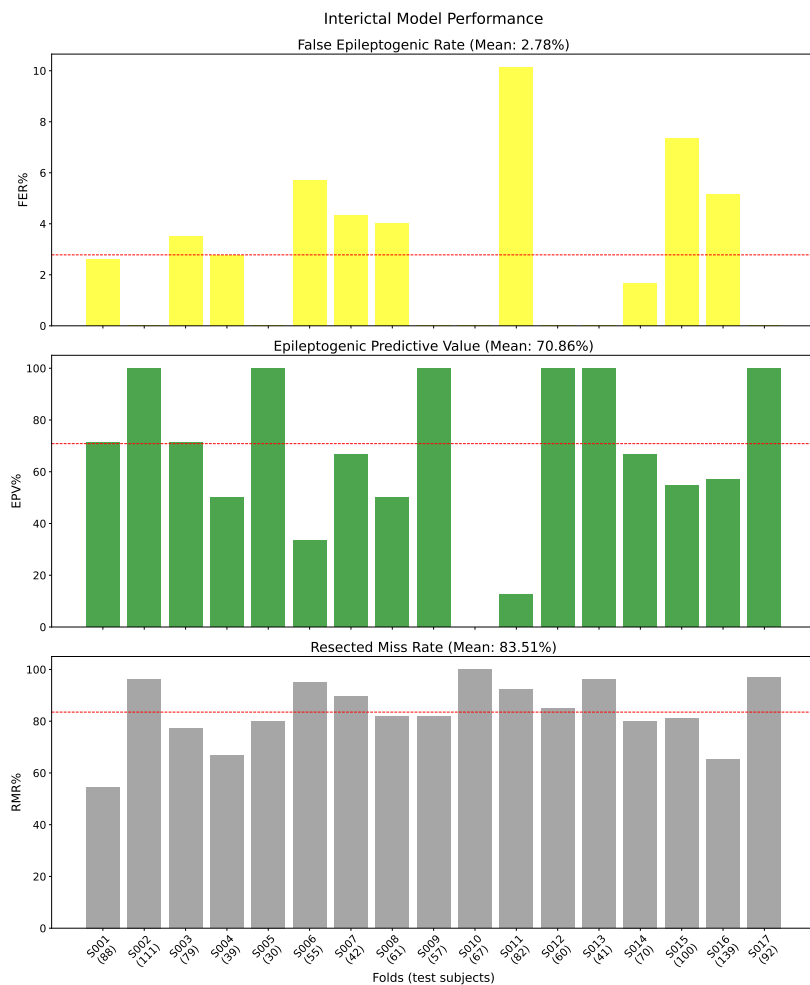Figure 4.3: Performance metrics of the Interictal model across the different subject-wise folds. The metrics are calculated on the test sets, which each represents a different subject. The mean value of each of the metrics is illustrated with a dashed red line and the exact value is shown in the title of each subplot. The number of contacts per subject is indicated in parentheses below the subject.

Figure 4.3 shows the performance of the Interictal Model across each fold associated to one of the 17 subjects. This model used a threshold of 0.43, as explained in section 4.2.2. Given that it has less information, this model has to be even more conservative, and as a result one of the subjects (S010) didn't have any contacts marked epileptogenic. Apart from this subject, six subjects were problematic for the model: S004, S006, S008, S011, S015 and S016. These had low Epileptogenic Predictive Values and/or high False Epileptogenic Rates. S006 and S011 proved to be particularly difficult with high False Discovery Rates between 70% and 90%. Nevertheless the mean False Discovery Rate was around 29%, thus, we can expect that around 71% of the contacts predicted epileptogenic will actually belong to the resection area. Around 3% of the contacts outside the resection area will be marked epileptogenic (97% of specificity) and 84% of the resected electrodes won't be marked epileptogenic (26% of sensitivity).

Hybrid Model

The use of both sets of features resulted in a 10% increase in the Epileptogenic Predictive Value, going as high as 80% (20% FDR). The threshold is set at 0.42. Every subject had at least one contact marked epileptogenic. As in the other models, six subjects were problematic, namely S004, S005, S008, S011, S015, and S016. These had high FER values and/or low EPV. Around 65% of the resected contacts will be marked non-epileptogenic, implying a 35% of sensitivity. The mean False Epileptogenic Rate is 3% which is more or less the same as in the Ictal and Interictal models, displaying 97 % of specificity.
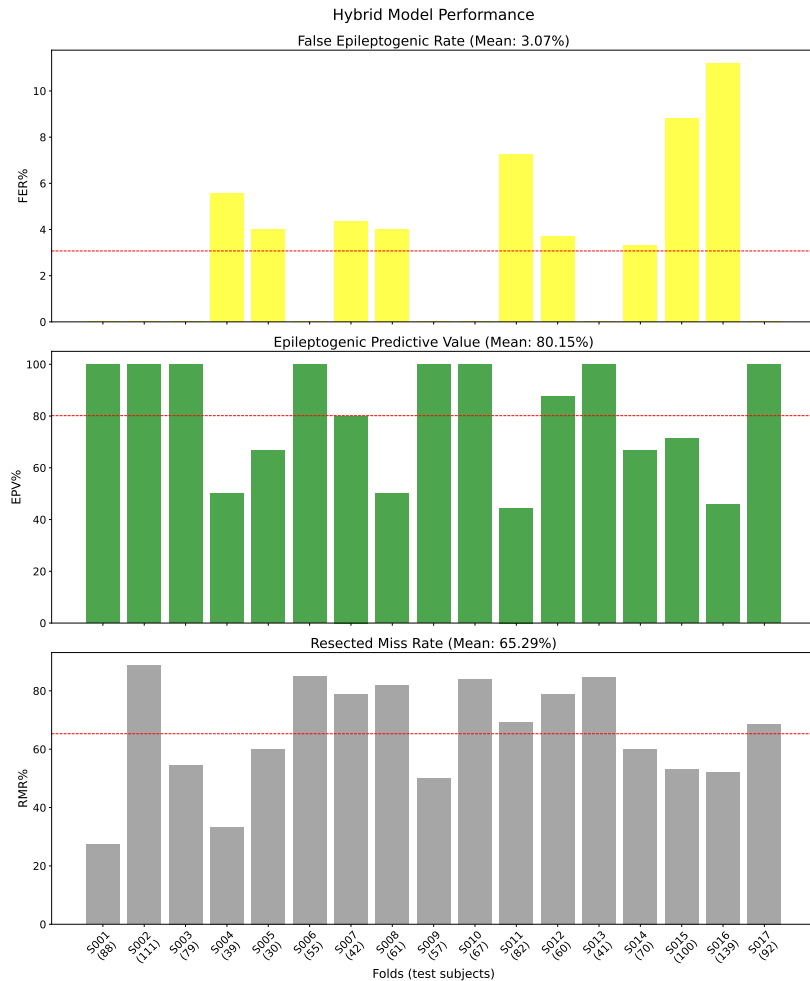
Figure 4.4: Performance metrics of the Hybrid model across the different subject-wise folds. The metrics are calculated on the test sets, which each represents a different subject. The mean value of each of the metrics is illustrated with a dashed red line and the exact value is shown in the title of each subplot. The number of contacts per subject is indicated in parentheses below the subject.

The Benchmark Model

The Bern-Barcelona dataset proved to be less problematic for the model. Using only interictal features it achieved around the same performance metrics as the Hybrid Model. This suggest that the statistical properties of this dataset are easier to handle, which can be observed from the fact that every fold behaves more or less the same in terms of the performance metrics. The Epileptogenic Predictive value sits at 82% (FDR=18%). While the False Epileptogenic Rate is actually a bit higher in comparison to the other models (6%) and implying a 94% of specificity. The sensitivity is similar to the ones of
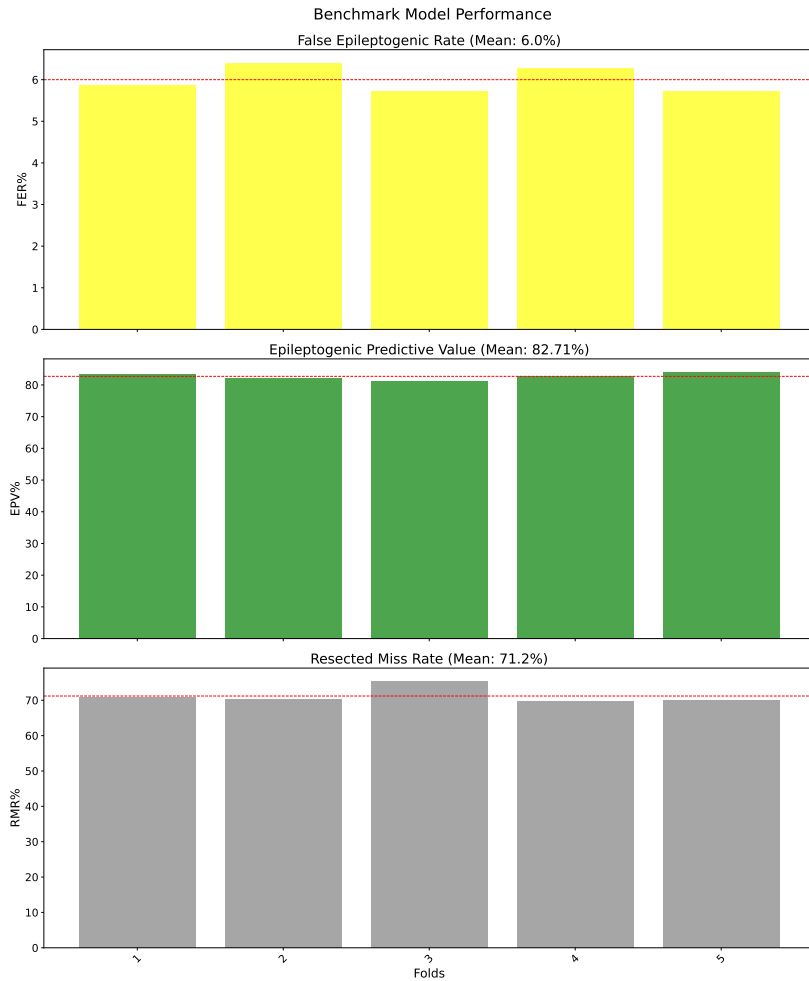
the other models as it is around 30%.

Figure 4.5: Performance metrics of the Benchmark model across the different folds. The metrics are calculated on the test sets. The mean value of each of the metrics is illustrated with a dashed red line and the exact value is shown in the title of each subplot.

4.2.2 Receiver operating characteristic curves

Figure 4.6 shows the ROC curves for all of the models. All ROCs were calculated from the predicted probabilities of the positive class in the corresponding test sets of the cross-validation folds. To the left of the figure are the ROC curves obtained from the hyperparameter optimization stage (using 11 validation subjects from Set B as explained in section 3.5.3). For each, the "Low FPR" threshold was selected and used as a threshold for the corresponding trained model to the right (which used the 17 training

subjects or the Bern-Barcelona dataset for the Benchmark model). Even though each trained model's ROC also shows its own optimal operating points, those were not used as operating points for the predictions, with the purpose of maintaining the test sets out of any parameter decision. Thus, they are included in the figures for completeness' sake. Concretely, the operating points for the models in Figures 4.2, 4.3, 4.4 and 4.5 are respectively 0.41, 0.43, 0.42 and 0.43, as shown in the "Low FPR" optimal points of their respective hyperparameter models. The hyperparameter validation set for the Benchmark model is the 11 validation subjects from set B using their interictal features, which is precisely the same as the set used in the hyperparameter phase of the Interictal Model. Thus, these two models have the same hyperparameter ROC curve and use the same operating point correspondingly.

Notably, all of the "Top-Left Corner" optimal points sit around a threshold of 0.2, while the chosen thresholds using the "Low FPR" criteria are around 0.4. The areas under the curve (AUCs) where around from 0.75 to 0.8 if the model used a feature set that included ictal data. The interictal feature set seems to provide lower AUCs that are around 0.7, except for the Benchmark Model which showed too an AUC around 0.8.
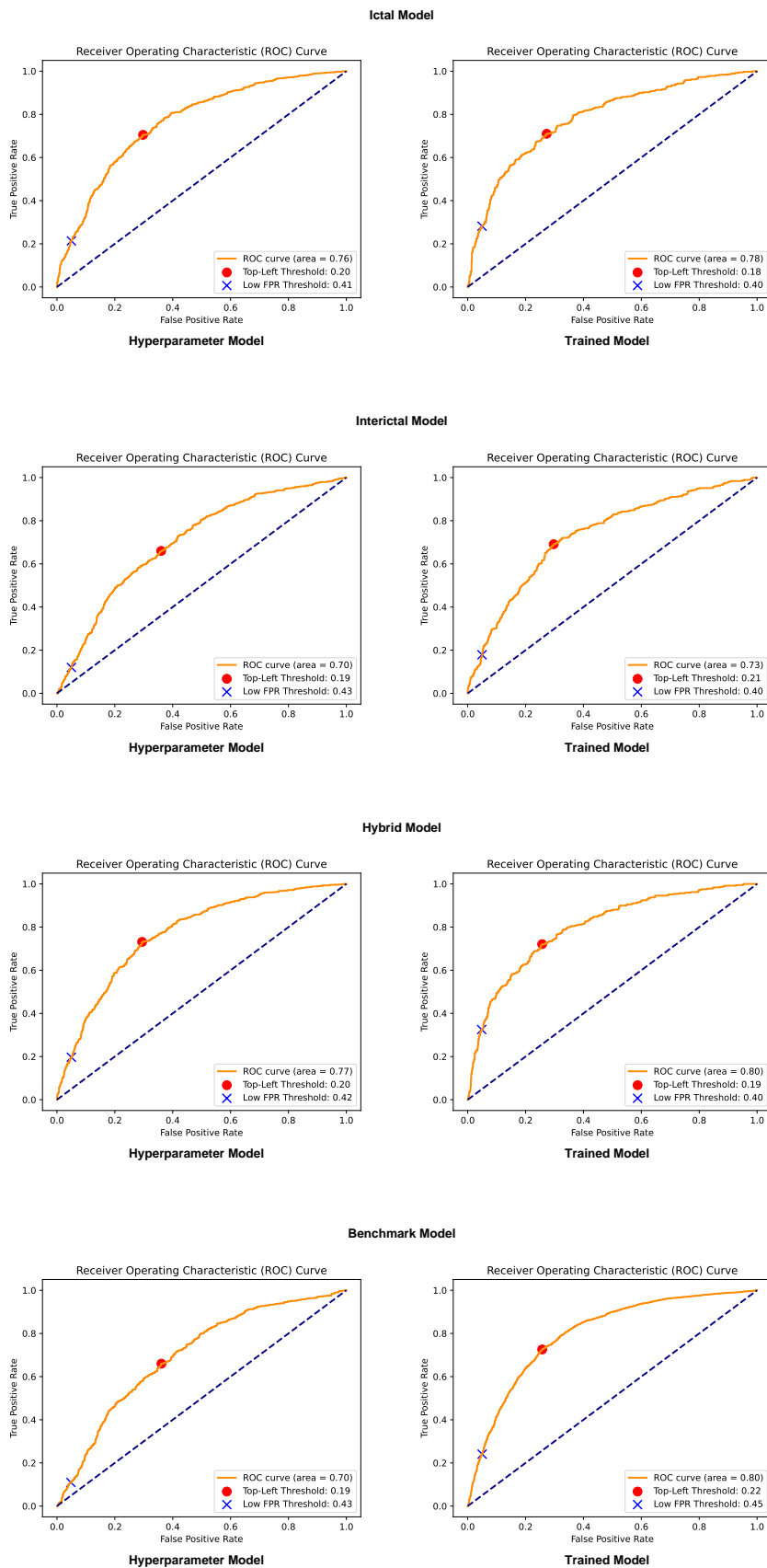
Figure 4.6: ROC curves of all of the models at the hyperparameter stage and after training.

