POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Pattern Recognition in Brain Networks to Characterize Preictal States

Tesi di Laurea Magistrale in
Biomedical Engineering - Ingegneria Biomedica

Author: **Laura Daniela Melgarejo Quiñones**

Student ID: 10850175
Advisor: Prof. Luca Mainardi, Prof.ssa Catalina Alvarado Rojas
Co-advisors: Dr. Guadalupe García Isla, Ing. Andrea Farabbi
Academic Year: 2022-23

Abstract

Epilepsy is one of the most common neurological disorders that affects over 50 million people worldwide. It is characterized by recurrent seizures, which are the result of excessive electrical discharges that generate disruptions in brain activity. In most epileptic patients the seizures are infrequent and so they are of unpredictable occurrence. Therefore, antiseizure medicines can reduce the number of seizure incidences in the patient. Unfortunately, for 30% of the people with epilepsy, seizures persist despite the use of this treatment, increasing the risk of injuries, premature death and reducing the patient's quality of life. The aim of the present work is to develop a patient-specific AI-based pipeline integrating functional connectivity analysis applied to epileptic brain networks, to accurately identify the non-seizure period (interictal) and the time interval immediately preceding the seizure onset (preictal state), and therefore detect a potential seizure onset.

In the present study, 17 patients, from the EPILEPSIAE database, with recordings available for at least 8-9 hours before the seizure onset were selected. The patient's recordings were sEEG. Feature Extraction, Feature Preprocessing, Feature Selection, Machine Learning and Deep Learning Application, and Model training and Evaluation, constitute the main pipeline blocks. Five Machine Learning and Deep Learning algorithms were selected for evaluation: Random Forest (RF), Support Vector Machines Classifier (SVC), XGBoost (XGB), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM). To enhance the performance of the classification algorithms to different temporal contexts, three preictal windows of 40, 60, and 80 minutes were evaluated. An F1 score greater than 60% was achieved by 11/17 patients, and with the preferred preictal window among subjects being the 80-minute interval.

Keywords: Epilepsy, Preictal states characterization, EEG, Graph Theory, Functional Connectivity, Machine Learning, Deep Learning.

Abstract in lingua italiana

L'epilessia è uno dei disturbi neurologici più comuni che colpisce oltre 50 milioni di persone in tutto il mondo. È caratterizzata da convulsioni ricorrenti, che sono il risultato di scariche elettriche eccessive che generano interruzioni dell'attività cerebrale. Nella maggior parte dei pazienti epilettici le convulsioni sono poco frequenti e quindi sono di insorgenza imprevedibile. Pertanto, i farmaci antiepilettici possono ridurre il numero di episodi di convulsioni nel paziente. Purtroppo, per il 30% delle persone con epilessia, le convulsioni persistono nonostante l'uso di questo trattamento, aumentando il rischio di lesioni, morte prematura e riducendo la qualità della vita del paziente. Lo scopo della present tesi è quello di sviluppare un *AI-based* pipeline specifica per ogni paziente che integri l'analisi della connettività funzionale applicata alle reti cerebrali epilettiche, per identificare con precisione il periodo non convulsivo (interictale) e l'intervallo di tempo immediatamente precedente l'insorgenza della crisi (stato preictale), e quindi rilevare una potenziale insorgenza convulsiva.

Nel presente studio, sono stati selezionati 17 pazienti, dal database EPILEPSIAE, con registrazioni disponibili per almeno 8-9 ore prima dell'insorgenza delle crisi. Le registrazioni del paziente erano sEEG. L'estrazione delle funzionalità, la pre-elaborazione delle funzionalità, la selezione delle funzionalità, l'applicazione di Machine Learning e Deep Learning, l'addestramento e la valutazione dei modelli costituiscono i principali blocchi del pipeline proposto. Per la valutazione sono stati selezionati cinque algoritmi di Machine Learning e Deep Learning: Random Forest (RF), Support Vector Machines Classifier (SVC), XGBoost (XGB), Convolutional Neural Networks (CNN) e Long Short-Term Memory (LSTM). Per migliorare il rapporto degli algoritmi di classificazione in diversi contesti temporali, sono state valutate tre finestre preictali di 40, 60 e 80 minuti. Un punteggio F1 score superiore al 60% è stato raggiunto da 11/17 pazienti e con la finestra preictale preferita tra i soggetti è l'intervallo di 80 minuti.

Parole chiave: Epilessia, Caratterizzazione degli stati preictali, EEG, Teoria dei grafi, Connettività funzionale, Machine Learning, Deep Learning.

Contents

Abstract **i**

Abstract in lingua italiana **iii**

Contents **v**

Introduction **1**

1 State of the Art **5**
1.1 Brain Networks and Connectivity Measurements in Epilepsy 5
1.1.1 The brain as a network 5
1.1.2 Functional Connectivity Methods 8
1.1.3 Epilepsy as a network disease 9
1.2 Seizure Prediction Methods 11
1.2.1 Spiking Rate Behavior in EEG signal for seizure prediction 12
1.2.2 Machine Learning and Deep Learning strategies for seizure prediction 13

2 Materials and Methods **17**
2.1 Pipeline 17
2.2 EPILEPSIAE database 19
2.2.1 Dataset exploratory analysis 19
2.3 Feature Extraction 20
2.3.1 Dataset preprocessing 21
2.3.2 Functional Connectivity Analysis 22
2.3.3 Network Feature Extraction 24
2.4 Feature preprocessing 26
2.4.1 Feature Median Filtering 27
2.4.2 Outlier Removal 27
2.4.3 Missing Values Imputation 28

2.4.4 Data Standardization 28
2.4.5 Data Augmentation 28
2.5 Feature Selection 29
2.6 Machine Learning and Deep Learning Application 30
2.6.1 Model Tuning 30
2.7 Model Training and Evaluation 39
2.7.1 Preictal window selection and Label assignment 39
2.7.2 Dataset splitting 40
2.7.3 Model Selection 44

3 Results **45**
3.1 Feature Extraction 46
3.1.1 Functional Connectivity Analysis and Network Feature Extraction . 46
3.2 Feature Preprocessing 47
3.2.1 Data Augmentation 48
3.3 Feature Selection 49
3.4 Model Training and Evaluation 49
3.4.1 Group 1 50
3.4.2 Group 2 53

4 Discussion **59**

5 Conclusions and Future Developments **63**

Bibliography **67**

A Appendix A **77**

B Appendix B **79**

List of Figures **81**

List of Tables **83**

Acknowledgements **85**

Introduction

Epilepsy is a neurological disease caused by abnormal electrical disruptions in the brain activity [1]. It is considered a chronic condition as it can incur recurrent epileptic seizures, convulsive or non-convulsive that can cause loss of consciousness, unusual behavior and sensations, and uncontrollable movement [2]. According to WHO, epilepsy affects about 1% of people worldwide, where 70% are adults and the remaining 30% are children. The most common type of epilepsy is called idiopathic epilepsy and can affect 6 out of 10 people, its cause is unidentifiable [3].

The occurrence of an epileptic seizure is due to chemical changes in the neurons of the brain, which communicate through positive and negative charges, generating electrical signals. These spontaneous changes produce excessive high-amplitude electrical activity that is not easily controllable and could end up in a seizure. In clinical terms, if two or more unprovoked seizures occur, the proper cause might be epilepsy. If so, the determination of the epileptic seizure onset is very useful to start the treatment of this disease [2][4].

Based on the epileptic seizure onset, two types have been defined: focal onset seizure, where only a specific portion of the brain is damaged causing mild or severe symptoms according to the spread of the electrical discharge, and generalized onset seizures which affect the whole brain simultaneously [4]. To lessen the negative effects of seizures, it can be helpful to track electrical activity in the brain, recognize developing epileptic states, and anticipate potential seizures.

A well-known method used for monitoring brain activity and the detection of epilepsy is based on the analysis of electroencephalogram (EEG) signals. EEG is one of the most effective and low-cost techniques used to measure electrical activity in the brain. It can be invasive or non-invasive. The information of the brain is recorded by means of electrodes placed on the scalp or cerebral cortex that detect changes of potential in the neurons. The EEG signal can be decomposed into five frequency bands: delta (δ, < 4 Hz), theta (θ, 4-8 Hz), alpha (α, 8-12 Hz), beta (β, 12–30 Hz) and gamma (γ, > 30 Hz). Each of the frequency bands is associated with a specific waveform and neural activity [5]. For example, delta waves are associated with deep sleep states, theta waves with deep

meditation state, alpha waves with awareness and relaxation, beta waves are predominant in large attention waking states, whilst gamma waves are highly associated with decision-making processes [6, 7]. In order to understand abnormal brain behavior, identify, and detect possible epileptic seizures the study of brain rhythms has become fundamental.

Generally, for epilepsy patients, different seizure states can be distinguished among EEG signals of their brain activity, including interictal (normal brain activity between seizures), preictal (immediately before the seizure), ictal (seizure period), and postictal (immediately after the seizure) [4, 8]. Distinct patterns can be shown between the different epileptic states, allowing the analysis of the progression of a seizure, as well as its potential occurrence and the mitigation of any negative effects that might affect the patient. Special attention has been given to the preictal state since it could give important information about any early alteration in the behavior of the brain activity that may end up in an epileptic event [9, 10]. The study of the EEG dynamics has also demonstrated that the change in epileptic states, especially between interictal and preictal periods, typically happens minutes to hours before the seizure onset, which can herald the occurrence of a seizure to the patient [9, 11–13].

Since the 1970s, automated analysis of continuous EEG recordings started to be used as a tool of support for the diagnosis of epilepsy. One of the first techniques was proposed by J. Gotman *et.al*, 1976. It consisted of the automatic recognition of interictal activity by quantifying spikes and sharp waves in the human scalp EEG [14]. Thereafter, Gotman improved his approach by building a method to identify epileptic seizures independently of the presence of clinical signs. Instead, it was based on seizure patterns identified during the interictal and ictal period by means of the decomposition of the EEG signal into elementary waves and the measure of their relative amplitude [15].

Furthermore, to analyze the seizures, different methods have been proposed to compute features from EEG signals in different domains: morphological, time, frequency, in time-frequency and spatial [16–19]. Automatic seizure detection for epileptic patients has the potential to reduce the amount of data that needs to be examined, which lowers the expense of continuous monitoring. Moreover, automatic seizure anticipation could improve the quality of life of patients, by preventing the occurrence of seizures.

In the recent years, epilepsy has been studied as a network disease. Multiple investigations have looked at the brains' functional connectivity behavior in different epileptic seizure stages by means of large-scale brain networks, also called graphs [20–23]. A brain network is composed by a set of **nodes** and a set of edges that connect a node. Each node represents a wide brain region (channels of EEG signals), while each **edge** represents the

existing interaction between nodes (inter-dependencies of EEG signals). A node with a high degree of connectivity is called a **hub node**, and a specialized group of nodes highly interconnected is known as a **module** [21]. Rings *et al.* showed that some nodes connected by an edge that carried predictive information could be involved in the generation of seizure precursors. Time-varying and centrality changes were analyzed to assess the predictive role of each node [24]. Further investigation must be done to understand the behavior of multiple features among brain networks.

Machine Learning and Deep Learning techniques have revolutionized the seizure prediction field as it address the high complexity of EEG signals allowing a multivariate analysis and high-order feature spaces [25–27]. In general, these solutions are composed of four main steps: **(i)** Pre-processing of the EEG signal, **(ii)** Feature Extraction from the EEG signal, **(iii)** Feature Selection, and **(iv** Implementation of machine or deep learning classifiers. Due to their robustness for modeling high-dimensional data, Support Vector Machine classifiers have become one of the most preferred solutions for classifying binary-class epileptic seizures [28]. In the aforementioned studies, good performance was obtained, a sensitivity between 90-100% and accuracy greater than 90%.

The present study focused on developing an efficient and patient-specific AI-based pipeline to accurately identify interictal and preictal states, in 17 epileptic patients from the *EPILEPSIAE* database, by integrating functional connectivity analysis through epileptic brain networks. The analysis was carried out for three different preictal windows: 40, 60, and 80 minutes. Five Machine Learning algorithms were evaluated: Random Forest (RF), Support Vector Machines Classifier (SVC), XGBoost (XGB), Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM).

1 | State of the Art

Epilepsy is a neurological disorder affecting around 70 million people worldwide and is characterized by the occurrence of spontaneous recurrent seizures [3]. The unpredictability of seizures has a significant impact on the quality of life for epileptic patients, particularly for those who do not respond to pharmacological treatment. It is still unclear how the brain may go from a normal state to a seizure, a process known as ictogenesis. The study of the transition from the interictal state to an ictal state (seizure) is been fundamental to understanding the dynamics of epilepsy [29]. In order to improve disease detection and therapy, scientists and doctors are very interested also in the study of the preictal state. Thanks to its distinct electrophysiological behavior, it gives useful information that could help anticipate or even prevent a seizure. Brain networks and connectivity measurements have become one of the most studied techniques, along with new machine-learning models to anticipate possible seizures in a patient.

1.1. Brain Networks and Connectivity Measurements in Epilepsy

1.1.1. The brain as a network

The brain can be analyzed as a complex network of interconnected components that produce diverse cognitive behaviors [30]. The analysis of brain networks in the last few years has gained relevance in the research field. Boccaletti *et al.* reviewed in [21] the importance of the study of complex networks to understand brain dynamics and its impact in the last decades. Their irregular, complex, and dynamic structure evolves over time and has shifted the focus from small network analysis to those of systems with thousands or even millions of nodes [21].

In their most general form, graphs can be defined as a collection of elements (nodes) and their pairwise interconnection (edges). The latest represents the strength of the connection between two nodes (i, j) therefore it can take binary or weighted values W_{ij}.