4.2.3   Confusion Matrices

Figures 4.7 and 4.8 show the confusion matrices for each model in their row and column normalizations respectively. For Figure 4.7, the row normalization is equivalent to a normalization by the total count of each Resection class (Inside or Outside). In this Figure all models behave similarly, 95% of contacts outside the resection are marked Non-Epileptogenic and around 30 % of contacts inside the resection are marked Epileptogenic. The deviation is on the Interictal model where only 13% of contacts inside resection are marked as Epileptogenic, making this model more conservative. More details can be found in Appendix C.1, where the confusion matrices per fold and normalized by the resection class count are shown.

For Figure 4.8 the column normalization is equivalent to a normalization by the total count of each Predicted Class (Epileptogenic or Non-Epileptogenic). Here the patterns are more diverse. The Ictal and Hybrid models have a similar behavior, with both Epileptogenic and Non-Epileptogenic Predictive Values around the 75% mark. The Non-Epileptogenic Predictive Value of the Benchmark Model is 10% lower (around 60%) and for the Interictal model it is only 5% less (70%). In contrast, the Epileptogenic Predictive Value for the Benchmark model is higher than that of any model (83%). On the other hand, the Interictal model has the lowest Epileptogenic Predictive Value (62%). More details can be found in Appendix C.2, where the confusion matrices per fold and normalized by the predicted class count are shown.

It is important to note that the values in Figures 4.7 and 4.8 are obtained by aggregating all the predictions into a single confusion matrix, and then calculating the metrics. This is different from what was done in the per-fold performance analysis (section 4.2.1), where the quantities are computed per fold and then averaged without previously aggregating the predictions into a single matrix.

Figure 4.7: Confusion Matrices of the models normalized by Resection class count (row sum). In parentheses the absolute number of samples is indicated. Each single sample in a confusion matrix correspond to a single contact of a single subject. All of the samples used in each confusion matrix are extracted from the collected predictions across the test sets of the corresponding cross-validation folds and model.
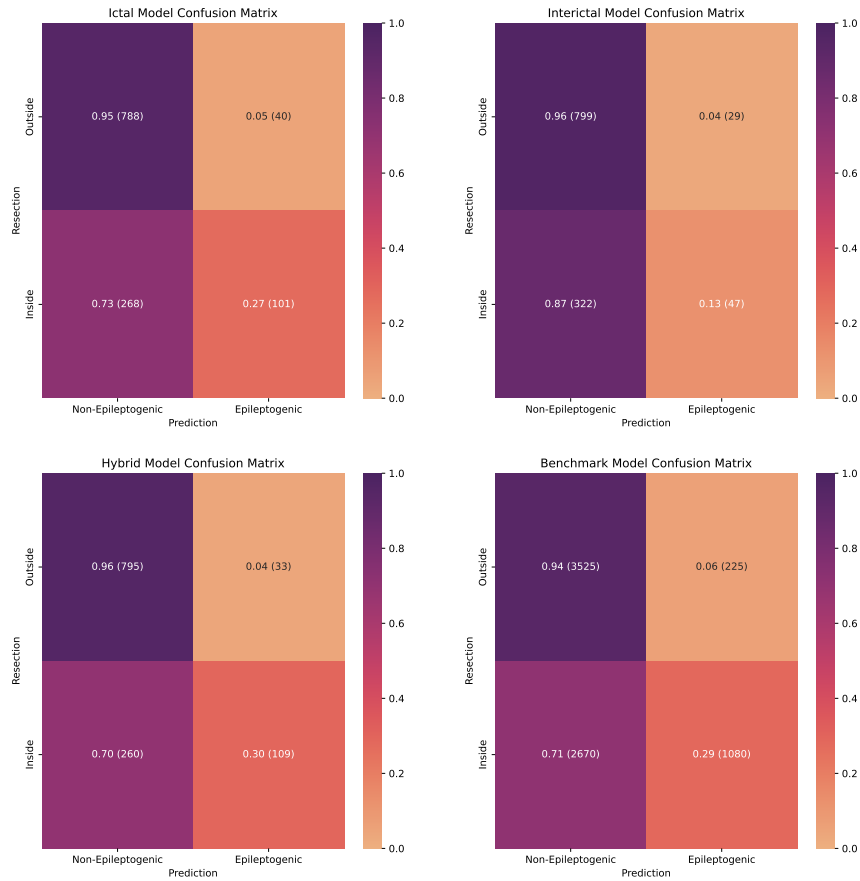
Figure 4.8: Confusion Matrices of the models normalized by predicted class count (column sum). In parentheses the absolute number of samples is indicated. Each single sample in a confusion matrix correspond to a single contact of a single subject. All of the samples used in each confusion matrix are extracted from the collected predictions across the test sets of the corresponding cross-validation folds and model.
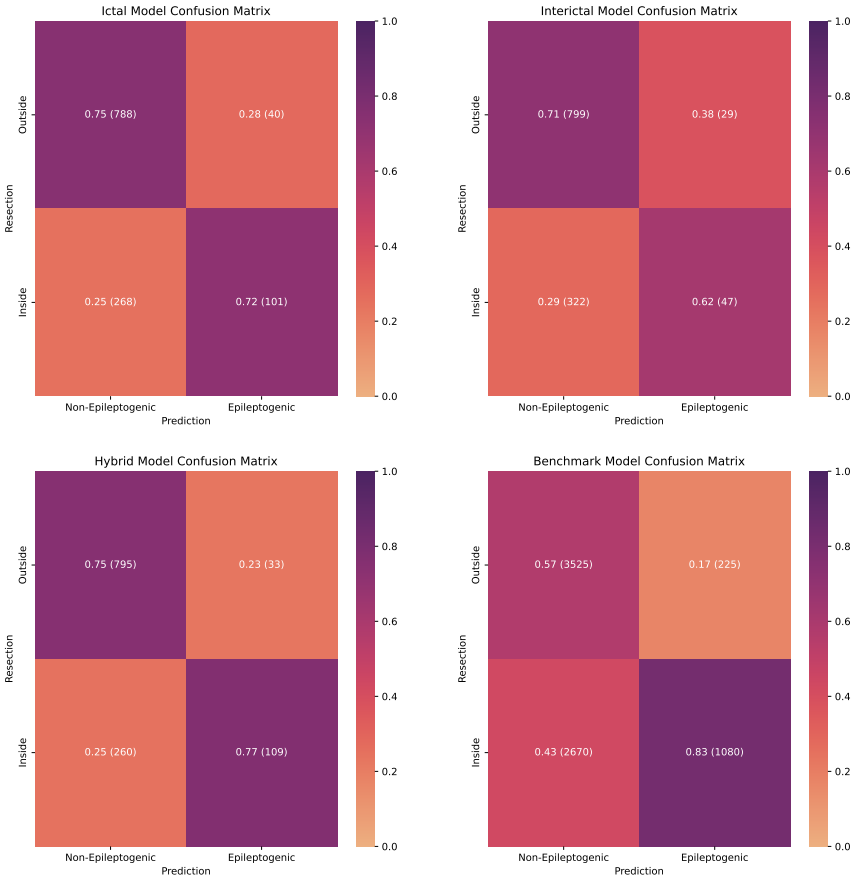
4.2.4 Summarized Performances

Table 4.4 summarizes the performance results of the previous sections. Notably from the models tested on the Cleveland Clinic Dataset, the best Epileptogenic Predicted Value performance (either from the mean across folds or the confusion matrices) is showcased by the Hybrid Model, followed by the Ictal and Interictal models in that order. This pattern is mirrored by the AUC metric. The False Epileptogenic Rates of all models trained on the Cleveland Clinic dataset are below 5% (whether from the fold-average or the confusion matrices). This is nonetheless counteracted by the low EPV value (62%) obtained from the confusion matrix of the Interictal model. The Resected Miss Rate tends to be relatively high, between 65% and 90%. With respect to the Benchmark Model, the performance is better than that found in the Interictal Model and even the Hybrid model. Nonetheless, this is potentially a result of the less challenging statistical properties of the Bern-Barcelona dataset.

Model	Mean across Folds			ROC	Confusion Matrices		
	EPV	FER	RMR	AUC	EPV	FER	RMR
Ictal	72.43%	4.23%	67.73%	0.78	72%	5%	73%
Interictal	70.86%	2.78%	83.51%	0.73	62%	4%	87%
Hybrid	80.15%	3.07%	65.29%	0.8	77%	4%	70%
Benchmark	82.71%	6%	71.20%	0.8	83%	6%	71%

Table 4.4: Summary of the performances metrics. The "Mean across Folds" columns corresponds to the average values across the fold of the metrics shown in section 4.2.1. The ROC column has the area under the curve of the trained models shown in section 4.2.2. The "Confusion Matrices" columns has the performance metrics extracted from section 4.2.3, which -in contrast to the fold averages- first accumulates the predictions in the matrix and then calculates the metrics.

4.3   Feature Importances

4.3.1   Ictal Model

Figure 4.9 shows the feature importances of the Ictal Model. Notably, the first 4 features decline in an almost linear way, after which there is more pronounced decline for the 5th feature, and then similar low values are shared by the last 3 features. The overall shape seems to be that of a half Bell curve. Three out of the first four features consist of standard deviations. Interestingly, some of the features show narrow boxplots (MSE Mean, 1/f Slope SD, MSE SD and PFD SD), while the other display more broad ones. From this, it is expected that some features interchange rankings for some of the subjects, e.g. the 1/f Slope SD with HFD SD. Nevertheless, the variability is low enough to consider the feature importance to be somewhat stable across the different subjects.
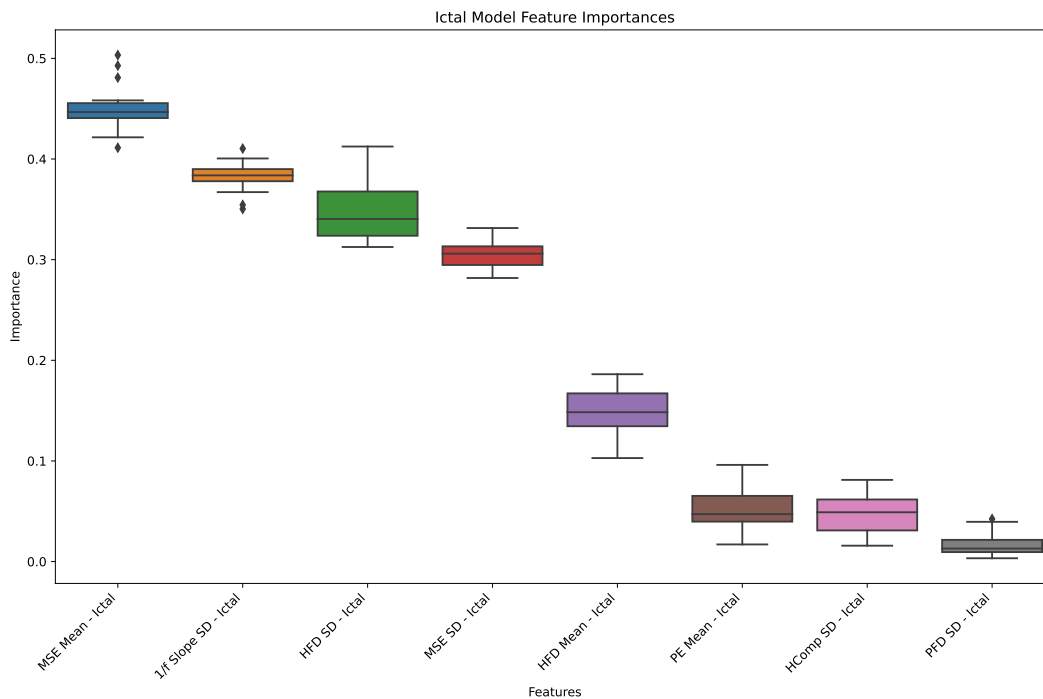
Figure 4.9: Feature Importances of the Ictal Model.

4.3.2   Interictal Model

Figure 4.10 shows the feature importances of the Interictal Model. Interestingly, the decline of the features seem to follow an almost exponential curve. The variability of the importances is also quite narrow. For the first 3 features, the variability is low and the features occupy well separated spaces of the plot; thus the ranking of the first three features should be completely stable across the 17 subjects. The other four features also exhibit low variability but they overlap in the feature importance space.

Figure 4.10: Feature Importances of the Interictal Model.

4.3.3   Hybrid Model

The shape of the feature importance curve for the Hybrid Model (showcased in Figure 4.11) resembles that of the Ictal model but with more features. In this case the Bell curve seems more clear. Importantly, the variability of the feature importances in conjunction with the overlap of the features in the importance space suggests that the features probably exchange ranking across the folds more often than not. Nonetheless, the pattern is clear enough to make inferences at least at the level of groups of features. For example, the first five features seem to create a first group, then the next 6 features can be use to do a second one, and finally the last four features could be clustered together. The first group described is dominated by ictal features, while the second one has majorly interictal ones. The last group also has mostly ictal features. There is also a mix of Mean and SD features across the ranking. Perhaps not surprisingly, the order in which ictal and interictal features appear seems to somewhat resemble the original order in which those features appeared when the periods were considered in isolation.
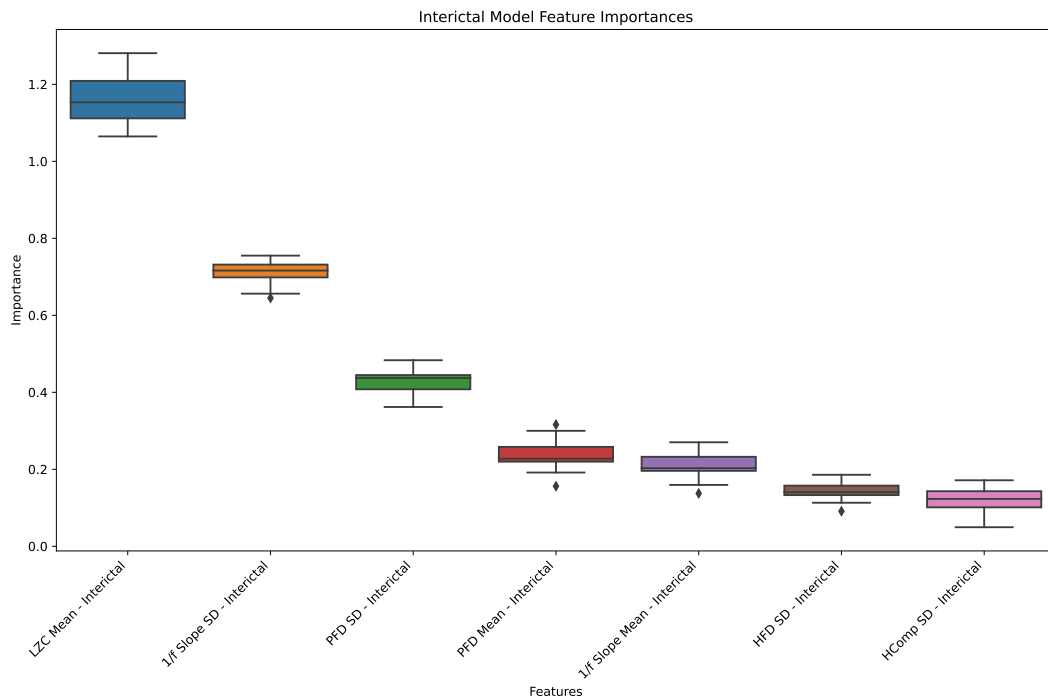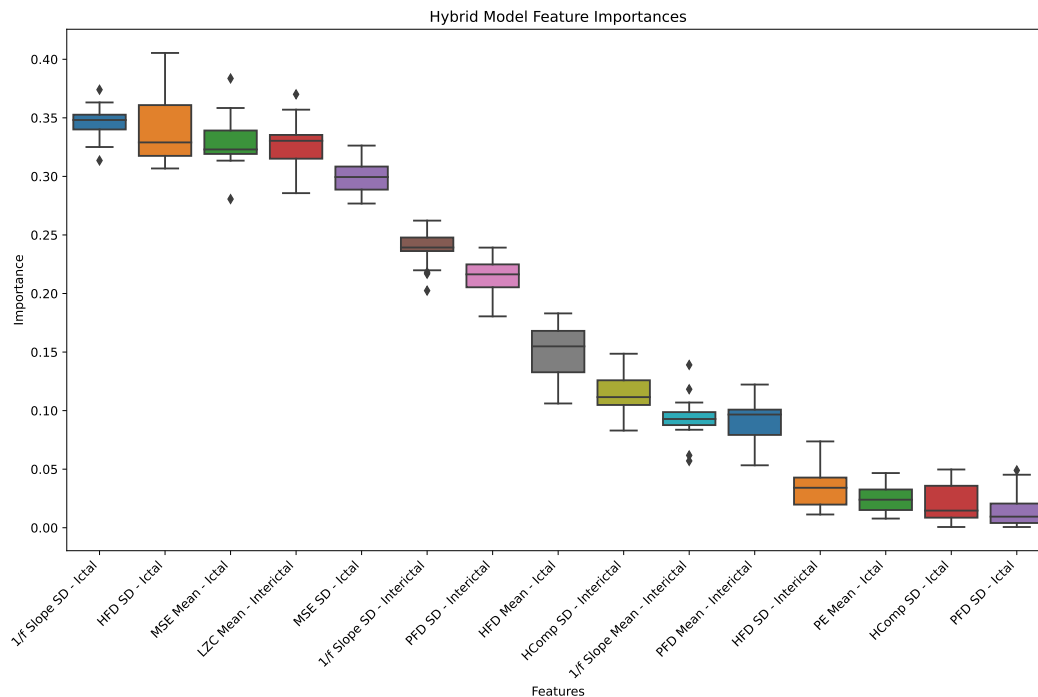

Figure 4.11: Feature Importances of the Hybrid Model.