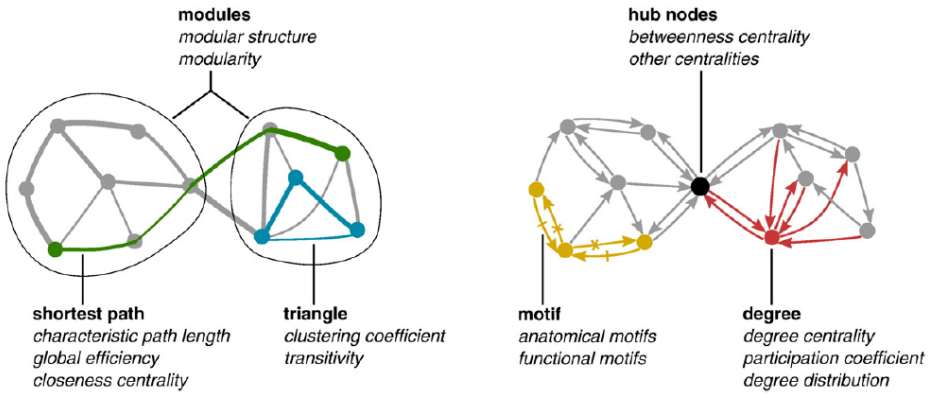


Figure 1.1: Measures of network topology. The left figure represents an undirected network. In the right figure, a directed network is shown. In green, are the shortest path lengths. Hub nodes, in black, lie on a high number of shortest paths. Taken from [31]

A node with a high degree of connectivity is called a **hub node**. A specialized group of nodes highly interconnected is known as a **module**. These elements can be represented as a connection matrix, allowing the study of pairwise relations between interacting brain regions [32]. Figure 1.1 shows different measures associated with network connectivity. Since graphs can be explored at various scales, graph measures can be calculated at a local (nodal) or a global (network-wide) level [32]. Some of the most used network features are **1)** Node strength (eq. 2.3), **2)** Clustering coefficient(eq. 2.4), **3)** Average shortest path (eq. 2.5), **4)** Betweenness centrality (eq. 2.6), and **5)** Eigenvector centrality (eq. 2.7).

The structural network of the brain can be represented through graph theory by means of nodes and edges. A *node* is a brain area defined by: **(i)** cellular architecture, **(ii)** local connectivity, **(iii)** output target projection and **(iv)** input projection source [33]. An *edge* is defined as an interconnected brain area.

With one of the earliest theories about brain networks, Sporns *et al.* [34] examined general concepts in the structural and functional organization of complex networks on a wide scale. This work analyzes a brain network as a set of nodes linked by connections that are mathematically presented as *graphs*. The complexity of the network is not only represented by its size but by existing interactions of the architecture and dynamics that give rise to complex emergent behaviors. Functional segregation (modularity) and functional integration are also discussed. Finally, they suggest that large-scale connection patterns emerge as a link between network distribution and human cognition.

Recent studies about large-scale brain networks have provided a better understanding of the neural basis of human cognition by showing how the interactions within distributed brain systems give rise to cognitive activities. Nodal measures based on node and edge

definition are used to characterize a network and obtain information about the brain's activity and function [32, 33, 35, 36].

Moreover, experimental modalities have been implemented using the analysis of brain connectivity models between regions. Neuroimaging techniques such as functional Magnetic Resonance (fMRI), Diffusion Tensor Imaging (DTI), Magnetoencephalography (MEG), and Electroencephalography (EEG) are reviewed as mapping methods of the human brain networks [37]. In particular, EEG is a very useful tool for mapping functional networks, because of its accessibility, low cost, and wide use in the context of diagnosis of epilepsy. Figure 1.2 shows the pipeline suggested to explore structural and functional brain networks using graph theory and neuroimaging techniques [37].

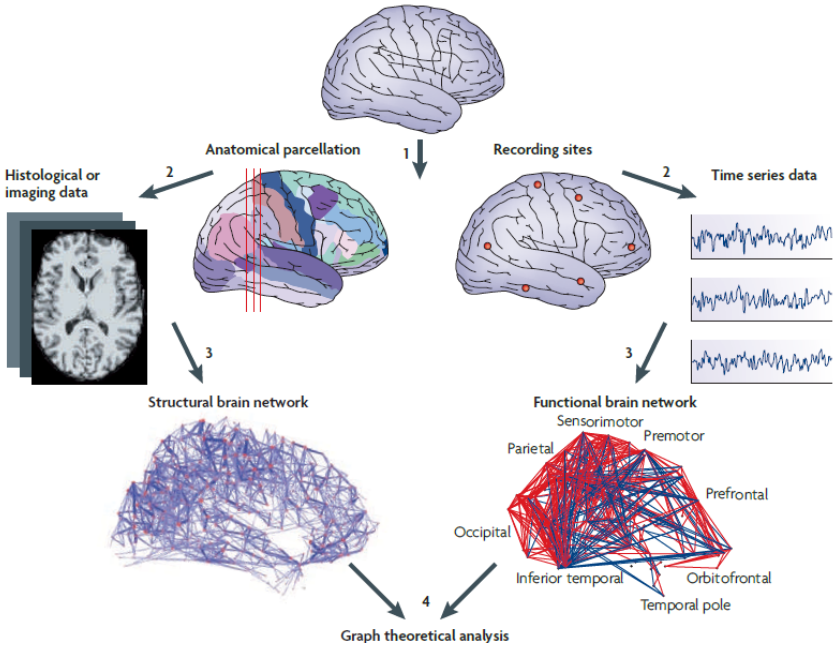


Figure 1.2: Structural brain network pipeline. Taken from [37]

Furthermore, van den Heuvel in [38] demonstrated that brain hubs are able to form so-called "*rich club*". A "*rich club*" is characterized by a bias towards *high-degree* nodes, also known as hubs, connected among themselves in a more dense way than *low-degree* nodes, i.e., nodes with low number of connections [38]. Centrality of the *rich club* was computed to a specific set of nodes/edges to examine its role in the global network structure. Highly connected and central brain hubs exhibited a robust tendency to be mutually interconnected.

1.1.2. Functional Connectivity Methods

Functional connectivity (FC) analyzes the correlation between distinct brain nodes or regions. It follows the premise that regions of the brain that are functionally related or involved in similar tasks are highly correlated, If so, they exhibit synchronized activity patterns. Whereas Structural Connectivity refers to anatomical connections between different nodes or regions in the brain.

Different brain functions require the physical connection of associated anatomical regions, as well as interactions by means of the synchronization of brain oscillations [39]. The first one, is known as structural connectivity, and the latter is known as functional connectivity. Multiple methods allow to measure the functional connectivity as relationships between the channels of EEG. In general, it describes statistical patterns associated to dynamic interactions among brain signals, also known as "*functional networks*" [35].

As presented in multiple studies [35, 40–42], linear and nonlinear measures such as cross-correlation, mutual information, phase synchrony or spectral coherence, of the amplitude among channels have been implemented to identify patterns of synchronous activity across different brain areas. In particular, Pearson's correlation is a robust method to quantify the degree of statistical interdependence's between EEG channels [43]. Functional connectivity analysis allows the assessment of functional interactions among brain network nodes.

Bressler *et al.* refers to functional connectivity as a tool for functional interdependence analysis, which aims to identify network edges as they evolve in time or frequency [33]. For time analysis, cross-correlation function is suggested, while spectral coherence or phase synchrony should be used in the frequency domain.

Finally, the brain is a connected system where multiple nodes can represent a region or a connection, graph analysis of functional brain networks is highly implemented to visualize the network's behavior [44]. This tool has become widely used in translational neuroscience to quantify brain dysfunction in terms of abnormal brain network reconfiguration.

Figure 1.3 shows the pipeline proposed to perform graph analysis for functional brain network modeling. Each node corresponds to a specific site in the brain, recorded by an electrode, while the weight of each link is estimated by measuring the functional connectivity of the activity between each of the nodes. Then, filtering is performed to choose the most important links that will constitute the brain graph. Finally, different metrics are computed to quantify the behavior of the brain graph and a statistical analysis

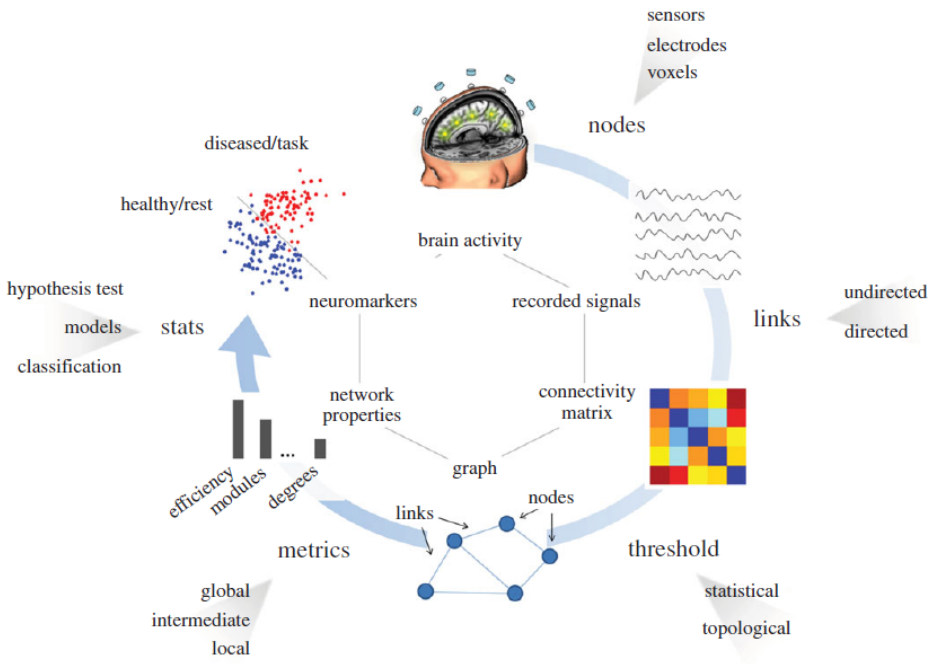


Figure 1.3: Functional Brain Networks modeling and analysis pipeline. Taken from [44]
.

leads to a possible classification of healthy and diseased patients.

1.1.3. Epilepsy as a network disease

Epilepsy is increasingly recognized as a network disease in which abnormal patterns of neural activity and connectivity play a relevant role. Contemporary research has exposed the complex nature of epileptic networks since seizure activity can propagate variably to any or all brain areas where anatomic connections exist and that are not necessarily part of the active network [45–48]. These investigations have demonstrated that epilepsy involves frequent modifications in large-scale brain networks that lead to abnormal synchronization and propagation of electrical activity.

In order to understand the evolution in time of the epilepsy network the study of the *"seizure onset"* has increased. Spencer in [45] suggested Intracranial EEG as the first important approach for the observation of *"seizure onset"* within the network. Advanced functional neuroimaging techniques such as PET and SPECT have been also studied. While PET scans are not sensitive to the epileptogenic zone early propagation areas, SPECT scans are more sensitive and have the unique ability to capture the status of blood flow when seizure activity starts providing a more precise epileptic network information [45]. Likewise, Centeo *et al.* point out Resting-State fMRI (RS-fMRI) as a tool to examine synchronized activity between regions without a specific task based on signal correlation.

An essential component of the neural network role is FC, which depicts the statistical correlation between activities in various brain areas. Today's cutting-edge research in epilepsy follows epileptic network analysis by means of FC. For example, van Diessen *et al.* suggested that disruptions in connectivity of large-scale brain networks for epileptic patients could be related to cognitive and behavioral impairments [48]. The reviewed framework in this study was the resting-state condition since it allowed the detection of the intrinsic activity of the brain. Nevertheless, multiple studies stand that there is no general consensus yet of whether functional connectivity increases, decreases, or remains constant for epileptic patients [48–50].

Emerging studies have found an increased average clustering coefficient for epileptic patients, which measures the tendency of nodes inside a network to cluster together [51–53]. This represents a more segregated network. Furthermore, a study in epileptic drug-resistant patients, showed a decrease in functional connectivity over the zones damaged by seizure propagation, as well as in the unaffected seizure side [54]. Morgan *et al.* demonstrated that disruption of resting-state cross-hippocampal functional connectivity increases over the years [55].

In general, FC analysis has shown an important definition of the network dynamic behavior during epileptic seizures. In order to gain insight into the temporal evolution of the disease, seizure dynamics have been studied along four main epileptic states[48].

1. **Interictal:** represents the non-seizure period.
2. **Preictal:** is the period immediately preceding the seizure onset.
3. **ictal:** corresponds to the actual occurrence of an epileptic seizure.
4. **Postictal:**it occurs after the seizure ends and the patient returns to baseline (normal brain activity).

To understand *"seizure dynamic"* variations driven by patient-specific cyclic patterns, it is necessary to monitor FC in epileptic networks for long periods of time. Nonetheless, given the necessity to explore differences in FC between regions for each patient, further investigation must be carried out. Examining the large-scale network in epileptic patients can help shed light on the causes of epilepsy and possibly on biomarker identification that could be used to predict epileptic seizures and guide a possible treatment.

1.2. Seizure Prediction Methods

Almost 30% of worldwide epilepsy patients cannot effectively treat the disease with anti-epileptic medicine due to drug-resistance [31]. Drug-resistant epilepsy (DRE) reduces the patient's quality of life increasing the risk of injuries, psychosocial dysfunction, and premature death [56]. The unpredictable nature of epileptic seizures is one of the most dangerous features of this disease. For clinicians, this has represented a major challenge for epilepsy treatment. As a result, seizure prediction has been particularly studied over the last decades. The ability to forecast a possible seizure for a patient with epilepsy could mean better precaution against injury as well as possible control over the incoming seizure.

Intracranial EEG (iEEG) has been one of the most preferred methods to monitor electrical activity in epileptic patients, here the electrodes are directly placed into the brain tissue which guarantees a high-quality signal recording. The work of Mormann *et al.* in 2007 reviewed multiple methods for seizure prediction that could lead to new epilepsy treatments towards seizure control *EEG-triggered on-demand therapy*. For example, by giving fast-acting anticonvulsant substances or by electrical stimulation in order to reset the dynamics of the patient's brain. First, it mentions iEEG as the most used technique at the time to collect neural behavior during epileptic seizures. This method allows the definition of a transitional preictal phase, corresponding a specific changes in brain activity. Seizure prediction aims to identify a preictal state defining a window of variable size before the seizure onset arises. The moving window analysis, while the iEEG signal acquisition is carried out, is one of the most used methods as it allows time profiling for one or multiple channels. The windows presented a typical range between 10 and 40 seconds. Finally, it recalls the importance of statistical validation on predictability over the designed algorithm to evaluate its performance [57].