4.3.4 Benchmark Model

Figure 4.12 shows the feature importance for the Benchmark model. Intriguingly, the first feature (LZC Mean) seems to vastly dominate the model, along with a second feature but in a significantly less amount. The next four features seem to cluster together and exchange ranking through the different folds. The last feature corresponds to the 1/f Slope mean, and it seems to be so in every fold as it is not superposed in the importance space with any other feature. Moreover, the order of the interictal features in this model differs from that found in the Interictal model, though both share the same most dominant feature (LZC Mean), which accentuates the relevance and reproducibility of this feature.

Figure 4.12: Feature Importances of the Benchmark Model.

Chapter 5

Discussion

5.1 About Feature Statistics and Selection
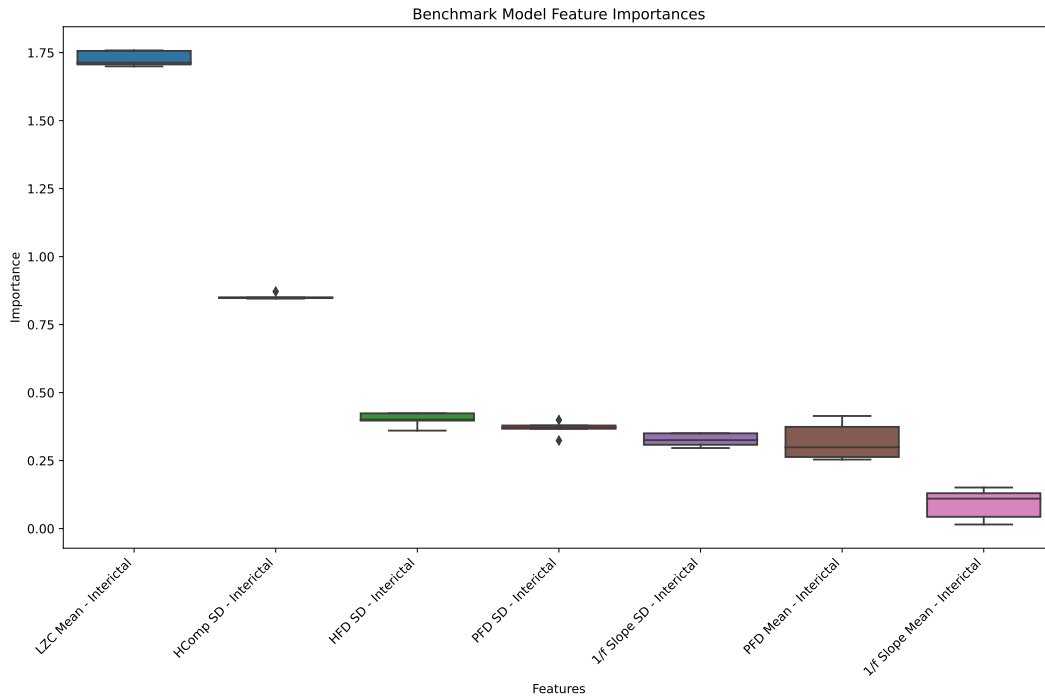
Both ictal and interictal tables share the presence of "fake" and "real" epileptogenic contacts within the Resected group, which hinders stronger statistical patterns. Moreover, the reduced statistical significance at the subject-level, when compared to the overall level, hints at the inherent variability across subjects and the possibility that some individuals deviate from general trends. Importantly, some features in both tables exhibit a stable pattern across subjects, such as the 1/f Slope standard deviation and the Multi-Scale Entropy mean in the ictal period, and the KFD and ZCR means in the interictal period. These offer intriguing avenues for further exploration and validation.

Delving more into each of the periods we note that the prominence of standard deviations in the upper part of the ictal table and means in the lower part may suggest that variability within features provides more discriminatory power than central tendencies in the ictal state. This could be a reflection of the evolving dynamic shown in the ictal period (Figure 2.2). Despite this, some mean-based features such as Multiscale Entropy and Approximate Entropy were able to stand out in the ictal period, which might indicate unique qualities or sensitivities of these features in capturing something beyond the transient dynamics. Furthermore, the dominance of mean features in the interictal period could reflect the stable dynamics inherent to this state, in contrast to the evolving dynamics observed during the ictal period. Curiously, in both tables the mean-based features tend to have negative t-values, which may indicate that the Resected group exhibits less complex dynamics, given that most of these features reflect an aspect of complexity.

Consistency between the selection carried out by the mRMR algorithm and the statistical characterization tables can be seen when observing the first features drawn for each period. In example, both MSE Mean and LZC Mean were around to the top of the ictal and interictal tables respectively. Nevertheless, the selection also encompasses features which don't reflect strong statistical patterns, such as the case of selecting HFD SD for the interictal period. The inclusion of such features indicates that they may have useful information that is not found in the most statistically dominant features, possibly exhibiting more complex patterns that are more easily captured from a machine learning framework [7], and which may have been highlighted by the XGBoost stage of the mRMR algorithm. Another possibility is that they don't carry as much information but they got selected as the statistically dominant features were deemed redundant.

One avenue not explored within this analysis involves the experiment of applying the mRMR algorithm per subject and then looking for features appearing in most of the subjects. This could have diminished the bias of over-represented subjects in the feature selection process. Similarly, it could also be possible to select the features based solely on the statistical results rather than the automated mRMR, though this would involved a careful consideration of the correlations between the features. At this, a promising approach would be to select the features with patterns that are most stable across subjects rather than the strength at the overall level.

5.2 About the spiking activity

In the results from the correlation of the features to the spike emulation parameters it is seen that a high number of features have correlations above 0.5. This means that if there is significant spiking in the signals, then it cannot be guaranteed that these features are predominantly reflecting criticality-related properties rather than the spikes themselves. Such a scenario potentially impedes the objective of this thesis. In the literature some researchers describe entropy features as their own set of properties of a signal (e.g. [35]), while others use them as ways to quantify biomarkers such as spiking and even high-frequency oscillations (e.g. [89]). From the analysis done in our work, we find that more features tend to be strongly correlated to the frequency content of the spikes than to the number of spikes. This may be a result of the features being computed on time windows. From this experiment it becomes apparent the importance of characterizing the amount of spikes in the dataset and possibly the need of improving the pre-processing steps, such that the windows extracted from the signal become spike-free.

5.3   About the models' performances

When focusing only on the main models (the ones trained on the Cleveland Clinic Dataset) it can be seen that the highest performing one is the Hybrid model, which naturally comes down to the fact that it has a greater amount of diverse information. If we pursue this association between model performance and information availability, it can also be seen that the ictal period contains more informative features in comparison to the interictal period. Nevertheless, the 10% performance increase in the Hybrid Model suggests that the interictal period has relevant information unavailable in the ictal stage.

Notably, there are some important disagreements between the subject-wise mean of performance metrics and their equivalents extracted from the confusion matrices. This is particularly worrying in the case of the EPV performance of the interictal model, which has as 8% decrease when computed on the overall matrix. Nevertheless, the opposite case (having lower performance in the subject-wise folds) would be even more worrying.

Grinenko et al. [26] developed a Support Vector Machine (SVM) model based on spectrotemporal features of the signal. Concretely, it was conceptualized as a fingerprint quantified by the extraction of properties from the normalized spectrogram of the ictal period by the interictal one. Given this, the most directly comparable model from our work to this SVM model corresponds to the Hybrid Model. The performances of the Hybrid model is indeed below that of the SVM model (SVM: 90.6 % PPV vs 80% PPV, and SVM: 0.7% FPR vs 3% FPR). The Hybrid model mostly fairs better in terms of the simplicity of the model. Notably, the SVM model leverages a set of features that quantify some biomarkers already conceptualized as possible delineators of the epileptogenic zone. In contrast, the model developed in the present work is more exploratory in nature. Curiously, the Hybrid model exhibits a lower FNR (SVM:84% FNR vs 65% FNR). It would be possible then to find an operating point of the Hybrid Model where the threshold results in an increased PPV at the cost of increasing the FNR to 84%.

Interestingly, the performance of the Benchmark Model -which is based on using only interictal data- is similar to that of the Hybrid Model (which uses both periods). This is probably a result of the benchmark dataset having more homogeneous statistical properties stemming from the fact that it has less subjects (5 vs 17) and that the corresponding folds of its training aren't done per subject, given that such information is missing from the dataset. From the confusion matrices shown in section 4.2.3 one can extract features such as Accuracy, Specificity and Sensitivity, which are the most

commonly evaluated. Concretely, this Benchmark Model shows an accuracy of 61.4%, which is much lower than many of those found in the literature [35]. It is important to note that these models are usually more complex (e.g. SVMs) which hindrances their interpretabillity. Moreover, many of them employ decomposition methods such as wavelets to further treat the signal. Importantly, some of these models show a lower specificity (e.g. 84% vs 94%) which suggests that they were tuned to optimize the overall accuracy rather than the positive predictive value. In addition, it is possible that some of those models didn't use cross-validation strategies, which further affects the estimated performance.

5.4 About feature importance

The dominant features in the ictal period correspond to the 1/f Slope standard deviation (SD), the Higuchi Fractal Dimension SD, and the Multiscale Entropy (MSE) Mean and SD. The prevalence of standard deviations may reflect the transient dynamics of the ictal stage in epileptogenic (or at least resected) contacts. Similar to what happened in Table 4.1, it is surprising to see a mean-based feature at the top of the feature importance. The fact that MSE reflects a complexity measure less-biased towards randomness through the use of multiple time-scales may have provided a basis for it to capture information beyond the transient dynamics of the ictal phase. It is important to note that in the ictal signals some pre-ictal spikes may be present, but MSE is relatively unaffected by the spikes (particularly if they have high frequency content) as can be seen in Figures 4.1 and Appendix B. Moreover, MSE is anchored to the concept of long-range temporal correlations, but given that we analyze the signal on short windows of 1 second it may be that it reflects this property relative to much shorter time scales. On the other hand, the interpretation of the 1/f Slope SD is less straightforward; its natural value (the mean) should reflect the Excitation/Inhibition balance in the brain. The standard deviation of such characteristic in the ictal phase could reflect transient changes of the system state of a particular location of the brain during a seizure. Likewise, the HFD represents a fractal dimension, which on one-dimensional signals could be portraying the long-range correlations properties of the system, though the same nuance of the MSE applies given the short windows. In general, and noting that from Table 4.1 we know that all standard deviations are larger among resected contacts, it could be that these bigger standard deviations imply a similar conclusion to that of the 1/f Slope standard deviation: the sudden changes in the local state of the system but in respect to their

correlations at different scales.

Nonetheless, a different cause could be identified. The work of Grinenko et al. [26] proposed that the epileptogenic regions exhibit special characteristics such as Multiband Fast Activity, Low Frequency Suppression and Pre-ictal Spikes. There may be some correlations between the complexity-based features used and those characteristics. In particular, it is possible that the standard deviation of some of the features in the ictal period reflect aspects of these biomarkers. Concretely, the combination of low frequency suppression and multiband fast activity could result in a mostly oscillatory signal that exhibits low complexity. Thus, if this signature is transient it will affect only some percentage of the segmented windows, being ultimately reflected on the standard deviation of the features. Moreover, one hindrance the HFD SD has is the fact that it is one of the features that reflect more strongly both the number of spikes and their frequency content. Importantly, the 1/f Slope SD does not show strong correlations for neither of these.

Moving into the interictal feature importance, we note that the mean of the Lempel-Ziv Complexity (LZC) has the largest weight with a considerate margin, both for the Interictal model and the Benchmark model. Interestingly, Toker et al. [84] proposed that LZC was maximized at the edge of chaos, which is a particular class of criticality where the transition happens between stable and chaotic phases of the system. Nevertheless, the interpretation of LZC becomes problematic in this project, as it can be thought of as a measure of Kolmogorov Complexity, which is maximized in randomness rather than in meaningful long-range correlations as happens for MSE [88]. Despite this, and in similarity with MSE, LZC wasn't particularly among the features most affected by spikes. Intriguingly, the next two more important features in the interictal model are the 1/f Slope SD and the Petrosian Fractal Dimension (PFD) SD, which mimics the same pattern seen in the ictal importances: one mean-based feature, the 1/f Slope SD, and one Fractal Dimension SD. In contrast to the Higuchi Fractal Dimension SD, the PFD SD has moderate correlations to the frequency content of the spikes, and low correlation with their number. This is surprising since PFD involves the sign changes of the derivative of the signal as one of the main concepts of its calculation, which could reflect spikes or the seizure itself. Apart from this and in regards to the 1/f Slope SD, its significance may be disputed in the interictal period because it does not show a stable pattern (only 61% of the subjects have a positive t value), though the exact same happens for the PFD SD (Table 4.2). This contrasts the stable patterns of the 1/f Slope SD and HFD SD in the ictal period where 100% and 89% of the subjects have the same pattern respectively.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

This project explored the many intricacies involved in the task of Epileptogenic Zone localization. Importantly, both ictal and interictal periods display challenges due to the presence of "fake" and "real" epileptogenic contacts within the Resected class. Despite this, our research underscores that while the interictal feature set might offer less information for the task, its incorporation into a hybrid model, which combines both ictal and interictal phases, enhances the performance beyond the ictal-only model. Feature importance evaluation showed that there were both commonalities and differences in the feature sets that were dominant in both periods. Notably, pivotal characteristics were identified in determining the fingerprint of the epileptogenic zone. This includes the prominence of the mean Multi-Scale Entropy during the ictal phase and the mean Lempel-Ziv Complexity in the interictal phase. Such characteristics show stable statistical results across subjects, and resonate with the brain criticality hypothesis, shedding light on temporal correlations across varying scales and the transitions between stable and chaotic brain phases. Furthermore, the present work emphasizes the importance of fractal dimension measures reflected in the top-3 critical features. Both interictal and ictal phases reveal that the standard deviation of the 1/f slope and fractal dimensions bear significant information, especially pronounced during the ictal phase. However, caution is advised in the interpretation of these findings due to potential metric distortions caused by spikes. Lastly, while this research ventured into analyzing model performance with a benchmark dataset, the results weren't groundbreaking. The diminished performance may be attributed to the simplicity of the utilized model and signal processing techniques.