After a decade, Kuhlmann *et al.*, presented an updated study about new seizure prediction strategies. The need for robust predictors led to the development of new databases that enabled researchers to have access to *long-term continuous* data to validate potential predictive algorithms [31]. *EPILEPSIAE* database provided access to multichannel *long-time* recording EEG data. Later studies have shown a high focus on the analysis of non-continuous seizure-free intervals (interictal) and pre-seizure periods (preictal) data. Finally, prospective seizure prediction systems are proposed, as shown in figure 1.4. The system is generally composed of intracranial implanted electrodes that transmit the EEG data wirelessly. The signal is then recorded and preprocessed. Finally, an algorithm extracts preictal biomarkers that allow seizure prediction.

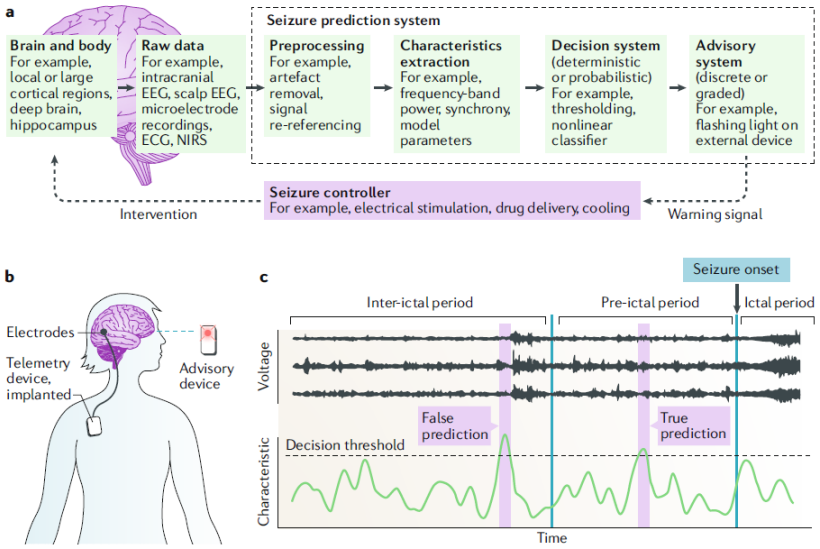


Figure 1.4: Seizure prediction system. **a)** Typical seizure prediction system scheme. **b)** Open-loop seizure prediction warning system **c)** Seizure prediction in practice. Taken from [31]

1.2.1. Spiking Rate Behavior in EEG signal for seizure prediction

As seizures are characterized by generalized spike-wave discharges [58], one of the first methods proposed for seizure prediction using EEG was based on electrographic changes represented as a burst of spikes along the transitional preictal period. In [59], Le Van Quyen *et al.* examined scalp-EEG recordings in 23 patients with temporal lobe epilepsy. preictal changes were evaluated 60 minutes before seizure onset. By means of non-linear similarity measures, long-term changes before a seizure were identified, which led to its recognition around seven minutes before seizure onset. A following study showed that a particular synchronization is observed between 5 and 24 hours prior to the actual seizure and that changes are frequently concentrated near the main epileptogenic zone [60].

Another study analyzed iEEG recordings from 21 patients with focal epilepsy [58]. Signal information was divided into interictal, preictal, ictal, and post-ictal. By means of a morphological filter, EEG spikes were identified. The study suggested that the spiking rate increases upon an incoming seizure and reaches its maximum value when the seizure arises. A sensitivity of 75.8% was achieved for seizure prediction with an average false prediction rate (FPR) of 0.09 FPR/h. A mean prediction time of 49.7 minutes was achieved. The computational cost of spike rate analysis for seizure prediction was very low.

1.2.2. Machine Learning and Deep Learning strategies for seizure prediction

In [31], Kuhlmann *et al.* states that the temporal relationship of spikes within seizures tends to add some uncertainty when predicting a seizure. Hence, new prediction strategies based on machine learning (ML) and deep learning (DL) algorithms with a high degree of accuracy have been proposed over the last few years to overcome this problem [26, 27, 61, 62]. Generally speaking, an automated seizure prediction system follows: **i)** EEG signal pre-processing, **ii)** Feature extraction **iii)** Dimensionality Reduction [63], **iv)** Classification process, i.e., the step where prediction is performed, and, finally, **v)** Decision-making process for predicting the epileptic seizure [62].

K-nearest-neighbor(KNN), Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) classifiers have been the most preferred algorithms to predict seizures in epileptic patients. In [64], an online seizure prediction system based on continuous monitoring of 10-minute sliding windows of EEG recordings was presented. There was an overlap of 50% between each consecutive window. The seizure prediction sought to analyze interictal and preictal baselines. Therefore, a gradient-based reinforcement learning algorithm was applied. The prediction range oscillated between 30, 90, and 150 minutes before seizure onset. With the last one, a sensitivity of 73% and a specificity of 67% was obtained.

Another study applied the RF algorithm for automated seizure detection using improved correlation-based feature selection (ICFS) [65]. In this work, an EEG signal was analyzed along the five physiological bands: delta, theta, alpha, beta, and gamma. Time, frequency, and entropy-based features were extracted from the raw EEG and the wavelet decomposed signal. Finally, feature selection was implemented by ICFS method. For the 8 cases presented in the study, specificity and sensitivity greater than 97% were achieved.

In 2019 [66], Wang *et al.* proposed an automatic seizure detection pipeline based on multiple time-frequency analysis using short-time Fourier transform to extract relevant signal features. An RF algorithm based on grid search optimization (GSO) was used obtain a three-class seizure classification: seizure, light-seizure, and non-seizure. Principal Component Analysis (PCA) was used for dimensionality reduction purposes. Multiple cross-validations are suggested to boost the model's accuracy. The accuracy obtained with the GSO strategy went from 88% to 96.7%

Another ensemble method evaluated for seizure prediction field in the literature is extreme Gradient Boosting (XGBoost). In [67], the XGBoost algorithm was implemented to per-

form a binary classification (seizure or background) for an EEG signal recording segment. To smooth the data, a sliding window of 19 segments applied a simple majority vote to avoid isolated and improbable seizure/non-seizure states. When the train, validation, and test set were evaluated, a sensitivity of 78.72% and a specificity of 33.54% were achieved.

SVM classifier (SVC) has been one of the most widely used ML algorithms to predict potential epileptic seizures. Due to the existing patient-specific seizure nature and the underlying dynamics in long-term EEG signals the development of patient-specific modeling with ML and DL algorithms is recommended. This is shown in [28], where iEEG recordings of dogs with epilepsy (which share biological similarity with human seizures) were used to develop SVM-based seizure forecasting subject-specific models. Interictal, preictal, and ictal states were discriminated by the algorithm. A 60-minute preictal window recording segment was used for the prediction. The average sensitivity was about 95%. The study suggested that a good prediction performance could be possible if a minimum amount of seizures (at least 5-7 seizures) are contained in the training data.

In [68], Gupta *et al.* implemented a three-stage method to process the EEG signal: *1)* A filterbank to decompose the EEG signal into the main five brain rhythms, *2)* Statistical modeling of the brain rhythms by means of a self-similar Gaussian random process using the Hurst exponent, *3)* an SVC was implemented to classify preictal, interictal and ictal EEG signal segments. No preictal analysis window was specified. The implemented model showed a sensitivity and specificity of 97%.

New DL methods have been implemented in recent studies due to their capability of automatically extracting detailed features from complex signals such as EEG, identifying useful patterns that a traditional ML technique might not be able to extract, and that could lead to better seizure prediction [27, 69, 70]. Convolutional Neural Networks (CNN) were applied in [69] as an automatic feature extraction technique to distinguish between interictal, preictal, and ictal periods along a the EEG signal. The study identified that the preictal phase transition towards a seizure occurred approximately 10 minutes before seizure onset and that it can be observed in the scalp EEG using automatically extracted features from the CNN. A sensitivity of 87.8% and a specificity of 0.142 FP/h were obtained which showed promising results.

Finally, Bongiorni *et al.* [27] presented Recurrent Artificial Neural Networks (RNN), of Long Short Term Memory (LSTM) type, as an approach for seizure prediction on segments of pediatric epilepsy patients. Short-Time Fourier Transform (STFT) was implemented to obtain spectral features that showed significant differences among the analyzed EEG periods. The results obtained suggested that there was a strong dependence between the

LSTM proposed technique and the specific patient EEG signal patterns. Therefore, no generalization about EEG patterns for several patients was possible. The RNN achieved a sensitivity of 61% and a specificity of 99% for ictal state classification, a window between 5 and 60 seconds before an epileptic seizure was defined.

2 | Materials and Methods

2.1. Pipeline

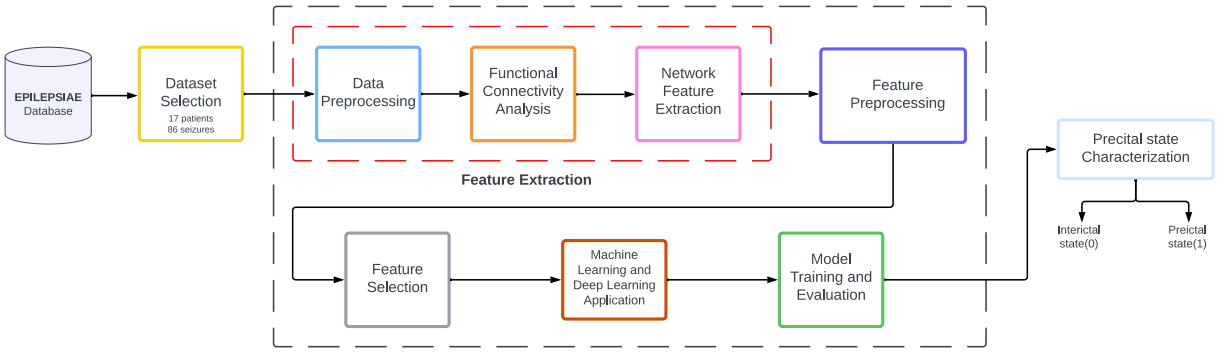


Figure 2.1: Preictal state characterization pipeline.

The designed preictal state characterization pipeline was composed by 6 main blocks: Epilepsiae dataset selection, Feature Extraction, Feature Preprocessing, Feature Selection, Machine Learning and Deep Learning Application, and Model training and Evaluation, as shown in figure 2.1.

1. **Epilepsiae dataset Selection**: 17 patients with sEEG recordings available for at least 8-9 hours before the seizure onset were selected randomly from the EPILEPSIAE database, for a total of 86 seizures. The patient selection was carried out thanks to exploratory analysis, as explained in section 2.2.1.

2. **Feature Extraction**: This block is divided in three parts, as follows.

   (a) **Dataset preprocessing**: The EEG signal was analyzed along seven frequency bands: $\delta, \theta, \alpha, \beta L, \beta H, \gamma L, \gamma H$. To reduce the amount of the data the P_{rms} of the 1-hour EEG signal recording was performed following equation 2.1. Additionally, signal concatenation of each sEEG recording block per patient was carried out to obtain an 8-9 hour recording per seizure instance. Finally, the identification of the activation time for the Seizure Onset Zone (SOZ) was carried out.

(b) **Functional Connectivity Analysis**: The connectivity matrix was computed calculating the Person's Correlation (see 2.2) between each pair of nodes for 1-minute windows. A total of 480-minute windows were obtained per seizure per patient.

(c) **Netwrok Feature Extraction**: Based on the connectivity matrix, 5 brain connectivity characteristics (*node strength, clustering coefficient, average shortest path, betweenness centrality, and eigenvector centrality*). The last characteristic was obtained using the signal's P_{rms} information. For most of the characteristics, the mean and standard deviation were computed, obtaining a total of 11 features computed for each of the 7 bands. Table 2.3 shows the description for each feature. In this step, the final feature dataset was obtained for each of the patients.

3. **Feature Preprocessing**: The splitting of the dataset was performed according to the number of seizure instances available per patient. Then, Feature Median Filtering was applied over the whole 480-minute feature window. After, 4 preprocessing steps were fitted to the train set and then applied to the validation and test sets: Feature Median Filtering, Outlier Removal, Missing Values Imputation, Data Standardization, and Data Augmentation (only for the Train set).

4. **Feature Selection**: This step was carried out in order to reduce the model's complexity and therefore increase its generalization capabilities. Low variance filtering, Mann-Whitney U test, and High correlation filtering were carried out. The feature selection was carried out over the train set, and the same features were then selected on the validation and test set to guarantee coherence.

5. **Machine Learning and Deep Learning Application**: 5 models (Random Forest, Support Vector Machine Classifier, XGBoost, Convolution Neural Network, and Long Short-Term Memory RNN) were trained to perform interictal and preictal period classification. The model tuning was performed using the Grid Search (GS) method.

6. **Model training and Evaluation**: This block is in charge of the label assignment process, group analysis division, and metric evaluation performed for each model in order to assess its performance. Metrics such as F1-score, Precision, and Recall were computed. In addition, Confusion Matrices were assessed for each model, along with the ROC and PR curves. The best-performing model and optimal window selected were chosen according to the F1 score value in the validation set. Section 2.7 presents the evaluated metrics.

The pipeline for the present Machine Learning application for the characterization of preictal states is presented in more detail in Appendix A.1.

2.2. EPILEPSIAE database

EPILEPSIAE was the selected database for this project. The database was part of an EU-funded project that focused on the creation of algorithms for seizure prediction based on high-quality (scalp and intracranial) long-term (>96h) EEG recordings of almost 300 epileptic patients [71]. Metadata information about start and end timestamps of the recordings, number of channels, number of samples, and others, for each patient was contained within the database. For the current investigation, 17 patients with stereo-EEG (sEEG) recordings, which is a type of iEEG where electrodes are placed in more depth inside the brain, were selected, for a total of 86 seizures. The selected period for analysis was between 8-9 hours before the activation of the seizure onset zone (SOZ).

2.2.1. Dataset exploratory analysis

Originally, 18 epilepsy patients with sEEG recordings lasting between 8-9h were randomly selected from the *EPILEPSIAE* database. An exploratory analysis was carried out in order to understand the number of seizures available per patient, as well as the number of electrodes and the number of sessions available for the selected subjects. A session refers to a full block of 8-9 hours of sEEG recording. Figure 2.2 presents the histogram of the number of seizures available per patient within the database.

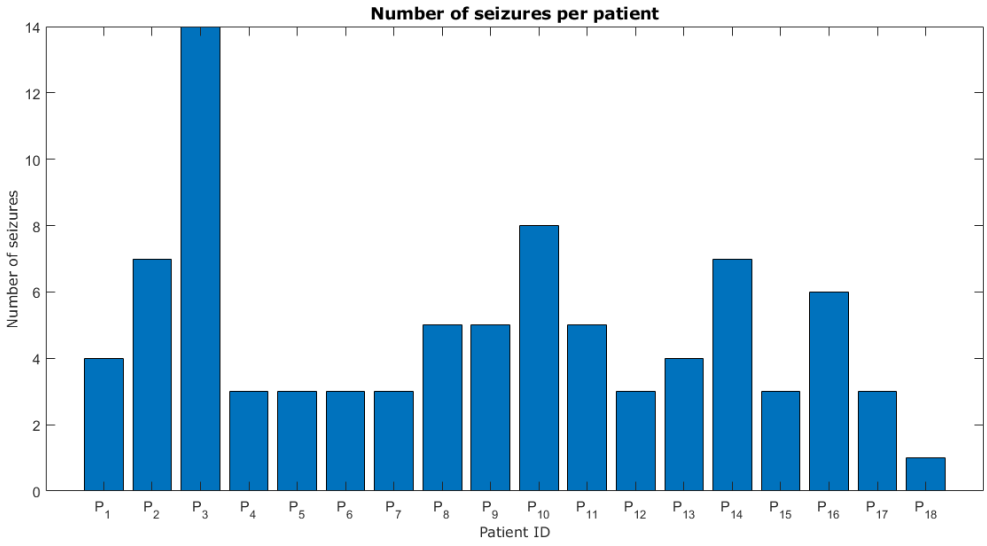


Figure 2.2: Number of seizures available per patient

P_{18} was pruned, as it only had one available seizure, which would not allow it to perform an accurate model training nor validation. Therefore, a total of 17 patients were selected to train patient-specific ML and DL classifiers to discriminate between interictal and preictal states. Table 2.1 shows the number of electrodes, number of seizures, and number of recording sessions (recording blocks) available for each of the selected patients.

Patient ID	Number of seizures	Number of electrodes	Number of sessions per patient
P_1	4	84-105	2
P_2	7	71-92	4
P_3	14	84-105	2
P_4	3	98-119	2
P_5	3	96	1
P_6	3	94-118	2
P_7	3	84-106	2
P_8	5	121	1
P_9	5	80	1
P_{10}	8	117	2
P_{11}	5	96-117	2
P_{12}	3	115	1
P_{13}	4	61	1
P_{14}	7	92-115	2
P_{15}	3	92	1
P_{16}	6	46-69	3
P_{17}	3	124	1

Table 2.1: Seizure patient information and Patient ID assigned for the present work

2.3. Feature Extraction

2.3.1. Dataset preprocessing

EEG Frequency Band	Low Cutoff Frequency (Hz)	High Cutoff Frequency (Hz)
Delta (δ)	0.3	4
Theta (θ)	4	8
Alpha (α)	8	12
BetaL (βL)	12	18
BetaH (βH)	18	30
GammaL (γL)	30	49
GammaH (γH)	51	99

Table 2.2: Frequency bands and cutoff frequency range selected to characterize the EEG signal

In order to remove artifacts that could affect the current analysis, a 50Hz filter was applied to eliminate line noise. Then, following the cutoff frequencies shown in table 2.2, physiological brain rhythms were filtered. Finally, to reduce the number of samples of the *sEEG* recordings and get an overall behavior of the signal, the RMS power was computed over 3s consecutive windows.

This was performed following equation 2.1, where $b-a$ represents the 3s sliding window, starting on a up to b, $Y[n]$ in the measure of the signal in time $n \in [a, b]$ [72].

$$P_{RMS} = \frac{1}{b-a} \sum_{n=a}^{b} |Y[n]|^2 \tag{2.1}$$

Each patient recording session consisted of multiple 1-hour EEG recording blocks one of which contained the seizure event. A concatenation of the files was performed in order to obtain the 8-9 hours EEG signal before seizure onset. Figure 2.3 shows an example of the EEG signal block concatenation per brain rhythm for patient P_3. For each session, the identification of the Seizure Onset Zone (SOZ) activation time, i.e., the time when the seizure originates, was performed.

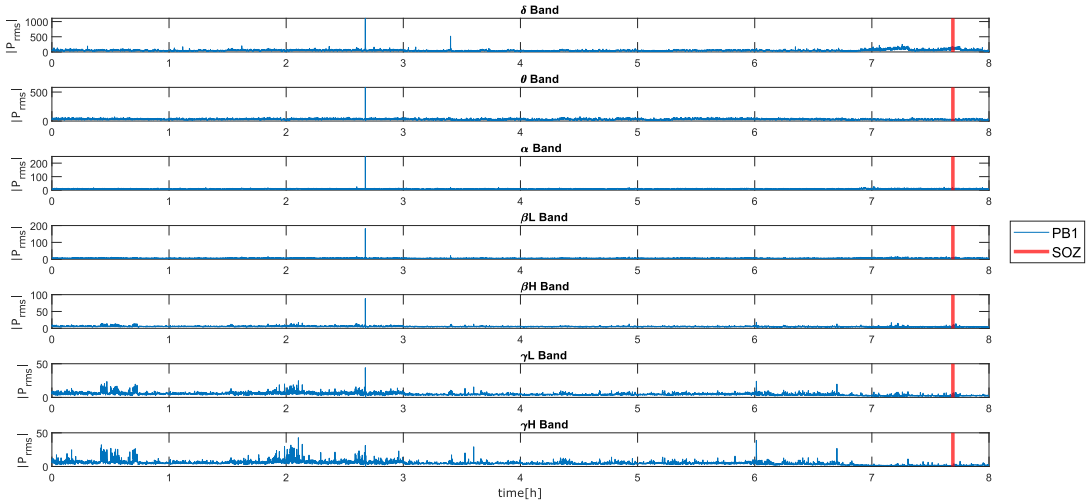


Figure 2.3: PB1 Electrode behavior for P_3 after band filtering and recording concatenation. In red, the SOZ marker.

2.3.2. Functional Connectivity Analysis

As mentioned in section 1.1.2, one of the main tools to analyze the behavior of brain networks in epileptic patients is based on FC. Specifically, Pearson's correlation is a robust and reliable approach for measuring the level of statistical dependence among EEG channels. Pearson's correlation coefficient (r_{ij}) measures the linear association within two variables X_i and X_j being $i \neq j$ [73].

Generally, correlation analysis begins with scatter diagrams, which depict the relationship between data pairs. Pearson's correlation coefficient values range from -1 to +1. When two variables have a positive correlation value it means a tendency for one to increase or decrease in tandem with another. A negative correlation coefficient suggests, on the other hand, a tendency for one variable's values to increase when those of the other variable decrease linearly, and vice versa. A low correlation between X_i and X_j is shown by a r_{ij} close to zero, whereas a strong linear correlation between the two variables is indicated by values close to -1 or +1 [43, 73]. The computation of the Pearson's correlation was carried out using the equation 2.2

$$\|r_{ij}\| = \left\| \frac{Cov(X_i, X_j))}{\sqrt{var(X_i)}\sqrt{var(X_j)}} \right\| \tag{2.2}$$

Taking the previous information into account, algorithm 2.1 explains the executed steps

to build the analyzed networks using Functional Connectivity Analysis.

Algorithm 2.1 Functional Connectivity Analysis

1: Assign each EEG channel as a network node
2: Divide the EEG signal in 1-minute windows with no overlap. Note: as each point of the signal represents a P_{RMS} value of 3s, each window must contain 20 samples.
3: Pearson's correlation coefficient is computed using equation 2.2 to calculate the functional connectivity between each pair of nodes. This is performed for each of the brain rhythms presented in Table 2.2.

As a result, for every 8 hours of EEG recording previous to seizure onset at least 480 windows, of 1 minute each, were obtained ($8h * 60min$). Therefore, 480 brain networks (one network per window) were built for the current analysis.

For each network, a matrix M_t of size $N * N$ was obtained. N represents the number of channels used on each sEEG recording. The M_t matrix is symmetric since the correlation between channels i and j is equal to the correlation between j and i. The main diagonal of M_t was nulled.

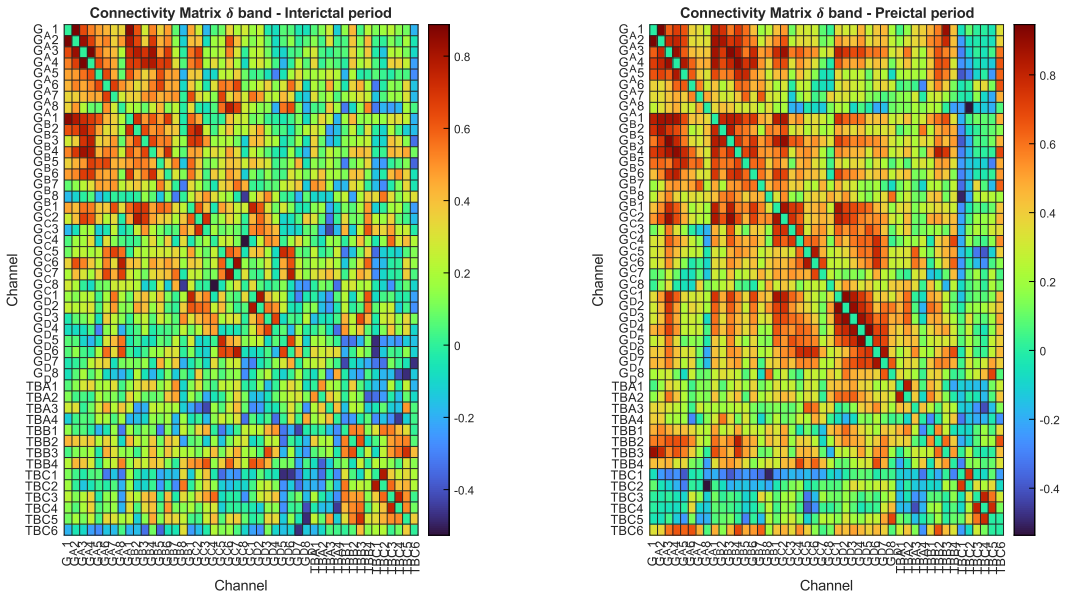


Figure 2.4: Connectivity Matrix for δ band for an interictal (*minute 50*) and preictal (*minute 450*) window for patient P_{16}

Figure 2.4 shows the connectivity matrix obtained for the last-minute window of a seizure of patient 16. P_{16} has a total of 6 seizures among the EEG recordings. It can be noted that

about 46 channels (electrodes) were set for the present recording. In addition, particular brain networks and areas that are involved in the propagation of epileptic activity could be analyzed from the connectivity matrix.

2.3.3. Network Feature Extraction

Based on the connectivity matrix computed for every 1-minute window, multiple network features were computed to build a unique patient feature matrix. Six network measures were computed per connectivity matrix for every single frequency band (see Table 2.2). Node strength, clustering coefficient, average shortest path, betweenness centrality, and eigenvector centrality were computed by means of the *MATLAB's "Brain Connectivity Toolbox"* (BCT). The following notation is explained to allow equations' interpretation:

- ***N***: is the set of all nodes in the network.
- ***n***: is the number of nodes.
- ***L***: is the set of all links in the network.
- ***l***: is the number of links. $l = \sum_{i,j \in N} a_{ij}$
- ***(i,j)***: is a link between nodes i and j, $(i, i \in N)$.
- a_{ij}: if the connection status between i and j:
  - $a_{ij} = 1$ when link (i,j) exists (when i and j are neighbors)
  - $a_{ij} = 0$ otherwise ($a_{ii} = 0$ for all i).
- k_i: Degree of the node, i.e., number of links connected to a node: $k = \sum_{j \in N} a_{ij}$

The following list shows the description and equation for each calculated feature using the BCT toolbox.

1. **Node Strength**: **Nodal** measure defined as the sum of the weights w_{ij} of all the edges that are connected to a specific node, i and j [36]. It is computed as follows:

$$S_i = \sum_{j \in n} w_{ij} \tag{2.3}$$

2. **Clustering Coefficient**: **Nodal** measure that quantifies the fraction of connections between a node and its neighbors, with relation to all possible existing links. It is also referred to as the tendency of nodes to form a triplet [36, 74]. It is computed

as follows:

$$C = \frac{1}{n}\sum_{i \in N} C_i = \frac{1}{n}\sum_{i \in N} \frac{2t_i}{k_i(k_i - 1)} \tag{2.4}$$

where C_i is the clustering coefficient of node i ($C_i = 0 for k_i < 2$).

3. **Average Shortest Path**: **Global** measure, it represents the average number of steps along the shortest paths for all possible pairs of nodes among the network [36]. The shortest path dis_{ij} is calculated using the Dijkstra's algorithm [75]. The average shortest path also characterizes the efficiency of information transport on a network, and is computed as follows:

$$L = \frac{1}{n}\sum_{i \in N} \frac{\sum_{j \in N, j \neq i} dis_{ij}}{n - 1} \tag{2.5}$$

4. **Betweenness Centrality**: **Nodal** measure described as the fraction of all shortest paths along the network that pass through a specific node. If a node has a high betweenness centrality value it means that the current node participates in a large number of shortest paths [36, 76]. The metric is calculated as follows:

$$b(i) = \sum_{j \neq k} \frac{n_{kj}(i)}{n_{kj}} \tag{2.6}$$

where n_{kj} is the number of shortest paths between h and j, and $n_{kj}(i)$ is the number of shortest paths between h and j that pass through i.