In general, given the promising insights unveiled in this exploration, there remains an optimistic horizon for subsequent studies that delve deeper into complexity features, potentially heralding groundbreaking advancements in our understanding of epilepsy within the framework of the brain criticality hypothesis.

6.2 Future Work

Potential directions for upcoming research could focus on addressing the existing constraints of the methods shown and advancing the techniques developed. Specifically:

- The interference of spikes obfuscate conclusions framed in the critical brain hypothesis. Logically, one point of improvement would be to include a method to detect the spikes, and subsequently select only spike-free windows. This could be done through a dynamic segmentation of the signal that allows for big overlaps, maintaining the amount of usable time windows.

- The machine learning model implemented in this project is simpler than those found in the literature. An avenue that could be explored to improve performances is to try models that are more complex, at the cost of sacrificing interpretability.

- Another of the main limitations is the uncertainty in regards to the epileptogenicity of the resected class. Two avenues can be explored here through the use of unsupervised learning: A) to find clusters of contacts with similar characteristics and relate the results to their location and resection class. And B) to leverage an unsupervised clustering approach within the resected class. This needs an assumption that is able to determinate which cluster is more likely epileptogenic. In our case this wouldn't be possible using complexity features but it can be achieved by leveraging other features only used to this purpose, such as those found by Grinenko et al. [26]. On the other hand, this would constitute a bias that constrains which features become important.

- The feature selection process could involve an approach that privileges the features with more stable statistical patterns across all of the subjects. This could in turn increase the subject-wise fold performance.

- When making a prediction, it could be possible to leverage information from surrounding contacts. Such effort would be motivated by the known network nature of epilepsy.

Appendix A

Clinical profiles of the patients

Table A.1: Clinical profile of the patients.

Sub ID[a]	Age (years)	Num. Epilepsy years	MRI lesion	Surgical pathology	Resection (or ablation) details	Outcome	Follow-up months
S001	43	37	FCD, insular/frontal operculum	FCD, type 2B	Anterior insula/frontal operculum	Seizure-free	13
S002	29	22	Negative	FCD, type 1C	Temporal-parietal-occipital	Seizure-free	49
S003	33	17	Hippo-campal sclerosis	Hippocampal sclerosis type 1	Anterior temporal lobe	Seizure-free	48
S004	17	8	Negative	No pathology	Laser ablation, superior frontal gyrus	Seizure-free	19
S005	16	1	Benign neoplasm, posterior parahip-pocampal gyrus	Low grade glial-glioneuronal neoplasm	Posterior parahip-pocampus gyrus and neoplasm	Seizure-free	39
S006	46	41	FCD, mesial frontal	Non-specific changes	Prefrontal lobe	Seizure-free	38
S007	5	1	Negative	FCD, type 2B	Superior frontal gyrus, superior frontal sulcus, frontal pole	Seizure-free	21
S008	63	14	Negative	FCD, type 1	Orbitofrontal	Seizure-free	44
S009	33	19	Gliotic postop changes	FCD, type 1B	Anterior temporal lobe	Seizure-free	40
S010	21	11	Negative	Grey matter het-erotopia, FCD type 1B	Occipital lobe	Seizure-free	12
S011	32	27	FCD, precentral gyrus	Non conclusive	Precentral gyrus	Seizure-free	77

S012	22	3	FCD, superior frontal sulcus	FCD type 2 B	Superior and middle frontal gyri, anterior cingulate	Seizure-free	78
S013	19	18	Negative	FCD type 1	Middle frontal gyrus	Seizure-free	48
S014	30	18	Negative	FCD type 2 B	Frontal operculum	Seizure-free	47
S015	20	11	Negative	FCD, type 1	Frontal lobe	Seizure-free	82
S016	65	25	Negative	FCD, 1C	Anterior temporal lobe	Seizure-free	39
S017	65	9	Negative	FCD, 1C	Anterior temporal lobe	Seizure-free	36
S101	25	11	Normal	Focal gliosis	L lateral temporal cortexectomy	Seizure-free	27
S102	17	7	Normal	FCD Type 1	L temporal polar and amygdala resection	Seizure-free	36
S103	30	12	Normal	FCD Type 1	R anterior temporal lobectomy	Seizure-free	36
S106	17	9	Normal	FCD Type 2B	R SMA/cingulate resection	Seizure-free	28
S108	37	32	Suspected, FCD	FCD Type 2B	R subcentral resection	Seizure-free	20
S111	48	6	Normal	FCD Type 1	L anterior temporal lobectomy	Seizure-free	28
S112	21	18	Normal	No due to laser surgery	L insular / temporal / frontal operculum laser ablation	Seizure-free	31
S113	24	17	Suspected FCD	FCD Type 1	R anterior temporal lobectomy	Seizure-free	29
S116	11	7	Prior resection, otherwise normal	Gliosis	R insular/ fronto-parietal and temporal operculum	Seizure-free	22
S118	33	13	Normal	Gliosis	R prefrontal resection	Seizure-free	19
S140	39	3	Normal	Focal perivascular gliosis	Anterior temporal lobectomy	Seizure-free	21
S215[b]	35	4	PNH	No due to laser surgery	Laser ablation, periventricular nodule	Seizures	
S219[c]	30	5	Normal	No due to laser surgery	Laser ablation, L cingulate/SMA	One-year seizure-free then seizure recurred	
S220	38	22	Normal	FCD Type 1	R posterior basal temporal resection	Seizures	

S221	41	39	Normal	Gliosis	R lateral temporo-parietal resection	Seizures	
S222	24	14	Normal	FCD Type 1	R basal posterior temporal resection	One-year seizure-free then seizure recurred	
S223	6	2	Normal	Inflammation, FCD Type 1	L anterior lateral temporal resection	Seizures	
S226	24	18	Normal	FCD Type 1	L prefrontal resection	Seizures	
S228[b]	25	12	Multiple areas of gliosis	Gliosis	R parieto-occipital resection	Seizures	
S231[c]	34	34	Normal	No due to laser surgery	Laser ablation, L frontal operculum	Seizures	
S232[b]	10	10	Bilateral occipital lesion	Ulegyria, inflammation	L parieto-occipital resection	Seizures	
S233	29	10	Heterotopic gray matter	FCD Type 1	R temporooccipital resection	Seizures	
S237	20	16	Normal	No due to laser surgery	Laser ablation, R angular gyrus	Seizures	
S238	35	35	PMG	No due to laser surgery	Laser ablation, L fronto-parietal operculum, subcentral gyrus	Seizures	

Abbreviations. FCD: focal cortical dysplasia; L: left; PMG: polymicrogyria; PNH: periventricular nodular heterotopia; R: right; SMA: supplemental motor area.

[a]Subjects from 101 to 140 were seizure-free (SF) after the surgery and subjects from 215 to 238 were nonseizure-free (NSF).

[b]Patients had seizures initiated from different area than fast activity, which influenced the surgery planning.

[c]Sparse implantation with inadequate sampling of the epileptogenic zone.

Appendix B
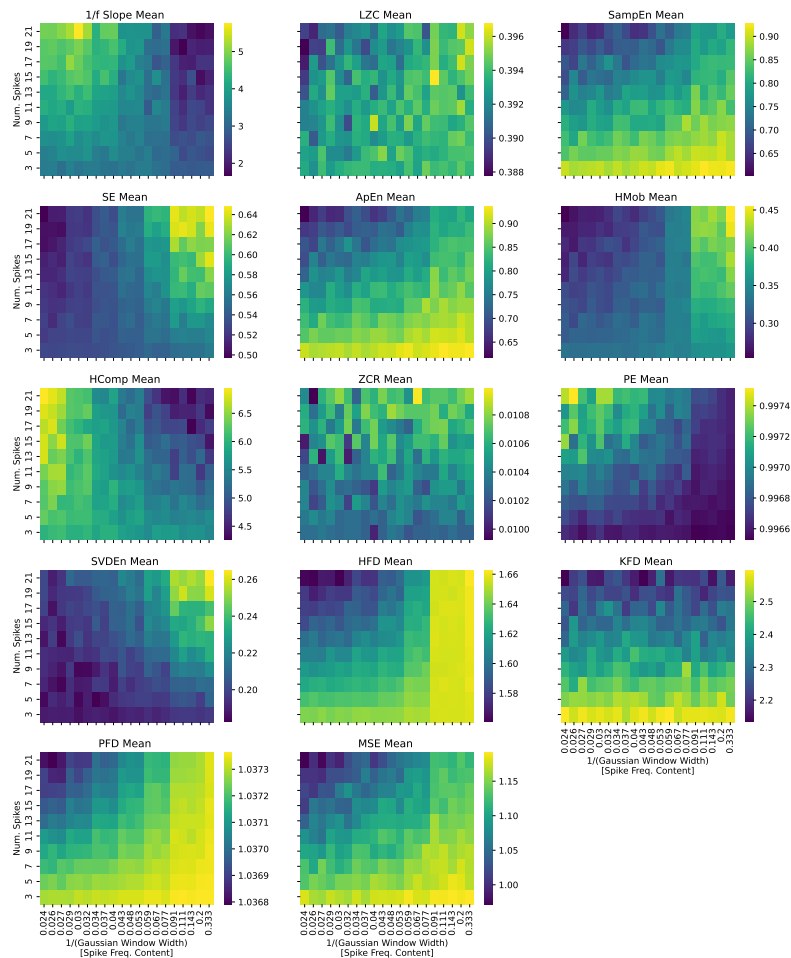
Complementary Spiking Analysis

Figure B.1: Heatmaps for the Mean of the Features across the two spiking parameters: Width of the Gaussian window and number of spikes. Note that the width has an inverse relationship to the frequency content of the spike. For ease of interpretation, the width axis corresponds to the inverse of the width so that a high frequency content is at the right border of the axis.
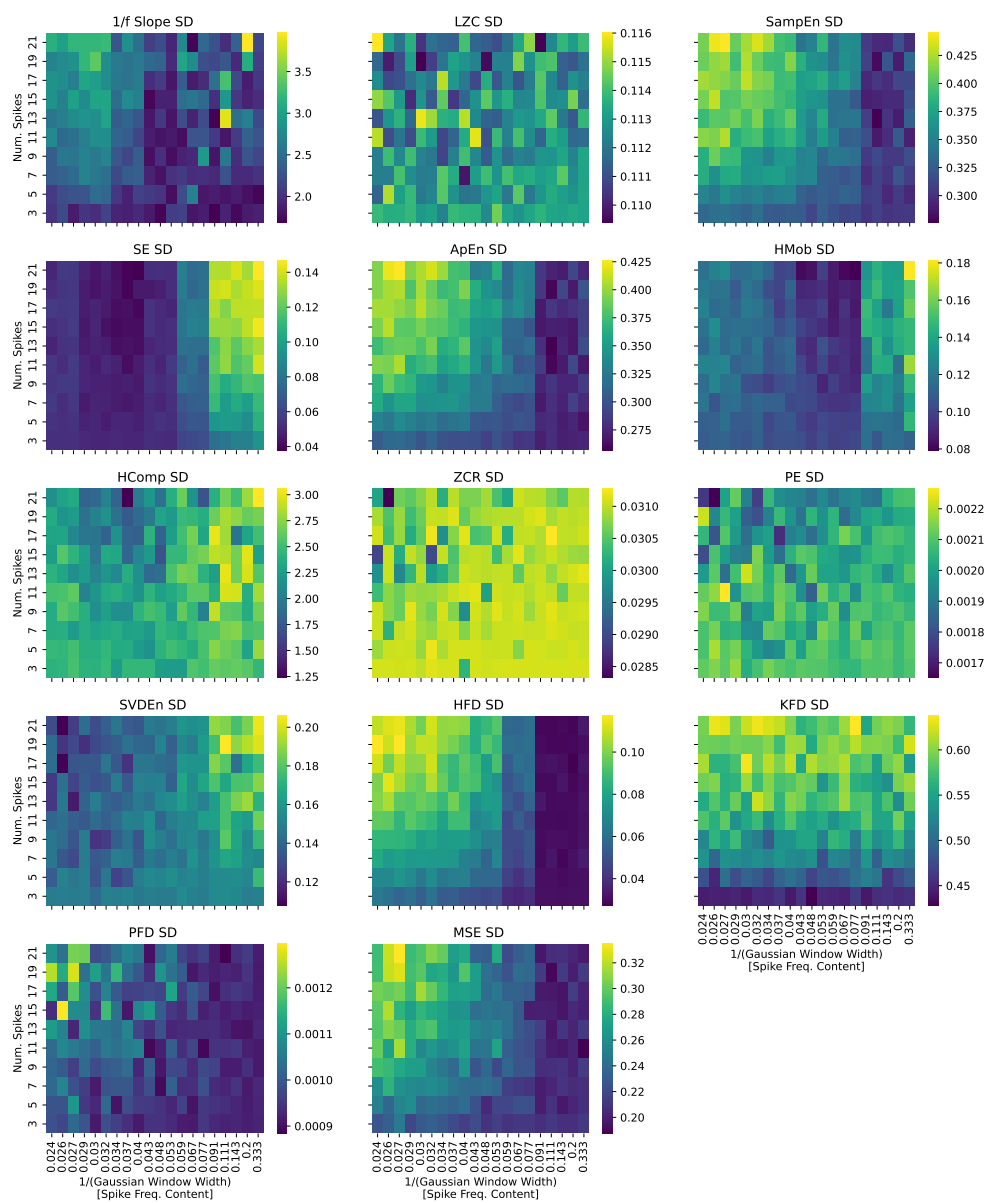
Figure B.2: Heatmaps for the Standard Deviation of the Features across the two spiking parameters: Width of the Gaussian window and number of spikes. Note that the width has an inverse relationship to the frequency content of the spike. For ease of interpretation, the width axis corresponds to the inverse of the width so that a high frequency content is at the right border of the axis.
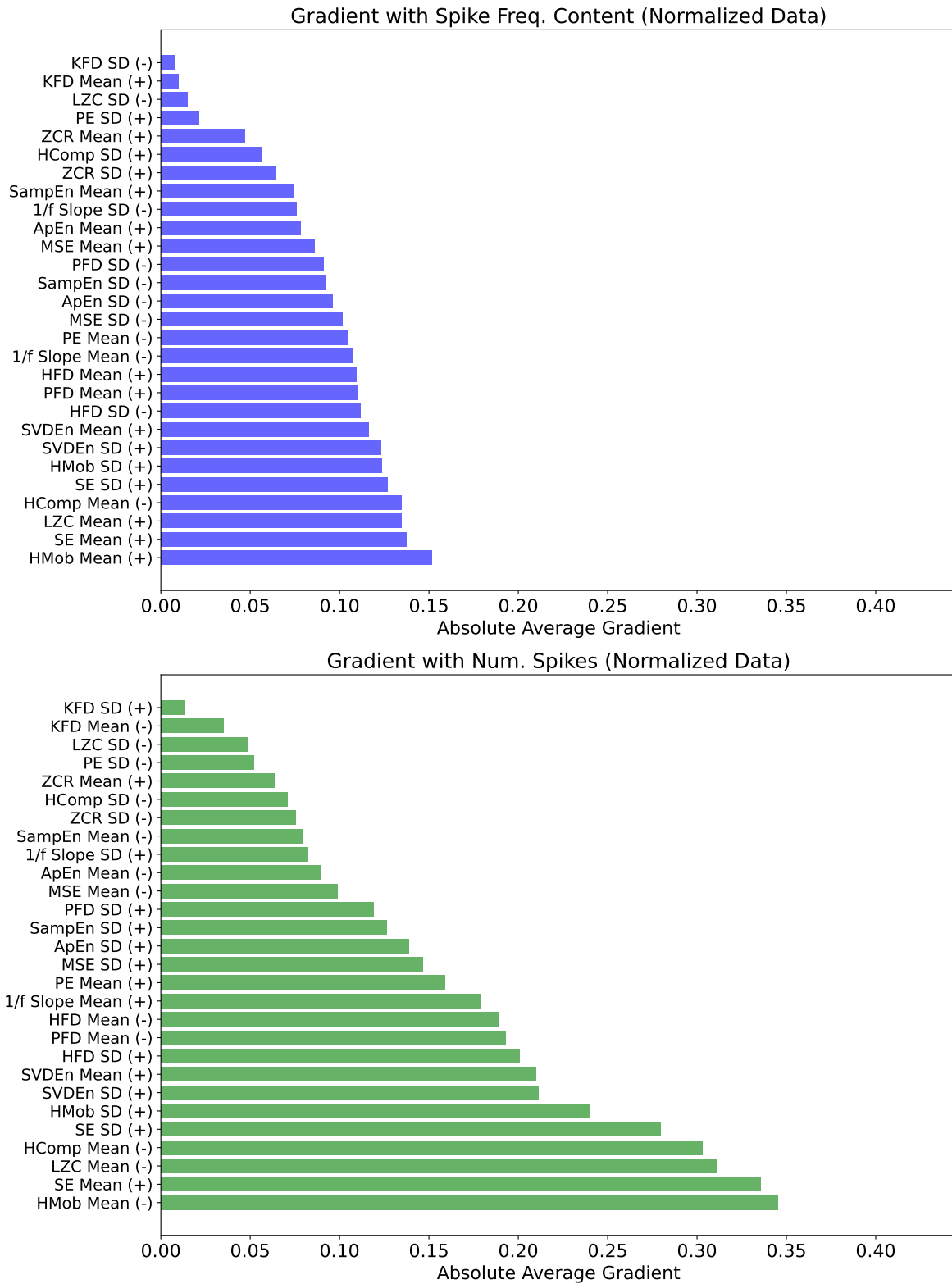
Figure B.3: Absolute Average gradient of each feature in respect to the spike parameters. Prior to the calculation of the gradient the features were z-scored to avoid scaling effects. To the right of each feature the sign of the gradient is indicated (±).