5. **Eigenvector Centrality Nodal**, a self-referential measure of centrality, nodes are weighted based on their degree of connections within the network. This feature is calculated by counting the number and the quality of connections. A node will have a high value of eigenvector centrality if it's connected to other nodes with high eigenvector centrality [36, 77].

$$x_i = \frac{1}{\lambda}\sum_{j=1}^{n} A_{ij}x_j \text{ , } where\ \lambda x = Ax \tag{2.7}$$

The remaining computed features pertained to the signal's power content (P_{RMS}). For each estimate, the mean (μ) and standard deviation (σ) were calculated. In total, 77 features ($11\ features * 7\ bands$) were obtained for every 1-minute window. Table 2.3 shows the final features extracted by each band.

Measurement	Measure Type	Function	Feature	Name
Node Strength	Nodal	$sum(M_t, 2)$	μ	<S>
			σ	σ_S
Clustering Coefficient	Nodal	clustering$_coef_wu(M_t)$	μ	<C>
			σ	σ_C
Average Shortest Path	Global	charpath(distance$_wei(M_t)$)	Average Shortest Path	sp
Betweenness Centrality	Nodal	betweenness$_wei(M_t)$	μ	<B>
			σ	σ_B
Eigenvector Centrality	Nodal	eigenvector$_$centrality$_und(M_t)$	μ	<ec>
			σ	σ_{ec}
P_{rms}	Signal amplitude	mean(),std()	μ	μ_{rms}
			σ	σ_{rms}

Table 2.3: Features extracted for each band

Finally, a vector X_i with 11 features per band (7 bands) was obtained, for a total of 77 features per 1-minute window, as follows.

$$X_i = [S_\alpha, \sigma_{S_\alpha}, ..., ec_\alpha, \sigma_{ec_\alpha}, ..., S_{\gamma H}, \sigma_{S_{\gamma H}}, ..., ec_{\gamma H}, \sigma_{ec_{\gamma H}}]_{1\times 77} \tag{2.8}$$

A patient-specific final dataset containing all the information previously mentioned was obtained per seizure instance, as shown in figure 2.5.

Connectivity Features Matrix		
PK	id	varchar
	new_id	varchar
	sze_num	integer
	minute	integer
	mean_Node_Strength_<BAND>	decimal
	mean_Node_Strength_<BAND>	decimal
	std_Node_Strength_<BAND>	decimal
	mean_Clust_coeff_<BAND>	decimal
	sttd_Clust_coeff_<BAND>	decimal
	Avg_Shortest_Path_<BAND>	decimal
	mean_Betweenness_Cen_<BAND>	decimal
	std_Betweenness_Cen_<BAND>	decimal
	mean_Eigenvector_Centr_<BAND>	decimal
	std_Eigenvector_Centr_<BAND>	decimal
	mean_P_rms_Delta	decimal
	std_P_rms_<BAND>	decimal
	sze_stamp	integer
	label	integer

Figure 2.5: Local Patient Dataset Columns

2.4. Feature preprocessing

The feature dataset preprocessing pipeline consisted of applying outlier removal, missing values imputation, and data standardization methods to the selected train set, in that same order. Finally, after handling the feature preprocessing for the training set, the same

outlier removal threshold, missing data imputer (KNN imputer), and standardization scaler were applied to the validation and test sets, just as previously mentioned.

To carry out the feature preprocessing in a robust manner, dataset-splitting was performed according to the patients' number of seizure instances available, as follows: For patients with at least 3 seizure recordings: a ***train set*** containing 70% of the seizures and a ***validation set*** containing the remaining 30%. For patients with more than 3 seizure recordings: a ***train set*** containing 70% of the seizures, a ***validation set*** with 20%, and, the ***test set*** with the remaining 10%. After performing the data splitting, the following preprocessing steps were fitted to the train set and then applied to the validation and test sets.

2.4.1. Feature Median Filtering

To reduce possible noise and smooth the data point inconsistencies, among each patient dataset, a median filter was applied with $k = 15$. The kernel $\boldsymbol{k}$ represents the size of the median filter window, in this case, the window took 15 samples.

2.4.2. Outlier Removal

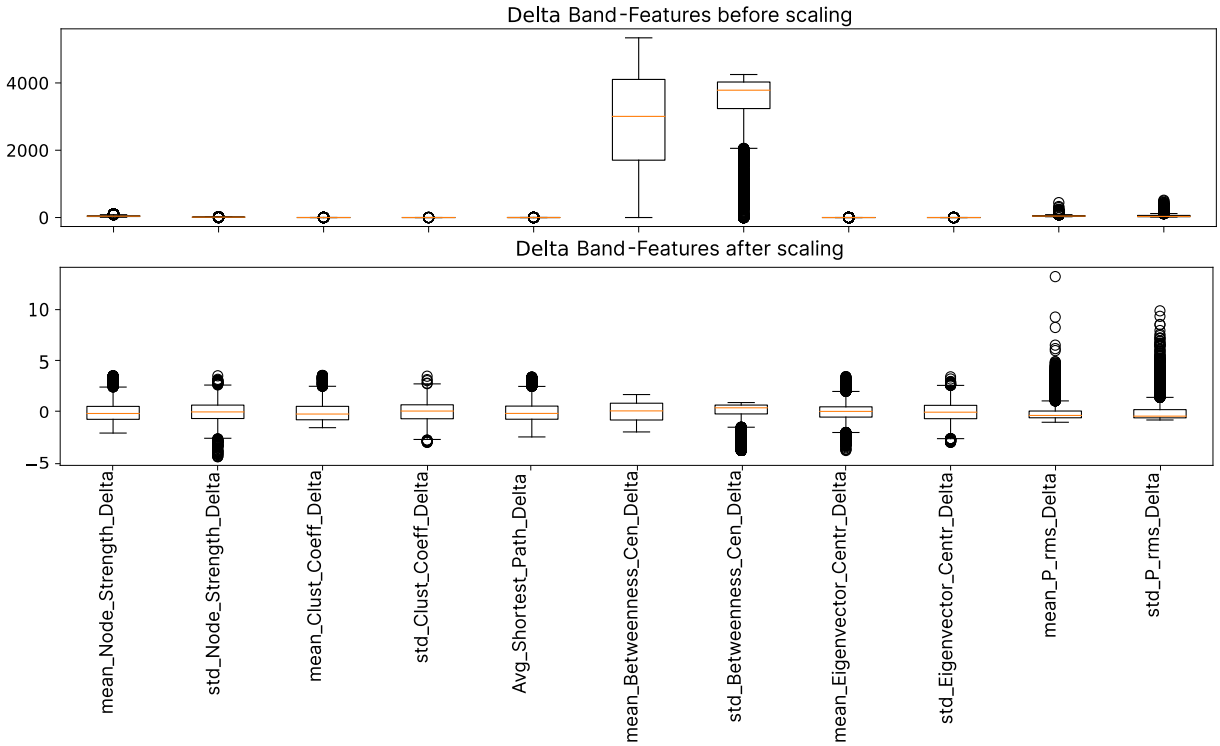


Figure 2.6: Boxplot of the feature behavior before and after outlier removal for δ Band

Identification of outliers serves the purpose of eliminating noisy data points that significantly deviate from their neighborhood. When detected, outliers may lead to instability in the data. Furthermore, outlier detection could drive pattern identification. For this purpose, the Z-score test was implemented. Figure 2.6 shows the behavior of the features in δ Band before and after outlier removal for P_3.

The z-score test measures the discrepancy between an experimental observation X from the most probable value, the mean (μ). The obtained result z is represented in terms of the standard deviation (σ). A positive Z indicates that the observation is located above the μ, while a negative Z value suggests that is below μ [78]. Equation 2.9 shows the computation of *z-score* measure. The proposed approach selects possible outliers based on a threshold ($Z = \pm 3$). Each identified outlier was replaced by a *NaN* value.

$$Z = \frac{X - \mu}{\sigma} \tag{2.9}$$

2.4.3. Missing Values Imputation

K-nearest neighbors (KNN) imputation algorithm was implemented to fill in missing data points. The algorithm replaces each missing observation with a value that relies on distance metrics between the k closest neighbors to the missing data point [79]. In the present work, k=5 was set. Therefore, the assigned value to the missing observation relied on the 5 closest neighbors. Euclidean distance (see equation 2.10) was selected as a metric to evaluate the distance.

$$d(x, z) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2} \tag{2.10}$$

2.4.4. Data Standardization

After applying KNN imputer to manage missing values, every observation in the feature dataset was standardized using z-scoring as described in equation 2.9. A *StandardScaler()* implemented in scikit-learn version 1.3.2. in Python was applied.

2.4.5. Data Augmentation

Data augmentation is a technique used in ML problems to increase the quality and diversity of the training dataset [80]. It generates altered variants of the existing data points

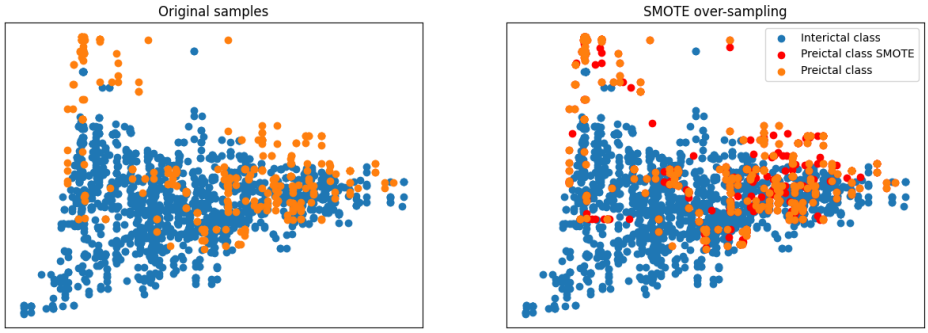


Figure 2.7: Example of synthetic data points generation using SMOTE

while keeping the target category. Data augmentation is also known as a regularization technique used to improve a model's performance. One of the main ML challenges is that may lead to overfit the training stage when the dataset is very small or highly imbalanced, reducing the generalization capability.

Due to the high imbalance between the interictal and preictal class among the patient's data, in order to avoid a possible bias in the prediction and a poor generalization, SMOTE technique was applied to generate new synthetic data points of the minority class in the training set [81]. The new synthetic samples are created along the observation segments by connecting any or all of the k nearest neighbors from the minority class. The choice of neighbors from the k-nearest neighbors depends on the desired level of oversampling. For the present application, k=5 was used. Figure 2.7 illustrates how the SMOTE technique was implemented in the present project. For the present work, undersampling was also performed for the majority class as another way to decrease the imbalance between classes.

2.5. Feature Selection

Feature selection is a very important step when classification or prediction is performed by means of ML or DL tools, it can improve the model's performance increasing the generalization capacity by reducing the model's complexity. This means that redundant or irrelevant features are removed from the dataset. Taking over the most important features could lead to a more interpretable and more accurate model. Also, computational resources and time for modeling as well as prediction are reduced. The present work proposes a three-step feature selection process: **1)** Low variance filtering, **2)** Mann-Whitney U test, and **3)** High correlation filtering.

1. **Low Variance Filtering**: All features whose variance is below 0.01 were pruned

from the feature set.

2. **Mann-Whitney U Test**: It is a non-parametric test used to compare if there is a difference in a dependent variable X among two independent groups (x and y), in the present case, interictal and preictal.
   A null hypothesis (H_0), stipulates that both groups come from the same population, and an alternative hypothesis (H_1), denies the idea that both groups come from the same distribution, are evaluated [82]. The test starts by calculating the U statistics for each group, as follows:

$$U_x = n_x n_y + ((n_x(n_x + 1))/2) - R_x \tag{2.11}$$

$$U_y = n_x n_y + ((n_y(n_y + 1))/2) - R_y \tag{2.12}$$

   Where n_x and n_y are the number of observations in the first and second groups, respectively. R_x and R_y are the sum of the ranks assigned to the first and second group observations, correspondingly. After calculating the U statistics a threshold $\alpha = 0.05$ was defined in order to reject or not the H_0. The rejection of the null hypothesis follows the *p-value* corresponding to the $min(U_x, U_y)$ such that:

$$Reject\ H_0 \text{ if } p \text{ of } min(U_x, U_y) < \alpha \tag{2.13}$$

3. **High Correlation Filtering**: Correlation between features was computed. Pairs with a correlation coefficient $r > 0.8$ were identified. The removed feature from each pair was selected based on the U statistic value, the one with the lowest $U_{x,y}$ value was pruned.

Additionally, the same candidate subset of features obtained after performing the feature selection process in the training set was extracted for the validation and test set.

2.6. Machine Learning and Deep Learning Application

2.6.1. Model Tuning

For the present project, the Supervised Learning method was selected to perform the classification of the periods of interest present along a seizure (interictal and preictal periods). Supervised ML techniques leverage knowledge acquired from past and current

data by using labeled data to predict or classify unseen events. The ML model builds a function to forecast the output values. Each ML algorithm compares the results obtained with the actual input data and the expected result, identifying errors and adapting to improve the outcome. In order to evaluate the generalization capability of the model, a test set, i.e., data never seen by it, should be assessed.

Five models were selected: three ML models (RF, SVC, and XGB) and two DL models (CNN and LSTM). A Parameter Grid was proposed for each of the models. The performance of the algorithms was evaluated using accuracy, F1 score, recall, precision, specificity, and G-mean metrics.

To perform the model hyperparameter tuning and select the model configuration that yields the best performance for the binary classification (interictal = 0, preictal = 1) a Grid Search (GS) was implemented. The GS approach evaluates every hyperparameter combination based on the Parameter Grid defined in advance for each model. The combination of hyperparameters selected is the one that maximizes the macro-averaged F1 score. Due to the high computational cost of GS, the number of hyperparameters was reduced.

Random Forest

RF is a type of ensemble bagging learning method, i.e., a method that combines the predictions of multiple models trained, in this case, classification trees, on different subsets of the training data to improve overall performance and reduce possible overfitting. The data classification is carried out by building and training multiple Decision Tree Classifiers. The algorithm recursively splits the input data based on thresholds among feature values to perform a prediction. This process is guided by a top-down induction approach that uses a *divide-and-conquer* partitioning scheme. The classification trees are built in the training phase of the model. Algorithm 2.2 illustrates in detail the top-down approximation of Decision Trees [83].

Algorithm 2.2 Top-down induction of decision trees. Adapted from[83]

1: In the initialization phase, each observation is placed in the root node of the tree. The root is included in the list of L nodes.
2: If the list L is empty the procedure is stopped, otherwise, a node J belonging to the list L is selected, is removed from the list, and is used as the node for analysis.
3: The optimal splitting rule is used to determine the best criterion for dividing each of the observations contained in node J. After the splitting rule is created, the observations are subdivided, resulting in descendant nodes. A stopping criteria, defined *a-priori*, is evaluated to define whether a node should keep branching or not. If the criteria are met, node J becomes a leaf, and the target class is assigned according to majority voting between the observations contained in the node.

There are three main components of the top-down induction of decision trees previously explained.

1. **Splitting rule**: It refers to the criteria selected to define the optimal splitting rule and perform an optimal branching along the classification tree. In the present RF implementation, to classify interictal and preictal states on an epileptic EEG signal, the **Gini Index** was chosen for this purpose. The Gini index is an impurity measure that evaluates the probability of misclassifying an observation within a node where a label was already assigned to that specific node. It is computed with the following expression:

$$Gini(q) = 1 - \sum_{h=1}^{H} P_h^2 \tag{2.14}$$

   The Gini index acquires its maximum value ($G_{max} = 1$) when all the examples among node J are distributed homogeneously among all classes. Otherwise, ($G_{min} = 0$) is obtained when all the observations belong to the same class. Another impurity measure used in the present work is the *Entropy*.

2. **Stopping criteria**: This criterion helps to establish whether the development of the tree branching should continue or node J should be considered as a leaf. For the present application *maximum depth, minimum number of samples per leaf and minimum samples split* were applied as criteria for this purpose.

3. **Pruning criteria**: It helps to avoid excessive growth of the decision tree during the training phase in order to reduce the chance of overfitting. For example, reducing the number of nodes after the tree has been generated is a way of *post-pruning*.

The hyperparameters chosen to evaluate the Random Forest's performance are shown in table 2.4

Hyperparameter	Description	Values
n_estimators	Number of trees in the forest	[10,30,100]
criterion	Function to measure the quality of the split	["entropy","gini"]
max_depth	Maximun depth of the trees	[5,7,10]
min_samples_split	Minimum number of samples required to split an internal node	[4,6,10]
min_samples_leaf	Minimum amount of samples required at a leaf node	[10,20,50]

Table 2.4: Random Forest Hyperparameter definition

Support Vector Machine Classifier

A Support Vector Machine Classifier or SVC is a type of separation method used for classification problems. It has been shown that is able to achieve good accuracy even with large problems thanks to its efficiency. An SVC model identifies a set of examples, known as *Support Vectors* (SV), which are the most representative observations of each class [83]. These, play the main role in the definition of the separating surface inside the feature space, as they are defined as data points located close to the decision boundary. SV helps to define the classification rule.

SVC uses high-dimensional spaces where data may be easily separated, especially when linear separation is not possible within the data points. This is carried out by means of a "kernel". A kernel function allows the mapping of the original observations into a feature space that is not explicitly computed and that involves an inherent linear segregation of the instances within growing functional spaces with multiple dimensions. The evaluated kernels for the present application are "polynomial" and "radial basis function" (rbf).

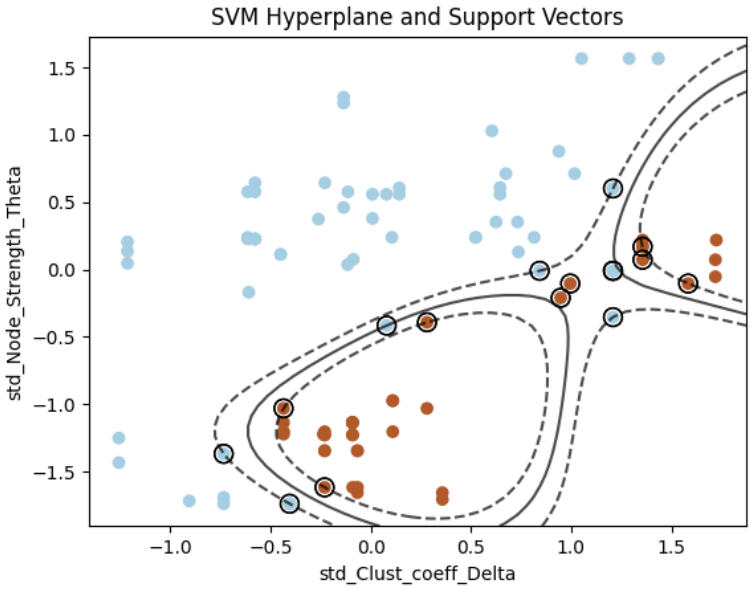


Figure 2.8: SVM Classifier Hyperplane and Support Vectors. Kernel="rbf"

Figure 2.8 shows an example of binary classification using SVM, the hyperplane definition based on the *support vectors* (in black contour). This case presents the separation between feature "$std\_Clust\_coeff\_Delta$" and "$std\_Node\_Strength\_Theta$" for patient P_1. The solid line represents the separation hyperplane. While the dashed line represents the canonical supporting hyperplanes. The kernel used is *RBF* which is based on a radial basis function represented by equation 2.15

$$K(X_1, X_2) = exp(-\frac{\|X_1 - X_2\|^2}{2\sigma^2}) \tag{2.15}$$

The hyperparameters chosen to evaluate the SVC performance are shown in table 2.5

Hyperparameter	Description	Values
kernel	Type of kernel to be used in the algorythm	["poly","rbf"]
C	Regularization parameter	[0.01,0.05,0.1]
gamma	Kernel coefficient for "rbf" and "poly"	["scale"]
degree	Degree of the polynomial kernel function	[2,3,10,20]

Table 2.5: SVM Classifier Hyperparameter definition

XGBoost

The eXtreme Gradient Boosting or, as it is widely known, XGBoost, is a streamlined and scalable implementation of the gradient boosting framework. XGBoost provides a parallel tree boosting that allows a faster performance during the model's training. One of the greatest advantages of the model is its scalability thanks to the underneath algorithmic optimizations that allow the handling of sparse data [84]. In comparison to Gradient Boosting learning algorithm, XGBoost introduces a regularization term (Ω) that controls the complexity of the model in order to avoid overfitting, as shown in equation 2.16 which represents the loss function for this algorithm.

$$L^{(t)} = \sum_{i=1}^{n} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \tag{2.16}$$

Where t is the number of the iteration, y_i represents the real label known from the training set and $\hat{y}_i^{t-1}$ is the predicted label. f_k represents an independent tree structure and $\Omega(f_t)$ is the regularization term that smooths the class weights learned by the XGBoost algorithm.

The hyperparameters chosen to evaluate the XGBoost model performance are shown in table 2.6

Hyperparameter	Description	Values
n_estimators	Number of trees in the forest	[20,50,100]
max_depth	Maximun depth of the trees	[5,10,100]
reg_alpha	L1 regularization term on weights	[40,100]
reg_lambda	L2 regularization term on weights	[0.1,0.2]
colsample_bytree	Subsample ratio of columns when constructing each tree	[0.5,1]
eta	Stop size shrinkage used in update to prevent overfitting	[0.01,0.05]

Table 2.6: XGBoost Hyperparameter definiton

Convolutional Neural Network

In general, a Convolutional Neural Network (CNN) is a form of Artificial Neural Network (ANN), which is a type of learning processing system inspired by the brain's biological

nervous systems. Is composed of an input node, followed by multiple layers, also known as hidden layers, comprised of interconnected computational nodes (neurons), and ends on an output node where the prediction result is presented. The hidden layers intertwine in a distributed manner to collectively learn from the input, aiming to optimize its ultimate output [85].

A CNN is similar to an ANN in the sense that the neurons that compose the network self-optimize through the learning process. Each neuron still processes an input signal or image through operations such as scalar product followed by a non-linear function. For example, an activation Rectified Linear Unit (ReLU) activation function is used in the present implementation as it increases the complexity of the CNN between layers by introducing non-linearity allowing a more complex representation of data. The last layer always contains a loss function associated with the classes, in this case, the *Softmax* function was proposed as the probabilities obtained always sum up 1, which guarantees the model's predictions accurately reflect a probability distribution over the interictal and preictal classes. Figure 2.9 includes the behavior of the activation functions used for the CNN structure proposed.

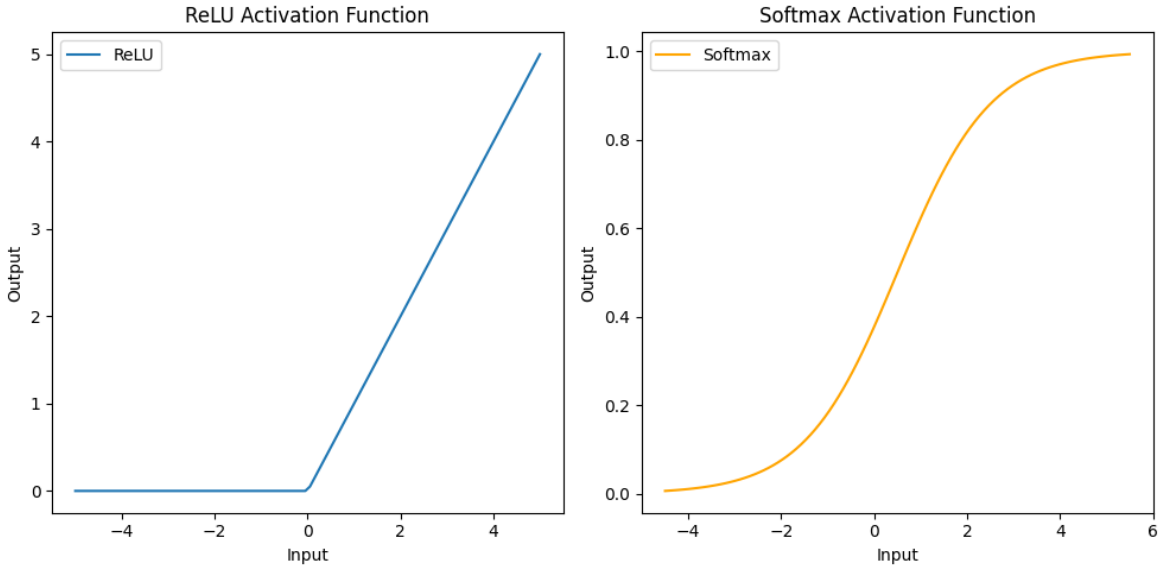


Figure 2.9: Activation functions used in the CNN implementation

CNNs are usually used in pattern identification within images, they are also a well-known forecasting technique in time-series analysis [86]. As the EEG signal presents a high temporal dependency among each sample, this approximation is also proposed. The CNN structure proposed is composed by a 2 *convolutional* + *ReLU* layers, 2 *Batch Normalization* layers that help prevent the overfitting, and, 1 *Dropout* Layer also for this purpose.

Finally, a *Flatten* layer, followed by a *Dense* layer and the *Output* layer with a *Softmax* loss function. as shown in figure 2.10

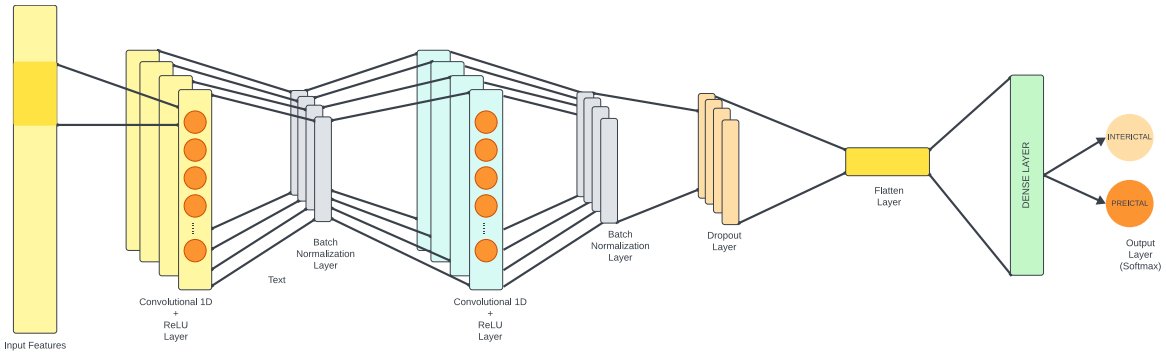


Figure 2.10: CNN proposed structured for preictal state characterization

The hyperparameters chosen to evaluate the CNN model performance are shown in table 2.7

Hyperparameter	Description	Values
batch_size	Type of kernel to be used in the algorithm	[64]
epochs	Regularization parameter	[100,200]
learning_rate	Kernel coefficient for "rbf" and "poly"	[0.001]

Table 2.7: CNN Hyperparameter definiton

Long Short-Term Memory - Recurrent Neural Network

A Long Short-Term Memory (LSTM) network is a derived form of recurrent neural network (RNN), which has the ability to learn order dependencies and patterns in sequence prediction problems from temporal information while preserving the dependencies among long-time sequences. It is highly used in complex problems such as machine translation and speech recognition [87]. The central feature of the LSTM is the constant error carousel (CEC), where information is stored in short-term memory for large periods of time. For the present work, an LSTM is proposed to efficiently classify the periods of interest in the epileptic EEG signal (interictal and preictal periods).