Appendix C

Confusion Matrices per Fold

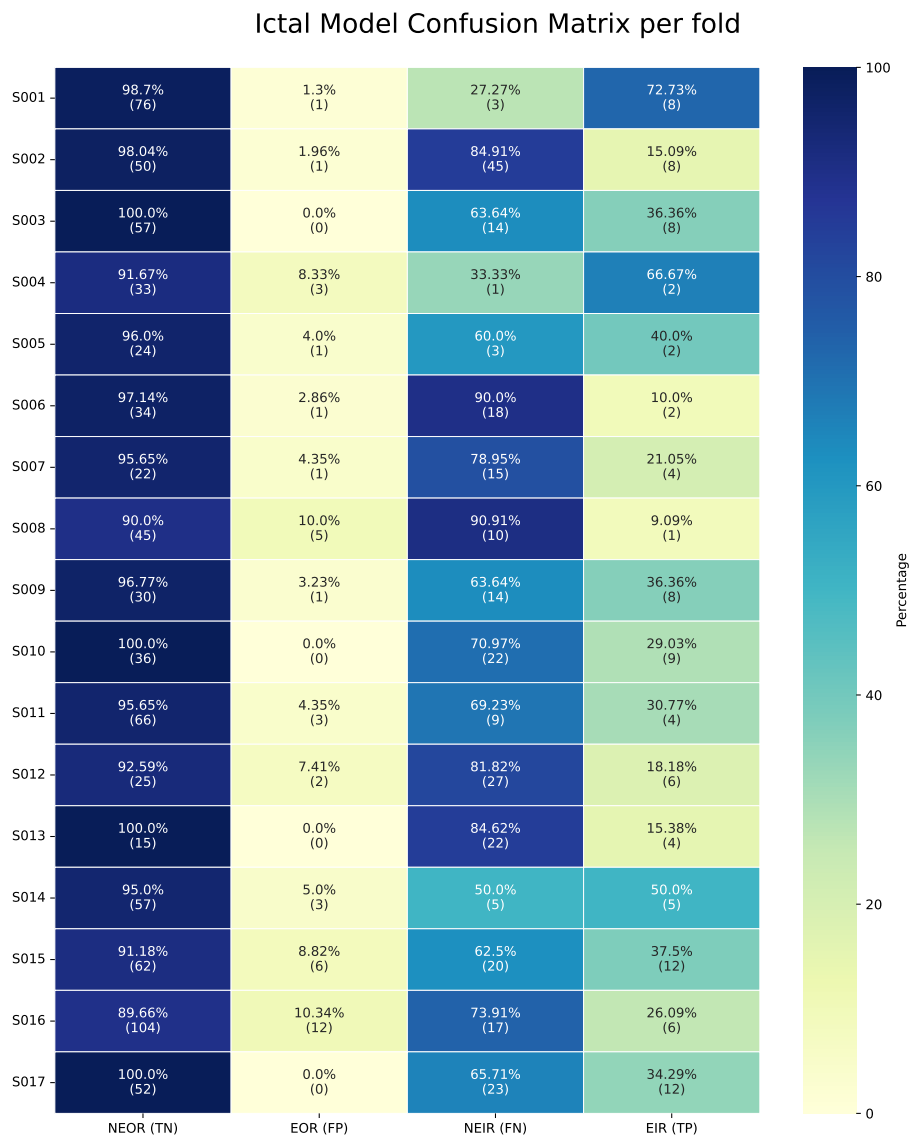## C.1   Normalized by Resection Class Count

Figure C.1: Confusion Matrix for the Ictal model across the different subject-wise folds, normalized by the resection class count.
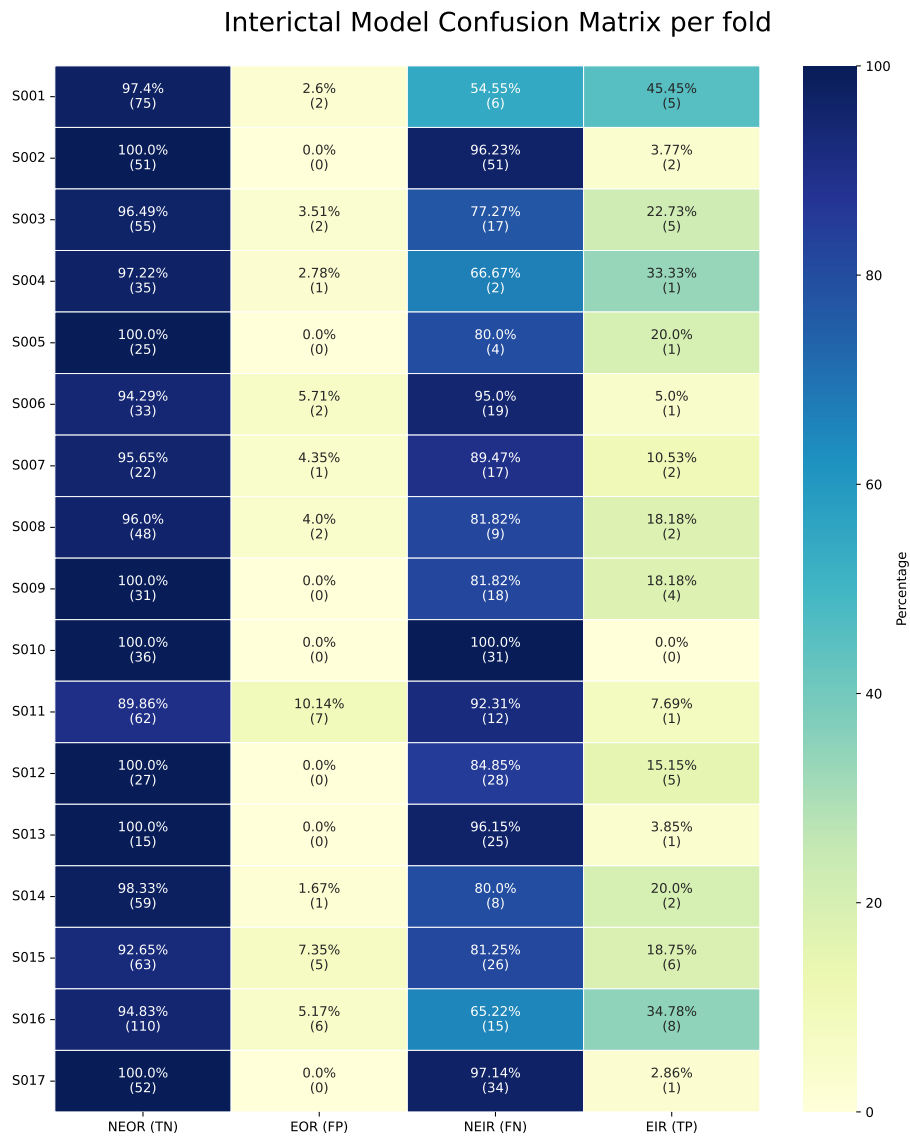
Figure C.2: Confusion Matrix for the Interictal model across the different subject-wise folds, normalized by the resection class count.
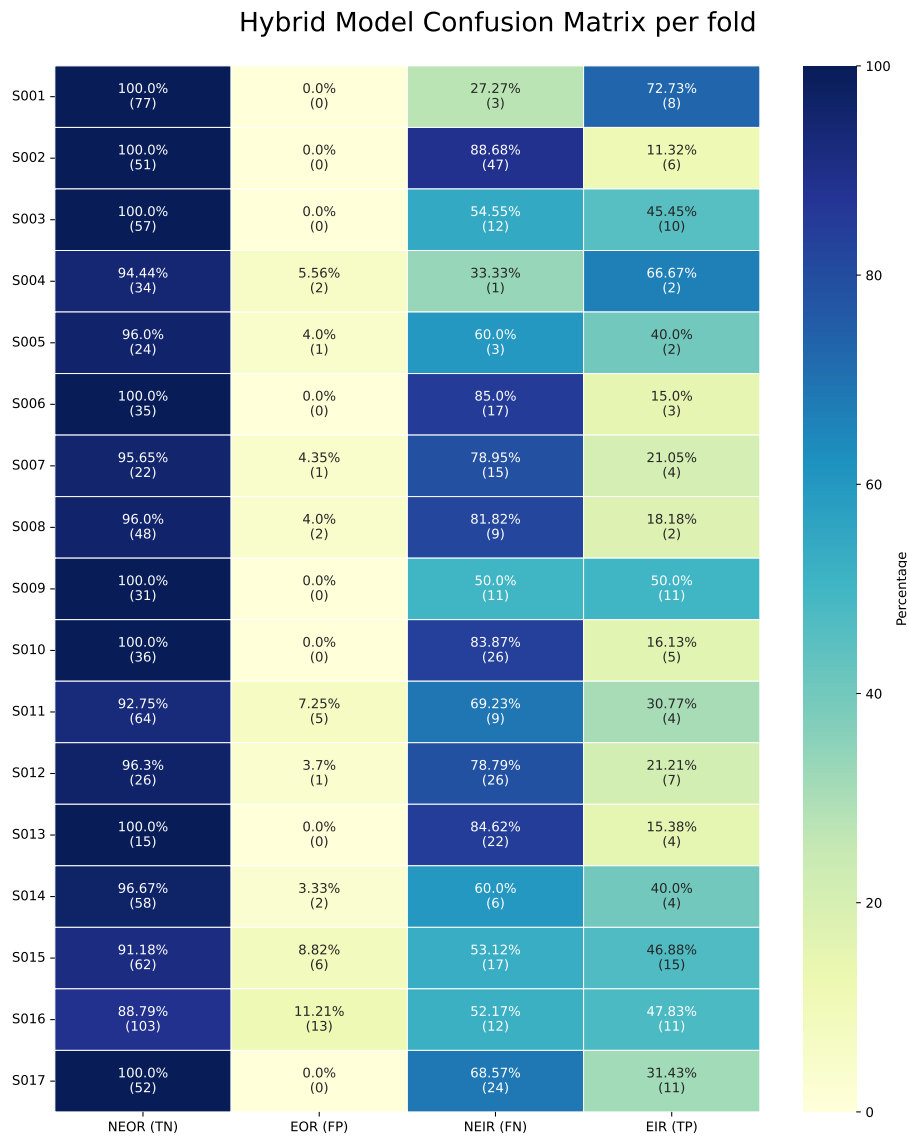
Figure C.3: Confusion Matrix for the Hybrid model across the different subject-wise folds, normalized by the resection class count.

Figure C.4: Confusion Matrix for the Benchmark model across the different folds, normalized by the resection class count.

C.2   Normalized by Predicted Class Count

Figure C.5: Confusion Matrix for the Ictal model across the different subject-wise folds, normalized by the predicted class count.

Figure C.6: Confusion Matrix for the Interictal model across the different subject-wise folds, normalized by the predicted class count.

Hybrid Model Confusion Matrix per fold

Figure C.7: Confusion Matrix for the Hybrid model across the different subject-wise folds, normalized by the predicted class count.
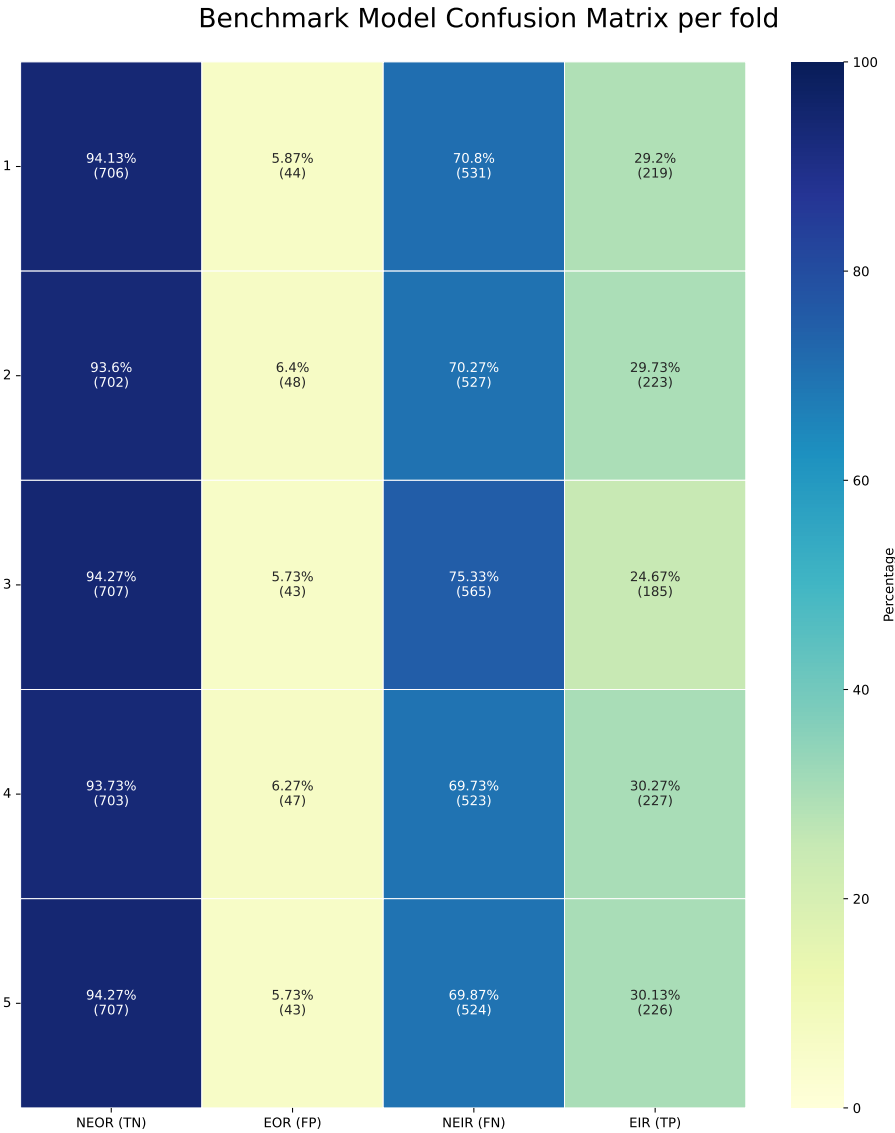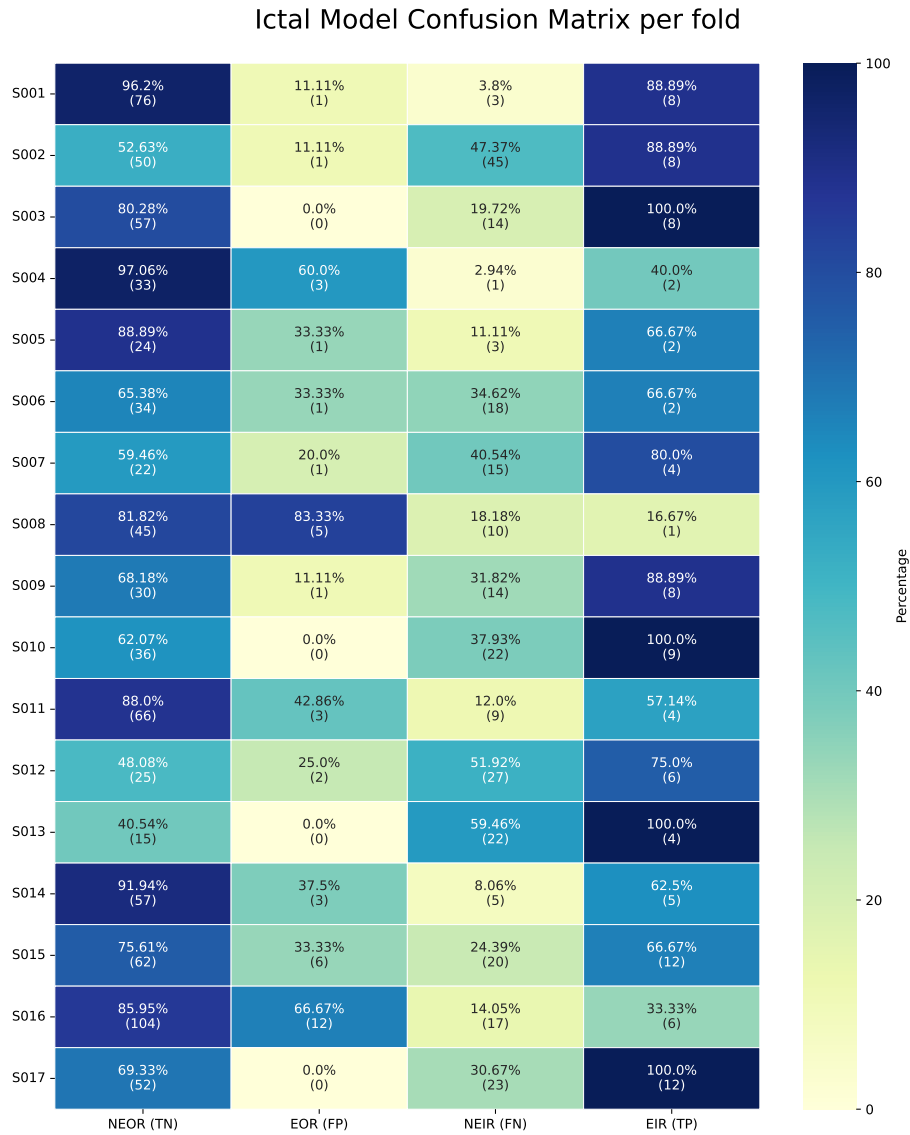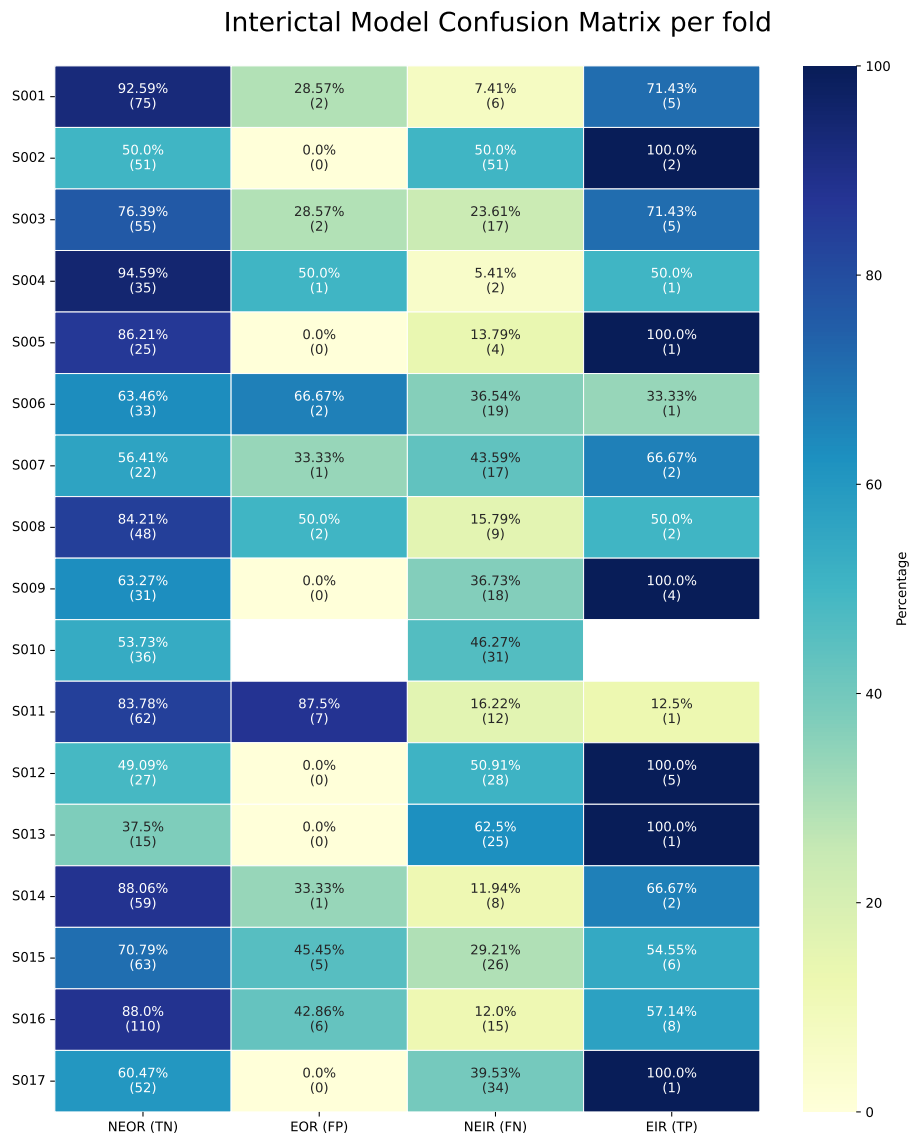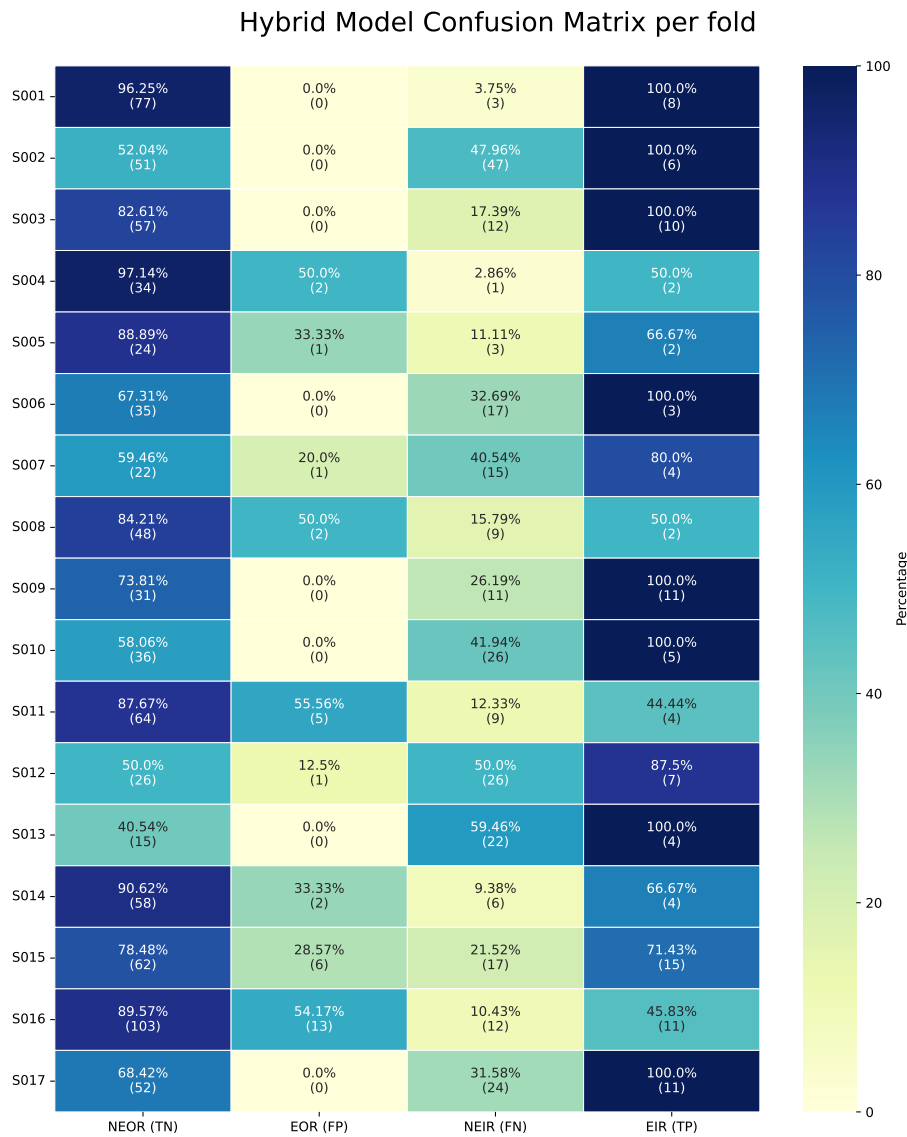
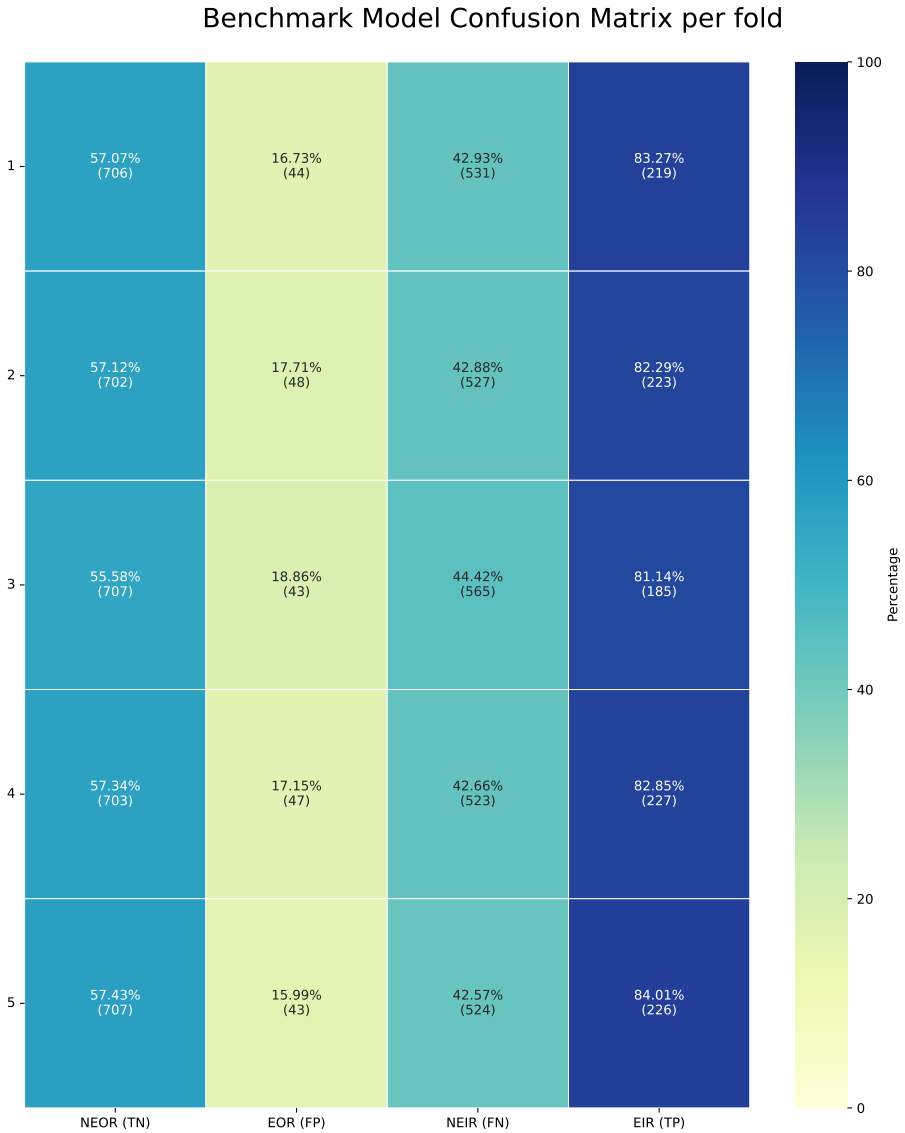Benchmark Model Confusion Matrix per fold

Figure C.8: Confusion Matrix for the Benchmark model across the different folds, normalized by the predicted class count.