LSTM models work by means of gates that are used to control the data stream within recurrent computations. In order to store long-term patterns of the input signal, four main components build the memory block present on each hidden layer within the network,

which enables the update of each block [88]. Figure 2.11 presents an example of an LSTM memory block structure.

1. **Input gate**: sigmoid threshold units with an activation function ranging between [0,1],. It controls the signal from the network to the memory cell and scales it.
2. **Forget gate**: is responsible for resetting the internal state of the memory cell when the stored information is no longer needed. This allows LSTM performance optimization.
3. **Cell state**: refers to the activation of a CEC by the input gate.
4. **Output gate**: it presents the output of the current hidden state. It learns how to manage access to contents of the memory cell, safeguarding other memory cells from disruptions.

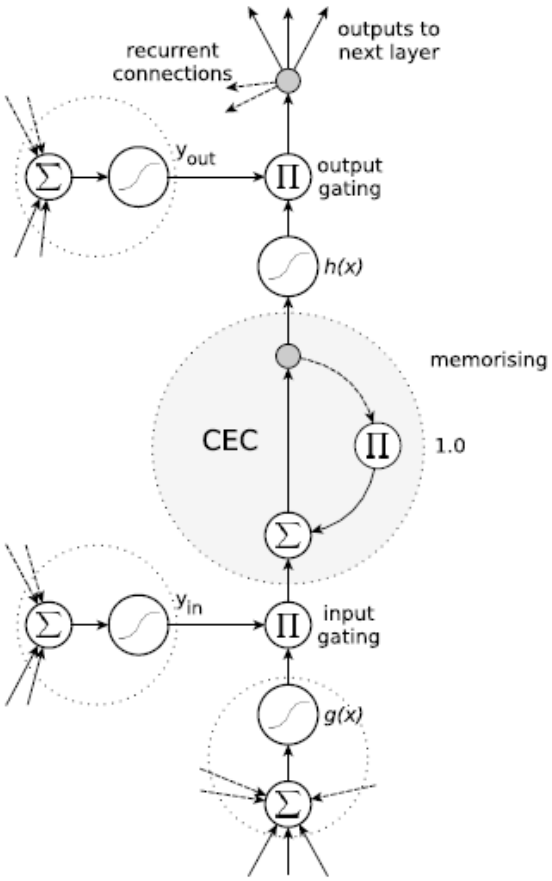


Figure 2.11: Standard LSTM memory block. Taken from [88]

The LSTM structure proposed for the present application is composed of 3 *LSTM* layers, interspersed with a *Dropout* layer to prevent possible overfitting in the training of the

model, a *Dense* layer and the *Output* layer with a loss sigmoid function to perform the binary classification. The proposed architecture is shown in figure 2.12

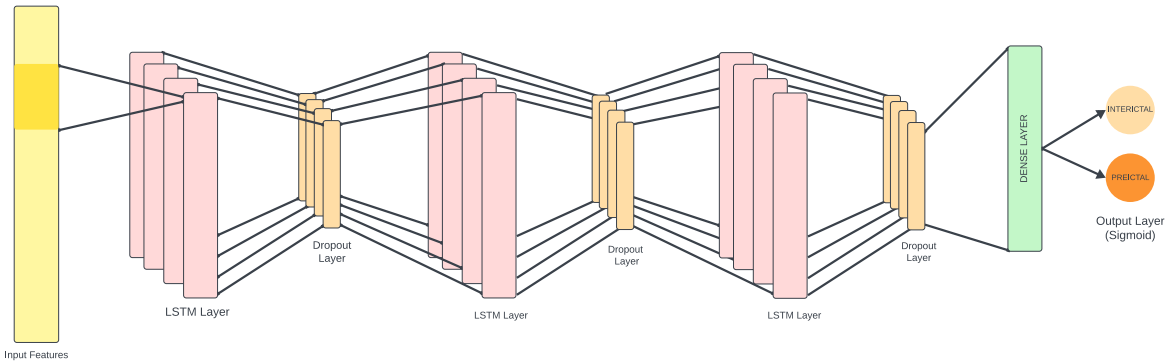


Figure 2.12: LSTM-RNN proposed structured for preictal state characterization

The hyperparameters chosen to evaluate the LSTM-RNN model performance are shown in table 2.8

Hyperparameter	Description	Values
batch_size	Type of kernel to be used in the algorithm	[64]
epochs	Regularization parameter	[200]
learning_rate	Kernel coefficient for "rbf" and "poly"	[0.001]

Table 2.8: Long Short-Term Memory RNN's Hyperparameter definiton

2.7. Model Training and Evaluation

2.7.1. Preictal window selection and Label assignment

To assess the performance of the models for interictal and preictal state classification when exposed to different temporal contexts, different preictal window lengths, i.e., distinct time intervals preceding the SOZ activation time were evaluated. The chosen preictal window lengths were of 40, 60, and 80 minutes. The training of the models was performed for the three preictal windows.

The label assignment for preictal state characterization, using the supervised learning

models proposed, was determined by the chosen preictal window length. Instances occurring outside the chosen preictal window were labeled as interictal state with **label** = 0, which represents a non-seizure state. Conversely, instances falling within the preictal window were labeled as preictal state with **label** = 1, indicating the potential start of a seizure. The preictal window labeling was assigned according to the "eeg onset", i.e., SOZ activation time, information present in the EPILEPSIAE database Metadata.

2.7.2. Dataset splitting

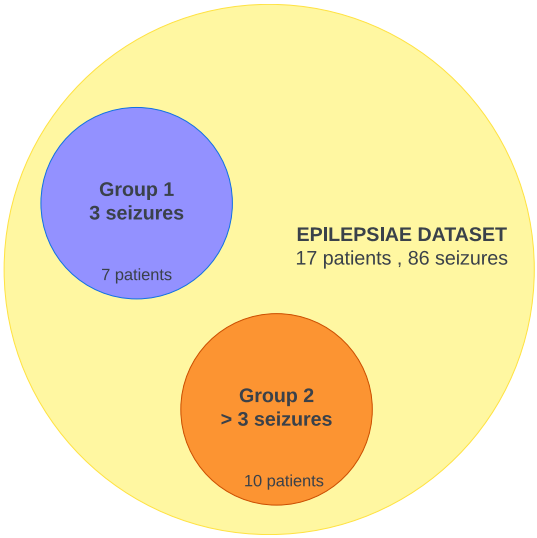


Figure 2.13: Group splitting of the available patients

As shown by the exploratory analysis performed (see figure 2.2) some of the patients only had 3 seizures available among the EEG recording blocks. Taking this into account, for the model training and validation at least 2/3 seizures were going to be used for training and the 1/3 remaining seizure for validation, leaving the model without a test seizure to perform a final evaluation of the model. Therefore, a group splitting to perform the model analysis was carried out, as shown in figure 2.13. **Group 1** (G1), contained the patients with only 3 seizures, it had a total of 7 patients. While **Group 2** (G2) contained all patients with > 3 seizures, for a total of 10 patients. For the latest group, it was possible to build a test set for a final evaluation of the tested models.

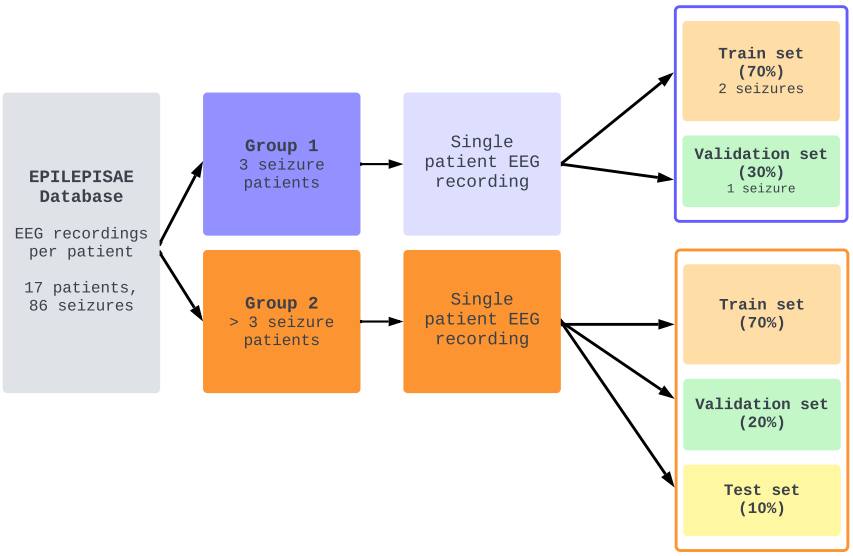


Figure 2.14: Dataset splitting for each group

Following the previous sections, in order to train the selected machine learning and deep learning models, a proper splitting of the dataset into train, validation, and test sets was proposed for each of the groups. To guarantee the correct performance of each model, it was important to ensure that each 480-minute feature window was complete on every set and that no window was duplicated on other sets, this would incur data leakage. To preserve seizure-window consistency, *Group-KFold* cross-validation was proposed to perform the partition of each set. Figure 2.14 shows the proposed dataset splitting.

For **G1**, the ***train set*** contained 70% of the seizures, while the remaining ***validation set*** contained 30%. For **G2**,the ***train set*** contained 70% of the seizures, the ***validation set*** 20%, and, the ***test set*** the remaining 10%.

Classification metrics

The Confusion Matrix (CM) is a performance report used to outline possible errors in a binary classification problem on a given dataset. The table presents a contrast between the labels predicted by the classification model and the actual labels of the data. Figure 2.15 shows the structure of a confusion matrix for a binary classification problem.

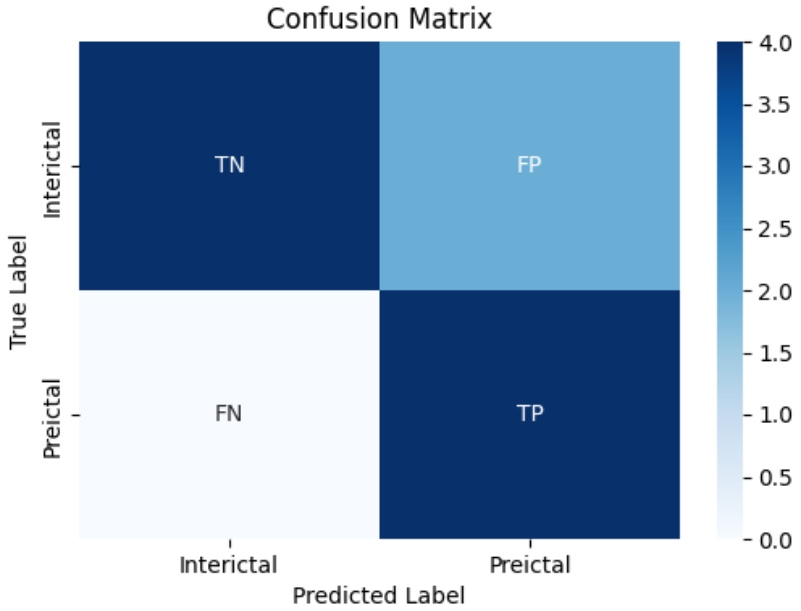


Figure 2.15: Confusion Matrix for Interictal (0) and Preictal (1) classification periods

The "TN" label stands for True Negative, which represents the number of correctly classified negative cases. Similarly, "TP" denotes True Positive, indicating the count of well-classified positive examples. A False Positive is represented with "FP", that is the count of actual negative examples wrongly classified as positive. Finally, "FN" stands for False Negative, which represents the number of real positive examples mistakenly classified as negative.

Based on the Confusion Matrix, overall metrics such as accuracy, F1-score, recall, specificity, and sensitivity can be obtained, which gives more information about the performance of the evaluated model. Table 2.9 presents the metrics computed to evaluate each model's classification performance.

Metric	Description	Formula
Accuracy	Proportion of well-classified instances out of the total instances	$\frac{TP+TN}{TP+TN+FP+FN}$
F1-score	Harmonic mean of precision and recall	$2 \times \frac{Precision \times Recall}{Precision+Recall}$
Precision	Proportion of TP predictions out of all positive predictions	$\frac{TP}{TP+FP}$
Recall (Sensitivity)	Proportion of TP predictions out of all true positive instances	$\frac{TP}{TP+FN}$
Specificity	Ability to correctly classify negative instances	$\frac{TN}{TN+FP}$

Table 2.9: Classification metrics computed based on the Confusion Matrix

Further classification metrics such as G-mean, and other metric tools like the ROC curve and Precision-Recall curve were also evaluated to assess the model´s performance.

- **Geometric mean of recall (G-mean)**: The geometric mean is a measure of central tendency that is calculated by multiplying together a set of numbers and then taking the nth root of the product. In the context of ML imbalanced problems, the G-mean of the recall works effectively in evaluating a model's performance of the target classes in a proper way. The G-mean of recall seeks to achieve the balance between majority and minority classes by calculating the square root of the product between recall and specificity[89].

$$G - mean = \sqrt{Recall \times Specificity} \tag{2.17}$$

- **ROC Curve**: The ROC curve is a graphical representation illustrating the balance between the true positive rate (sensitivity) and the false positive rate (one minus true negative rate, so the specificity) for a classifier across different decision thresholds. This curve is a valuable tool in assessing the performance of a classifier as the threshold for decision-making varies in order to: (1) establish an optimal decision threshold that minimizes error rates or misclassification costs, considering specific class and cost distributions; (2) identify areas where one classifier demonstrates superior performance compared to another; (3) identify regions where a classifier performs worse than chance; and (4) acquire well-calibrated estimates of the class

posterior [90].

- **Precision-Recall curve (PR)**: The precision-recall curve is a graphical representation that illustrates the trade-off between precision and recall for different thresholds of a binary classification model. It's obtained by smoothly changing the decision threshold from one to zero and calculating for each value of the cutoff threshold the Precision and Recall. The precision is the ratio of true positive predictions to the total number of positive predictions made by the classifier while the ratio of true positive predictions to the total number of actual positive instances. Is particularly useful when dealing with imbalanced datasets and it is used to find a balance that optimizes both precision and recall, depending on the specific requirements of the application [91].