References

[1] M. S. Akter, M. R. Islam, Y. Iimura, H. Sugano, K. Fukumori, D. Wang, T. Tanaka, and A. Cichocki. Multiband entropy-based feature-extraction method for automatic identification of epileptic focus based on high-frequency components in interictal iEEG. Scientific Reports, 10(1), Apr. 2020. doi: 10.1038/s41598-020-62967-z. URL https://doi.org/10.1038/s41598-020-62967-z. (page 12)

[2] O. Alter, P. O. Brown, and D. Botstein. Singular value decomposition for genome-wide expression data processing and modeling. Proceedings of the National Academy of Sciences, 97(18):10101–10106, Aug. 2000. doi: 10.1073/pnas.97.18. 10101. URL https://doi.org/10.1073/pnas.97.18.10101. (page 17)

[3] L. P. Andrade-Valenca, F. Dubeau, F. Mari, R. Zelmann, and J. Gotman. Interictal scalp fast oscillations as a marker of the seizure onset zone. Neurology, 77(6):524–531, July 2011. doi: 10.1212/wnl.0b013e318228bee2. URL https://doi.org/10.1212/wnl.0b013e318228bee2. (page 1)

[4] C. Bandt and B. Pompe. Permutation entropy: A natural complexity measure for time series. Physical Review Letters, 88(17), Apr. 2002. doi: 10.1103/physrevlett. 88.174102. URL https://doi.org/10.1103/physrevlett.88.174102. (page 15)

[5] C. Baumgartner and S. Pirker. Video-EEG. In Clinical Neurophysiology: Basis and Technical Aspects, pages 171–183. Elsevier, 2019. doi: 10.1016/b978-0-444-64032-1. 00011-4. URL https://doi.org/10.1016/b978-0-444-64032-1.00011-4. (page 8)

[6] E. Beghi, G. Giussani, E. Nichols, F. Abd-Allah, J. Abdela, A. Abdelalim, H. N. Abraha, M. G. Adib, S. Agrawal, F. Alahdab, A. Awasthi, Y. Ayele, M. A. Barboza, A. B. Belachew, B. Biadgo, A. Bijani, H. Bitew, F. Carvalho, Y. Chaiah, A. Daryani, H. P. Do, M. Dubey, A. Y. Y. Endries, S. Eskandarieh, A. Faro, F. Farzadfar, S.-M. Fereshtehnejad, E. Fernandes, D. O. Fijabi, I. Filip, F. Fischer, A. K. Gebre, A. G. Tsadik, T. G. Gebremichael, K. E. Gezae, M. Ghasemi-Kasman, K. G. Weldegwergs, M. G. Degefa, E. V. Gnedovskaya, T. B. Hagos, A. Haj-Mirzaian, A. Haj-Mirzaian, H. Y. Hassen, S. I. Hay, M. Jakovljevic, A. Kasaeian, T. D. Kassa, Y. S. Khader, I. Khalil, E. A. Khan, J. Khubchandani, A. Kisa, K. J. Krohn, C. Kulkarni, Y. L. Nirayo, M. T. Mackay, M. Majdan, A. Majeed, T. Manhertz, M. M. Mehndiratta, T. Mekonen, H. G. Meles, G. Mengistu, S. Mohammed, M. Naghavi, A. H. Mokdad, G. Mustafa, S. S. N. Irvani, L. H. Nguyen, M. R. Nixon, F. A. Ogbo, A. T. Olagunju, T. O. Olagunju, M. O. Owolabi, M. R. Phillips, G. D. Pinilla-Monsalve, M. Qorbani, A. Radfar, A. Rafay, V. Rahimi-Movaghar, N. Reinig, P. S. Sachdev, H. Safari, S. Safari, S. Safiri, M. A. Sahraian, A. M. Samy, S. Sarvi, M. Sawhney, M. A. Shaikh, M. Sharif, G. Singh, M. Smith, C. E. I. Szoeke, R. Tabarés-Seisdedos,

M.-H. Temsah, O. Temsah, M. Tortajada-Girbés, B. X. Tran, A. A. T. Tsegay, I. Ullah, N. Venketasubramanian, R. Westerman, A. S. Winkler, E. M. Yimer, N. Yonemoto, V. L. Feigin, T. Vos, and C. J. L. Murray. Global, regional, and national burden of epilepsy, 1990–2016: a systematic analysis for the global burden of disease study 2016. The Lancet Neurology, 18(4):357–375, Apr. 2019. doi: 10.1016/s1474-4422(18)30454-x. URL https://doi.org/10.1016/s1474-4422(18)30454-x. (page 1)

[7] D. Bzdok, N. Altman, and M. Krzywinski. Statistics versus machine learning. Nature Methods, 15(4):233–234, Apr. 2018. doi: 10.1038/nmeth.4642. URL https://doi.org/10.1038/nmeth.4642. (page 72)

[8] T. Cecchin, R. Ranta, L. Koessler, O. Caspary, H. Vespignani, and L. Maillard. Seizure lateralization in scalp EEG using hjorth parameters. Clinical Neurophysiology, 121(3):290–300, Mar. 2010. doi: 10.1016/j.clinph.2009.10.033. URL https://doi.org/10.1016/j.clinph.2009.10.033. (page 21)

[9] G. J. Chaitin. On the simplicity and speed of programs for computing infinite sets of natural numbers. Journal of the ACM, 16(3):407–422, July 1969. doi: 10.1145/321526.321530. URL https://doi.org/10.1145/321526.321530. (page 20)

[10] B. S. Chang and D. H. Lowenstein. Epilepsy. New England Journal of Medicine, 349(13):1257–1266, Sept. 2003. doi: 10.1056/nejmra022308. URL https://doi.org/10.1056/nejmra022308. (page 4)

[11] M. Costa, A. L. Goldberger, and C.-K. Peng. Multiscale entropy analysis of complex physiologic time series. Physical Review Letters, 89(6), July 2002. doi: 10.1103/physrevlett.89.068102. URL https://doi.org/10.1103/physrevlett.89.068102. (page 19)

[12] M. Costa, A. L. Goldberger, and C.-K. Peng. Multiscale entropy analysis of biological signals. Physical Review E, 71(2), Feb. 2005. doi: 10.1103/physreve.71.021906. URL https://doi.org/10.1103/physreve.71.021906. (page 19)

[13] S. Cranstoun, G. Worrell, J. Echauz, and B. Litt. Self-organized criticality in the epileptic brain. In Proceedings of the Second Joint 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society] [Engineering in Medicine and Biology. IEEE. doi: 10.1109/iembs.2002.1134468. URL https://doi.org/10.1109/iembs.2002.1134468. (page 9)

[14] A. Delgado-Bonal and A. Marshak. Approximate entropy and sample entropy: A comprehensive tutorial. Entropy, 21(6):541, May 2019. doi: 10.3390/e21060541. URL https://doi.org/10.3390/e21060541. (page 14)

[15] O. Devinsky, A. Vezzani, T. J. O'Brien, N. Jette, I. E. Scheffer, M. de Curtis, and P. Perucca. Epilepsy. Nature Reviews Disease Primers, 4(1), May 2018. doi: 10.1038/nrdp.2018.24. URL https://doi.org/10.1038/nrdp.2018.24. (page 1)

[16] T. Donoghue, M. Haller, E. J. Peterson, P. Varma, P. Sebastian, R. Gao, T. Noto, A. H. Lara, J. D. Wallis, R. T. Knight, A. Shestyuk, and B. Voytek. Parameterizing neural power spectra into periodic and aperiodic components. Nature Neuroscience,

23(12):1655–1665, Nov. 2020. doi: 10.1038/s41593-020-00744-x. URL https://doi.org/10.1038/s41593-020-00744-x. (pages 2 and 36)

[17] S. Elgohary, S. Eldawlatly, and M. I. Khalil. Epileptic seizure prediction using zero-crossings analysis of EEG wavelet detail coefficients. In 2016 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB). IEEE, Oct. 2016. doi: 10.1109/cibcb.2016.7758115. URL https://doi.org/10.1109/cibcb.2016.7758115. (page 21)

[18] J. Engel. Excitation and inhibition in epilepsy. Canadian Journal of Neurological Sciences / Journal Canadien des Sciences Neurologiques, 23(3):167–174, 1996. doi: 10.1017/S0317167100038464. (page 10)

[19] J. Engel. Approaches to refractory epilepsy. Annals of Indian Academy of Neurology, 17(5):12, 2014. doi: 10.4103/0972-2327.128644. URL https://doi.org/10.4103/0972-2327.128644. (page 1)

[20] K. M. Fiest, K. M. Sauro, S. Wiebe, S. B. Patten, C.-S. Kwon, J. Dykeman, T. Pringsheim, D. L. Lorenzetti, and N. Jetté. Prevalence and incidence of epilepsy. Neurology, 88(3):296–303, Dec. 2016. doi: 10.1212/wnl.0000000000003509. URL https://doi.org/10.1212/wnl.0000000000003509. (page 1)

[21] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. Engel, L. Forsgren, J. A. French, M. Glynn, D. C. Hesdorffer, B. Lee, G. W. Mathern, S. L. Moshé, E. Perucca, I. E. Scheffer, T. Tomson, M. Watanabe, and S. Wiebe. ILAE official report: A practical clinical definition of epilepsy. Epilepsia, 55(4):475–482, Apr. 2014. doi: 10.1111/epi.12550. URL https://doi.org/10.1111/epi.12550. (page 4)

[22] R. Gao, E. J. Peterson, and B. Voytek. Inferring synaptic excitation/inhibition balance from field potentials. NeuroImage, 158:70–78, 2017. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2017.06.078. URL https://www.sciencedirect.com/science/article/pii/S1053811917305621. (page 10)

[23] A. Gramfort, M. Luessi, E. Larson, D. A. Engemann, D. Strohmeier, C. Brodbeck, R. Goj, M. Jas, T. Brooks, L. Parkkonen, and M. S. Hämäläinen. MEG and EEG data analysis with MNE-Python. Frontiers in Neuroscience, 7(267):1–13, 2013. doi: 10.3389/fnins.2013.00267. (pages 35 and 36)

[24] K. M. Grande, S. K. Z. Ihnen, and R. Arya. Electrical stimulation mapping of brain function: A comparison of subdural electrodes and stereo-eeg. Frontiers in Human Neuroscience, 14, 2020. ISSN 1662-5161. doi: 10.3389/fnhum.2020.611291. URL https://www.frontiersin.org/articles/10.3389/fnhum.2020.611291. (page 8)

[25] P. Grassberger. Randomness, information, and complexity, 2012. (page 12)

[26] O. Grinenko, J. Li, J. C. Mosher, I. Z. Wang, J. C. Bulacio, J. Gonzalez-Martinez, D. Nair, I. Najm, R. M. Leahy, and P. Chauvel. A fingerprint of the epileptogenic zone in human epilepsies. Brain, 141(1):117–131, 12 2017. ISSN 0006-8950. doi: 10.1093/brain/awx306. URL https://doi.org/10.1093/brain/awx306. (pages 1, 7, 41, 43, 73, 75, and 77)

[27] A. Hagemann, J. Wilting, B. Samimizad, F. Mormann, and V. Priesemann. Assessing criticality in pre-seizure single-neuron activity of human epileptic cortex. PLOS Computational Biology, 17(3):e1008773, Mar. 2021. doi: 10.1371/journal.pcbi.1008773. URL https://doi.org/10.1371/journal.pcbi.1008773. (page 9)

[28] B. J. He. Scale-free properties of the functional magnetic resonance imaging signal during rest and task. The Journal of Neuroscience, 31(39):13786–13795, Sept. 2011. doi: 10.1523/jneurosci.2111-11.2011. URL https://doi.org/10.1523/jneurosci.2111-11.2011. (page 10)

[29] B. J. He. Scale-free brain activity: past, present, and future. Trends in Cognitive Sciences, 18(9):480–487, 2014. ISSN 1364-6613. doi: https://doi.org/10.1016/j.tics.2014.04.003. URL https://www.sciencedirect.com/science/article/pii/S1364661314000850. (pages 2 and 10)

[30] J. Hesse and T. Gross. Self-organized criticality as a fundamental property of neural systems. Frontiers in Systems Neuroscience, 8, 2014. ISSN 1662-5137. doi: 10.3389/fnsys.2014.00166. URL https://www.frontiersin.org/articles/10.3389/fnsys.2014.00166. (page 9)

[31] T. Higuchi. Approach to an irregular time series on the basis of the fractal theory. Physica D: Nonlinear Phenomena, 31(2):277–283, 1988. ISSN 0167-2789. doi: https://doi.org/10.1016/0167-2789(88)90081-4. URL https://www.sciencedirect.com/science/article/pii/0167278988900814. (page 23)

[32] B. Hjorth. EEG analysis based on time domain properties. Electroencephalography and Clinical Neurophysiology, 29(3):306–310, Sept. 1970. doi: 10.1016/0013-4694(70)90143-4. URL https://doi.org/10.1016/0013-4694(70)90143-4. (page 21)

[33] K. IIDA and H. OTSUBO. Stereoelectroencephalography: Indication and efficacy. Neurologia medico-chirurgica, 57(8):375–385, 2017. doi: 10.2176/nmc.ra.2017-0008. URL https://doi.org/10.2176/nmc.ra.2017-0008. (page 8)

[34] T. Inouye, K. Shinosaki, H. Sakamoto, S. Toi, S. Ukai, A. Iyama, Y. Katsuda, and M. Hirano. Quantification of eeg irregularity by use of the entropy of the power spectrum. Electroencephalography and Clinical Neurophysiology, 79(3):204–210, 1991. ISSN 0013-4694. doi: https://doi.org/10.1016/0013-4694(91)90138-T. URL https://www.sciencedirect.com/science/article/pii/001346949190138T. (page 17)

[35] M. R. Islam, X. Zhao, Y. Miao, H. Sugano, and T. Tanaka. Epileptic seizure focus detection from interictal electroencephalogram: a survey. Cognitive Neurodynamics, May 2022. doi: 10.1007/s11571-022-09816-z. URL https://doi.org/10.1007/s11571-022-09816-z. (pages 1, 12, 39, 72, and 74)

[36] L. Jehi. The epileptogenic zone: Concept and definition. Epilepsy Currents, 18(1):12–16, 2018. doi: 10.5698/1535-7597.18.1.12. URL https://doi.org/10.5698/1535-7597.18.1.12. PMID: 29844752. (page 5)

[37] H. Jiang, V. Kokkinos, S. Ye, A. Urban, A. Bagić, M. Richardson, and B. He. Interictal SEEG resting-state connectivity localizes the seizure onset zone and predicts seizure outcome. Advanced Science, 9(18):2200887, May 2022. doi: 10.1002/advs.202200887. URL https://doi.org/10.1002/advs.202200887.
(page 10)

[38] B. C. Jobst, F. Bartolomei, B. Diehl, B. Frauscher, P. Kahane, L. Minotti, A. Sharan, N. Tardy, G. Worrell, and J. Gotman. Intracranial EEG in the 21st century. Epilepsy Currents, 20(4):180–188, July 2020. doi: 10.1177/1535759720934852. URL https://doi.org/10.1177/1535759720934852.
(page 8)

[39] M. J. Katz. Fractals and the analysis of waveforms. Computers in Biology and Medicine, 18(3):145–156, Jan. 1988. doi: 10.1016/0010-4825(88)90041-8. URL https://doi.org/10.1016/0010-4825(88)90041-8.
(page 22)

[40] G. Kaushik, P. Gaur, R. R. Sharma, and R. B. Pachori. EEG signal based seizure detection focused on hjorth parameters from tunable-q wavelet sub-bands. Biomedical Signal Processing and Control, 76:103645, July 2022. doi: 10.1016/j.bspc.2022.103645. URL https://doi.org/10.1016/j.bspc.2022.103645.
(page 21)

[41] A. Kolmogorov. On tables of random numbers. Theoretical Computer Science, 207 (2):387–395, Nov. 1998. doi: 10.1016/s0304-3975(98)00075-9. URL https://doi.org/10.1016/s0304-3975(98)00075-9.
(page 20)

[42] P. Kwan and J. W. Sander. The natural history of epilepsy: an epidemiological view. Journal of Neurology, Neurosurgery & Psychiatry, 75(10):1376–1381, 2004. ISSN 0022-3050. doi: 10.1136/jnnp.2004.045690. URL https://jnnp.bmj.com/content/75/10/1376.
(page 1)

[43] S. Lagarde, S. Buzori, A. Trebuchon, R. Carron, D. Scavarda, M. Milh, A. McGonigal, and F. Bartolomei. The repertoire of seizure onset patterns in human focal epilepsies: Determinants and prognostic values. Epilepsia, 60(1):85–95, Nov. 2018. doi: 10.1111/epi.14604. URL https://doi.org/10.1111/epi.14604.
(page 4)

[44] Z. J. Lau, T. Pham, S. H. A. Chen, and D. Makowski. Brain entropy, fractal dimensions and predictability: A review of complexity measures for EEG in healthy and neuropsychiatric populations. European Journal of Neuroscience, 56(7):5047–5069, Sept. 2022. doi: 10.1111/ejn.15800. URL https://doi.org/10.1111/ejn.15800.
(page 8)

[45] A. Lempel and J. Ziv. On the complexity of finite sequences. IEEE Transactions on Information Theory, 22(1):75–81, Jan. 1976. doi: 10.1109/tit.1976.1055501. URL https://doi.org/10.1109/tit.1976.1055501.
(page 20)

[46] J. Li, O. Grinenko, J. C. Mosher, J. Gonzalez-Martinez, R. M. Leahy, and P. Chauvel. Learning to define an electrical biomarker of the epileptogenic zone. Human Brain Mapping, 41(2):429–441, Oct. 2019. doi: 10.1002/hbm.24813. URL https://doi.org/10.1002/hbm.24813.
(page 43)

[47] V. Lopantsev, M. Both, and A. Draguhn. Rapid plasticity at inhibitory and excitatory synapses in the hippocampus induced by ictal epileptiform discharges. European Journal of Neuroscience, 29(6):1153–1164, Mar. 2009. doi: 10.1111/j.1460-9568.2009.06663.x. URL https://doi.org/10.1111/j.1460-9568.2009.06663.x. (page 10)

[48] H. O. Lüders. Textbook of Epilepsy Surgery. 07 2008. ISBN 9781841845760. (page 4)

[49] H. O. Lüders, J. Engel Jr, and C. Munari. General principles. Surgical treatment of the epilepsies, pages 137–153, 1993. (page 4)

[50] H. O. Lüders, I. Najm, D. Nair, P. Widdess-Walsh, and W. Bingman. The epileptogenic zone: general principles. Epileptic disorders, 8(2):1–9, 2006. (pages 4, 5, and 6)

[51] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and A. S H Chen. Neurokit2: A python toolbox for neurophysiological signal processing, 2020. URL https://github.com/neuropsychology/NeuroKit. (page 36)

[52] D. Makowski, A. S. Te, T. Pham, Z. J. Lau, and S. H. A. Chen. The structure of chaos: An empirical comparison of fractal physiology complexity indices using NeuroKit2. Entropy, 24(8):1036, July 2022. doi: 10.3390/e24081036. URL https://doi.org/10.3390/e24081036. (page 8)

[53] B. Mandelbrot. How long is the coast of britain? statistical self-similarity and fractional dimension. Science, 156(3775):636–638, May 1967. doi: 10.1126/science.156.3775.636. URL https://doi.org/10.1126/science.156.3775.636. (page 22)

[54] J. McCagh, J. E. Fisk, and G. A. Baker. Epilepsy, psychosocial and cognitive functioning. Epilepsy Research, 86(1):1–14, Sept. 2009. doi: 10.1016/j.eplepsyres.2009.04.007. URL https://doi.org/10.1016/j.eplepsyres.2009.04.007. (page 1)

[55] C. Meisel, A. Storch, S. Hallmeyer-Elgner, E. Bullmore, and T. Gross. Failure of adaptive self-organized criticality during epileptic seizure attacks. PLoS Computational Biology, 8(1):e1002312, Jan. 2012. doi: 10.1371/journal.pcbi.1002312. URL https://doi.org/10.1371/journal.pcbi.1002312. (page 9)

[56] K. J. Miller, L. B. Sorensen, J. G. Ojemann, and M. den Nijs. Power-law scaling in the brain surface electric potential. PLoS Computational Biology, 5(12):e1000609, Dec. 2009. doi: 10.1371/journal.pcbi.1000609. URL https://doi.org/10.1371/journal.pcbi.1000609. (pages 2 and 10)

[57] S. Neeman and A. Maharshak. Order and disorder, entropy in math, science, nature and the arts. In Proceedings of the 3rd WSEAS/IASME International Conference on ENGINEERING EDUCATION, pages 11–13, 2006. (page 12)

[58] S. Noachtar and J. Rémi. The role of EEG in epilepsy: A critical review. Epilepsy &amp Behavior, 15(1):22–33, May 2009. doi: 10.1016/j.yebeh.2009.02.035. URL https://doi.org/10.1016/j.yebeh.2009.02.035. (page 8)

[59] S. Noachtar, P. A. Winkler, and H. O. Lüders. Surgical therapy of epilepsy. In Neurological Disorders, pages 235–244. Elsevier, 2003. doi: 10.1016/b978-012125831-3/50216-1. URL https://doi.org/10.1016/b978-012125831-3/50216-1. (page 8)

[60] J. O'Byrne and K. Jerbi. How critical is brain criticality? Trends in Neurosciences, 45(11):820–837, Nov. 2022. doi: 10.1016/j.tins.2022.08.007. URL https://doi.org/10.1016/j.tins.2022.08.007. (pages 2 and 8)

[61] W. H. Organization et al. Epilepsy: a public health imperative. World Health Organization, 2019. (page 1)

[62] N. Päivinen, S. Lammi, A. Pitkänen, J. Nissinen, M. Penttonen, and T. Grönfors. Epileptic seizure detection: A nonlinear viewpoint. Computer Methods and Programs in Biomedicine, 79(2):151–159, Aug. 2005. doi: 10.1016/j.cmpb.2005.04.006. URL https://doi.org/10.1016/j.cmpb.2005.04.006. (page 21)

[63] F. Panzica, G. Varotto, F. Rotondi, R. Spreafico, and S. Franceschetti. Identification of the epileptogenic zone from stereo-eeg signals: A connectivity-graph theory approach. Frontiers in Neurology, 4, 2013. ISSN 1664-2295. doi: 10.3389/fneur.2013.00175. URL https://www.frontiersin.org/articles/10.3389/fneur.2013.00175. (page 1)

[64] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011. (page 41)

[65] H. Peng, F. Long, and C. Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8):1226–1238, Aug. 2005. doi: 10.1109/tpami.2005.159. URL https://doi.org/10.1109/tpami.2005.159. (pages 31 and 38)

[66] A. Petrosian. Kolmogorov complexity of finite sequences and recognition of different preictal eeg patterns. In Proceedings Eighth IEEE Symposium on Computer-Based Medical Systems, pages 212–217, 1995. doi: 10.1109/CBMS.1995.465426. (page 23)

[67] S. M. Pincus and A. L. Goldberger. Physiological time-series analysis: what does regularity quantify? American Journal of Physiology-Heart and Circulatory Physiology, 266(4):H1643–H1656, 1994. doi: 10.1152/ajpheart.1994.266.4.H1643. URL https://doi.org/10.1152/ajpheart.1994.266.4.H1643. PMID: 8184944. (page 14)

[68] S. M. Pincus, I. M. Gladstone, and R. A. Ehrenkranz. A regularity statistic for medical data analysis. Journal of Clinical Monitoring, 7(4):335–345, Oct. 1991. doi: 10.1007/bf01619355. URL https://doi.org/10.1007/bf01619355. (page 13)

[69] B. S. Raghavendra, D. N. Dutt, H. N. Halahalli, and J. P. John. Complexity analysis of EEG in patients with schizophrenia using fractal dimension. Physiological Measurement, 30(8):795–808, June 2009. doi: 10.1088/0967-3334/30/8/005. URL https://doi.org/10.1088/0967-3334/30/8/005. (page 22)

[70] S. Raghu, N. Sriraam, and A. S. Hegde. Features ranking for the classification of epileptic seizure from temporal EEG. In 2016 International Conference on Circuits, Controls, Communications and Computing (I4C). IEEE, Oct. 2016. doi: 10.1109/cimca.2016.8053309. URL https://doi.org/10.1109/cimca.2016.8053309. (page 21)

[71] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. American Journal of Physiology-Heart and Circulatory Physiology, 278(6):H2039–H2049, June 2000. doi: 10.1152/ajpheart.2000.278.6.h2039. URL https://doi.org/10.1152/ajpheart.2000.278.6.h2039. (page 14)

[72] F. Rosenow and H. O. Lüders. Presurgical evaluation of epilepsy. Brain, 124 (9):1683–1700, 09 2001. ISSN 0006-8950. doi: 10.1093/brain/124.9.1683. URL https://doi.org/10.1093/brain/124.9.1683. (pages 4 and 5)

[73] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5): 206–215, May 2019. doi: 10.1038/s42256-019-0048-x. URL https://doi.org/10.1038/s42256-019-0048-x. (page 41)

[74] M. J. Sánchez-Granero, M. Fernández-Martínez, and J. E. Trinidad-Segovia. Introducing fractal dimension algorithms to calculate the hurst exponent of financial time series. The European Physical Journal B, 85(3), Mar. 2012. doi: 10.1140/epjb/e2012-20803-2. URL https://doi.org/10.1140/epjb/e2012-20803-2. (page 22)

[75] R. Seshadri. Autoviml/featurewiz: Use advanced feature engineering strategies and select the best features from your data set fast with a single line of code. https://github.com/AutoViML/featurewiz, 2020. (page 38)

[76] C. E. Shannon. A mathematical theory of communication. The Bell System Technical Journal, 27(3):379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x. (page 12)

[77] R. Solomonoff. A formal theory of inductive inference. part II. Information and Control, 7(2):224–254, June 1964. doi: 10.1016/s0019-9958(64)90131-7. URL https://doi.org/10.1016/s0019-9958(64)90131-7. (page 20)

[78] T. Strydom, G. V. D. Riva, and T. Poisot. SVD entropy reveals the high complexity of ecological networks. Frontiers in Ecology and Evolution, 9, June 2021. doi: 10.3389/fevo.2021.623141. URL https://doi.org/10.3389/fevo.2021.623141. (page 18)

[79] C. Symonds. Excitation And Inhibition In Epilepsy. Brain, 82(2):133–146, 06 1959. ISSN 0006-8950. doi: 10.1093/brain/82.2.133. URL https://doi.org/10.1093/brain/82.2.133. (page 10)

[80] F. Takens. Detecting strange attractors in turbulence. In Lecture Notes in Mathematics, pages 366–381. Springer Berlin Heidelberg, 1981. doi: 10.1007/bfb0091924. URL https://doi.org/10.1007/bfb0091924. (page 16)

[81] E. Tan, S. Algar, D. Corrêa, M. Small, T. Stemler, and D. Walker. Selecting embedding delays: An overview of embedding techniques and a new method using persistent homology. Chaos: An Interdisciplinary Journal of Nonlinear Science, 33 (3), Mar. 2023. doi: 10.1063/5.0137223. URL https://doi.org/10.1063/5.0137223. (page 16)

[82] F. Tang, A. M. S. Hartz, and B. Bauer. Drug-resistant epilepsy: Multiple hypotheses, few answers. Frontiers in Neurology, 8, 2017. ISSN 1664-2295. doi: 10.3389/fneur.2017.00301. URL https://www.frontiersin.org/articles/10.3389/fneur.2017.00301. (page 1)

[83] M. Tanveer, R. B. Pachori, and N. V. Angami. Classification of seizure and seizure-free EEG signals using hjorth parameters. In 2018 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, Nov. 2018. doi: 10.1109/ssci.2018.8628651. URL https://doi.org/10.1109/ssci.2018.8628651. (page 21)

[84] D. Toker, I. Pappas, J. D. Lendner, J. Frohlich, D. M. Mateos, S. Muthukumaraswamy, R. Carhart-Harris, M. Paff, P. M. Vespa, M. M. Monti, F. T. Sommer, R. T. Knight, and M. D'Esposito. Consciousness is supported by near-critical slow cortical electrodynamics. Proceedings of the National Academy of Sciences, 119(7), Feb. 2022. doi: 10.1073/pnas.2024455119. URL https://doi.org/10.1073/pnas.2024455119. (pages 20 and 75)

[85] V. N. Vakharia, J. S. Duncan, J.-A. Witt, C. E. Elger, R. Staba, and J. Engel. Getting the best outcomes from epilepsy surgery. Annals of Neurology, 83(4): 676–690, Apr. 2018. doi: 10.1002/ana.25205. URL https://doi.org/10.1002/ana.25205. (page 5)

[86] R. Vallat. Antropy: Entropy and complexity of (eeg) time-series in python. https://github.com/raphaelvallat/antropy, 2022. (page 36)

[87] A. Vera-González. Pathophysiological mechanisms underlying the etiologies of seizures and epilepsy. In Epilepsy. Exon Publications, Apr. 2022. doi: 10.36255/exon-publications-epilepsy-pathopysiology. URL https://doi.org/10.36255/exon-publications-epilepsy-pathopysiology. (page 4)

[88] F. von Wegner, M. Wiemers, G. Hermann, I. Tödt, E. Tagliazucchi, and H. Laufs. Complexity measures for EEG microstate sequences - concepts and algorithms. May 2023. doi: 10.21203/rs.3.rs-2878411/v1. URL https://doi.org/10.21203/rs.3.rs-2878411/v1. (page 75)

[89] Y. Wang, Y. Yang, S. Li, Z. Su, J. Guo, P. Wei, J. Huang, G. Kang, and G. Zhao. Automatic localization of seizure onset zone based on multi-epileptogenic biomarkers analysis of single-contact from interictal SEEG. Bioengineering, 9(12):769, Dec. 2022. doi: 10.3390/bioengineering9120769. URL https://doi.org/10.3390/bioengineering9120769. (page 72)

[90] B. L. Welch. The generalization of `student's' problem when several different population variances are involved. Biometrika, 34(1/2):28, Jan. 1947. doi: 10.2307/2332510. URL https://doi.org/10.2307/2332510. (page 38)

[91] C. Witton, S. V. Sergeyev, E. G. Turitsyna, P. L. Furlong, S. Seri, M. Brookes, and S. K. Turitsyn. Rogue bioelectrical waves in the brain: the hurst exponent as a potential measure for presurgical mapping in epilepsy. Journal of Neural Engineering, 16(5):056019, Aug. 2019. doi: 10.1088/1741-2552/ab225e. URL https://doi.org/10.1088/1741-2552/ab225e. (page 9)

[92] G. A. Worrell, S. D. Cranstoun, J. Echauz, and B. Litt. Evidence for self-organized criticality in human epileptic hippocampus. NeuroReport, 13(16):2017–2021, Nov. 2002. doi: 10.1097/00001756-200211150-00005. URL https://doi.org/10.1097/00001756-200211150-00005. (page 9)

[93] A. C. Yang and S.-J. Tsai. Is mental illness complex? from behavior to brain. Progress in Neuro-Psychopharmacology and Biological Psychiatry, 45:253–257, Aug. 2013. doi: 10.1016/j.pnpbp.2012.09.015. URL https://doi.org/10.1016/j.pnpbp.2012.09.015. (page 8)

[94] A. Zandi, G. Dumont, M. Javidan, and R. Tafreshi. An entropy-based approach to predict seizures in temporal lobe epilepsy using scalp EEG. In 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, Sept. 2009. doi: 10.1109/iembs.2009.5333971. URL https://doi.org/10.1109/iembs.2009.5333971. (page 21)

[95] R. Zanetti, U. Pale, T. Teijeiro, and D. Atienza. Approximate zero-crossing: a new interpretable, highly discriminative and low-complexity feature for EEG and iEEG seizure detection. Journal of Neural Engineering, 19(6):066018, Nov. 2022. doi: 10.1088/1741-2552/aca1e4. URL https://doi.org/10.1088/1741-2552/aca1e4. (page 21)

[96] Y. Zhang, J. Hao, C. Zhou, and K. Chang. Normalized lempel-ziv complexity and its application in bio-sequence analysis. Journal of Mathematical Chemistry, 46(4):1203–1212, Dec. 2008. doi: 10.1007/s10910-008-9512-2. URL https://doi.org/10.1007/s10910-008-9512-2. (page 20)

[97] V. Zimmern. Why brain criticality is clinically relevant: A scoping review. Frontiers in Neural Circuits, 14, Aug. 2020. doi: 10.3389/fncir.2020.00054. URL https://doi.org/10.3389/fncir.2020.00054. (page 8)