2.7.3. Model Selection

In order to evaluate each of the ML models, and preserve the seizure-window consistency within the dataset splitting, a 3 Group-KFold cross-validation was proposed to perform the splitting of each set. GroupKFold ensures that entire groups of related data points, meaning that the data points chosen for the split belong to the same 480-feature vector, and therefore remain intact for training, validation, and test set during each iteration of the cross-validation process.
To manage the trade-off between precision and recall for the predicted values a custom prediction threshold for the validation test to assess the classification for the interictal and preictal states was set. This threshold was computed over each fold. As imbalanced data was being used, the threshold selected was the one that maximized the F1 score for the predicted values in the validation set. This threshold was then applied to the test set, if possible. Finally, The best-performing model and best preictal window selected were the ones that also maximized the F1 macro score in the Validation set. The evaluation metrics mentioned in 2.7.2 were computed by contrasting the real and predicted label values.

3 | Results

The preceding chapters have outlined the theoretical framework, methodology, and data collection steps carried out in the present investigation. The study's findings are presented in this chapter, detailing the observed outcomes and their implications. The purpose of the study was to design a pipeline to perform *patient-specific* pattern recognition in brain networks in order to characterize and understand the behavior of *preictal* states in epileptic patients. The epileptic state classification was performed by implementing different machine learning classifiers such as RF, SVM Classifier, XGB, CNN and LSTM (see 2.6.1), to identify ***Interictal*** and ***Preictal*** labels varying the length of the analyzed *preictal* time window (40, 60 and 80 minutes).

In the first step, as presented in Section 2.3.1, each EEG recording block was concatenated until obtaining a full 8-9h recording, for each of the EEG bands. Then the connectivity matrix was computed using Pearson's Correlation and multiple Brain Network features were computed (see Table 2.3). After, the preprocessing steps explained in Section 2.4 in order to prepare the data for the ML pipeline were executed. Finally, RF, SVC, XGB, CNN, and LSTM were evaluated for each of the patients.

The result analysis was performed based on the two study groups proposed in 2.7.2. **Group 1** (G1) contains the patients with only 3 seizures among the dataset, while **Group 2** (G2) the patients with more than 3 seizures. The best-performing model and best preictal window selected were the ones that maximized the F1 macro score in the Validation set.

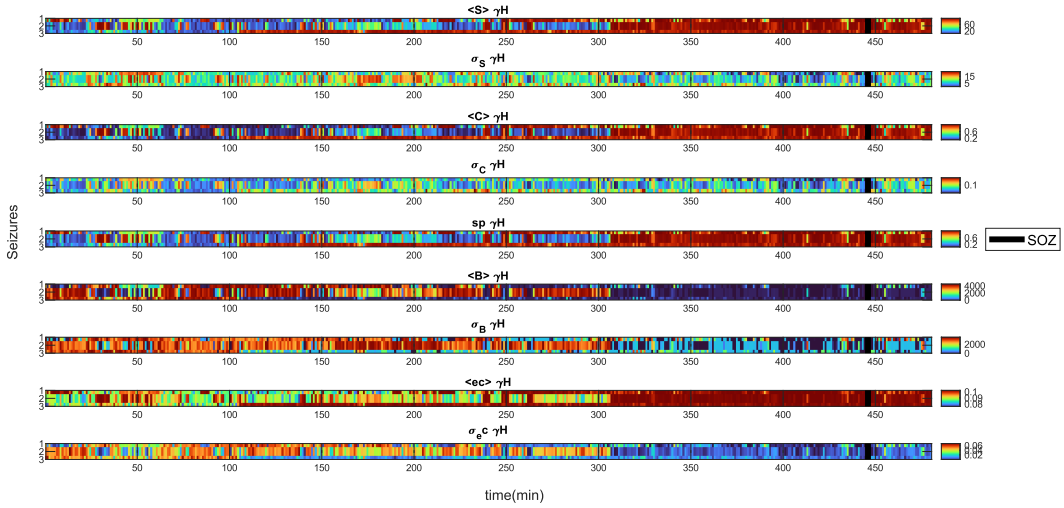
3.1. Feature Extraction

3.1.1. Functional Connectivity Analysis and Network Feature Extraction

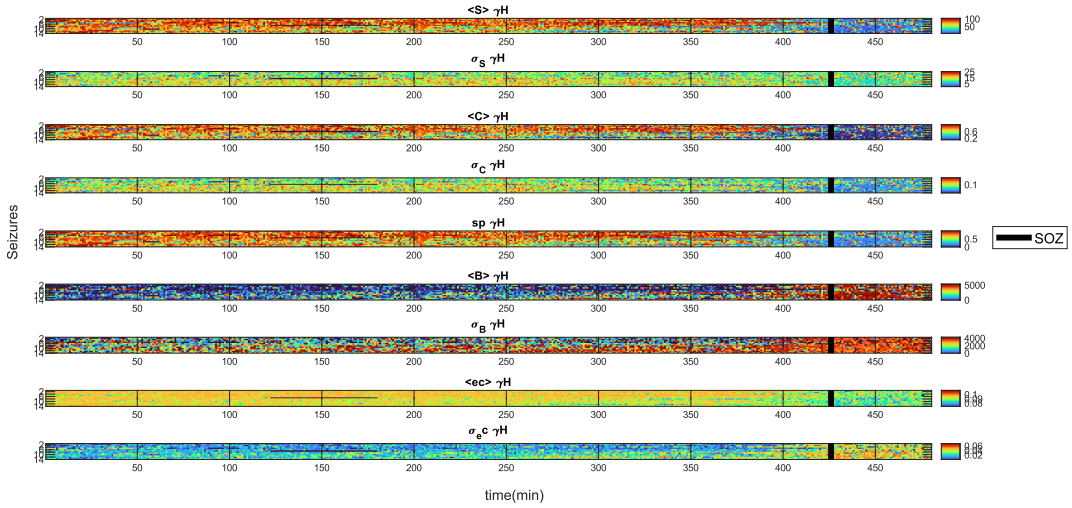
Once the 8-9h EEG recording block was obtained, the 11 Brain Network features were computed based on the connectivity matrix extracted per frequency band. This allowed to gain a closer insight into the behavior of the network in time from which it is expected to be able to extract information that may indicate a potential seizure onset. Figure 3.1a and Figure 3.1b show an example of the time evolution of the first 9 features for Group 1 P_5 patient and Group 2 P_3 patient for γH band. Each pixel among the *heatmap* indicates the magnitude of each feature per minute instance, allowing the identification of possible patterns or trends in the data.

In the case of P_5, for features such as Node Strength ($< S >$), Clustering Coefficient ($< C >$), Average Shortest Path (sp), and Eigenvector Centrality ($< ec >$), the magnitude of the measurement increased its value as it got closer to SOZ activation time. For the Betweenness Centrality ($< B >$) feature, the magnitude seemed to decrease as it got closer to the SOZ activation time. On the other hand, P_3 showed that for features such as $< S >$, $< C >$, sp, and $< ec >$, the magnitude value is reduced as it gets closer to SOZ activation time. For feature $< B >$, the magnitude seems to increase as it gets closer to the seizure onset zone activation time. Compared to figure 3.1a, the behavior of some of these characteristics is the opposite for P_5. This variability in the brain dynamics during seizure events across patients highlights the complex and *patient-specific* nature of epileptic events.

Finally, every patient inside G1 and G2, 480 feature vectors were obtained per every seizure instance. The total number of feature vectors used to train and validate each of the models for each group was computed as follows: $480 \times \#of seizures = \#\ of\ feature\ vectors$



(a) Group 1 P_5 patient's network feature evolution in time for in γH band compared with all 3 seizures.



(b) Group 2 P_3 patient's network feature evolution in time for in γH band compared with all 3 seizures.

Figure 3.1: Network feature evolution in time in γH band.

3.2. Feature Preprocessing

Median Filtering, Outlier removal, KNN imputation for missing values, Data standardization, and Data augmentation, were applied for both groups, G1 and G2.

3.2.1. Data Augmentation

In order to increase the quality and diversity of the training dataset, as well as reduce the high imbalance existing between classes for the present problem, the SMOTE technique was applied (see Section 2.4.5). A reduction of the majority class was also implemented by an undersampling method.
Figure 3.2 shows the increase for the preictal class after implementing the SMOTE technique in both cases. The same approach was applied to patients in G2. To compensate for the remaining imbalance, custom class weights were computed for each class and then were configured for each of the models to avoid possible incongruities. The percentage of oversample and undersample ratio for the instances depended on the length of the current preictal window evaluated (40, 60, or 80 minutes), as the number of samples for each of the classes, interictal (0) and preictal (1), changes with the selected window. This approach allowed the reduction of the imbalance gap between the classes.

Once all Brain Network features were computed, and the splitting into train and validation set was carried out, as described in Section 2.3.1 other preprocessing steps, such as feature median filtering, outlier removal, missing values imputation, data standardization, data augmentation, and feature selection, were implemented for each of the patients' dataset.

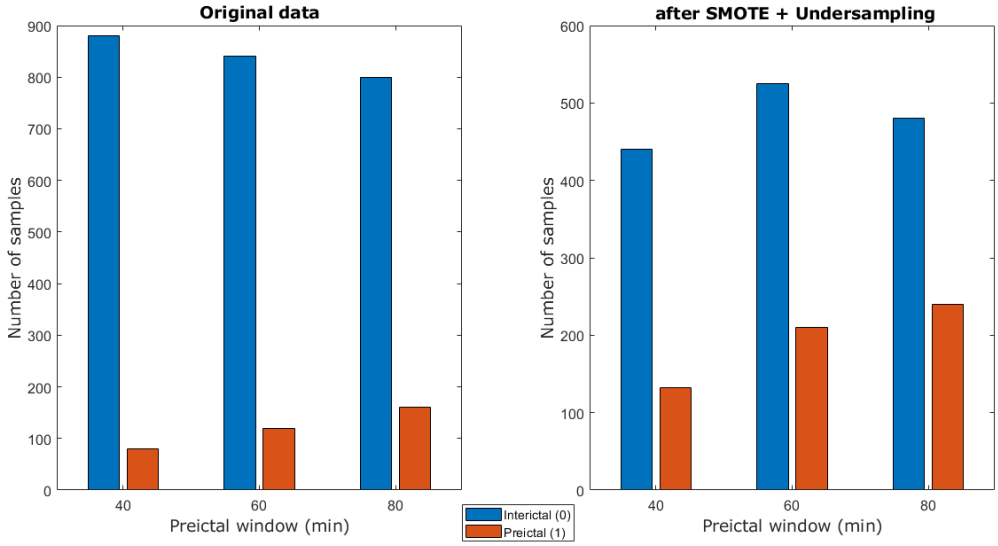


Figure 3.2: SMOTE and Undersampling sample behavior in Train set for patients in G1, with Majority (0) and Minority (1) Class.

3.3. Feature Selection

As detailed in Section 2.5, the feature selection involved three procedures: i) Low Variance Filtering, ii) Mann-Whitney U Test, and iii) High Correlation Filtering. This iterative process was applied to each of the 3-Folds within the implemented GroupKFold method for G1 and G2. Figure 3.3 shows the Top 5 of the most frequently selected features within the 3-Kfoid process, for G1 and G2. It can be noted that features $< ec >$ and σ_{ec} of γH band, as well as, $< C >$ and σ_C of δ band, are the most recurrent selected features during the training process across patients for both groups, which suggests their potential high significance in distinguishing the distribution of the two class groups (interictal and preictal).

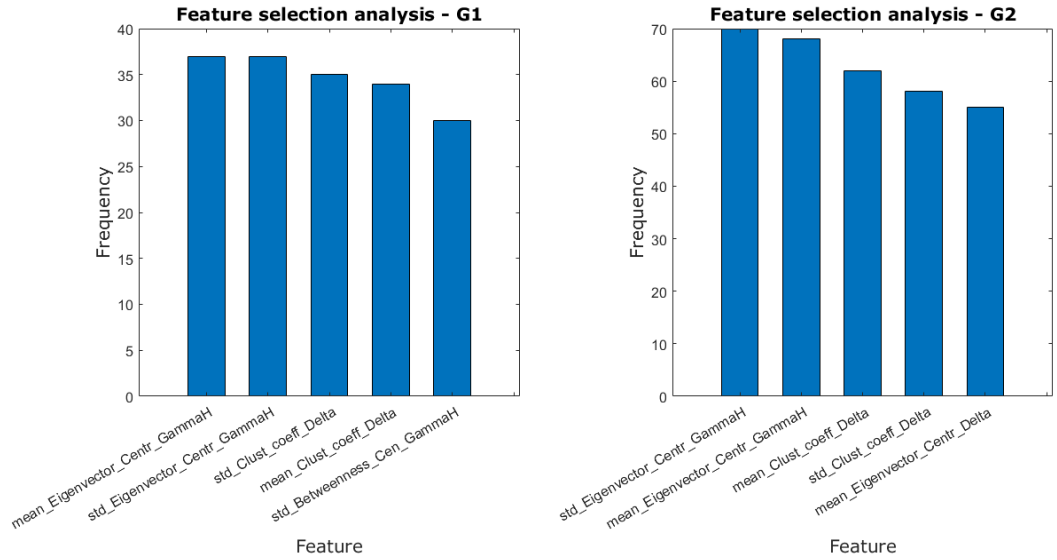


Figure 3.3: Top 5 - Feature frequency selection in training set over the 3-GroupKFold for G1 and G2.

3.4. Model Training and Evaluation

Once the final patient dataset was obtained, each of the models proposed in Section 2.6, RF, SVM Classifier, XGB, CNN, and LSTM were trained and evaluated with a validation set. Multiple hyperparameters were proposed for each of the models, the best-performing model parameters were selected by using the Grid Search method for parameter tuning (see Section 2.6.1). A different model was then obtained for each of the 3-folds, and each was evaluated over the validation set proposed. The model with the highest *F1 score* value and not overfitting in the validation set was selected for each of the preictal windows evaluated. Additionally, as mentioned in Section 2.7, a custom prediction threshold for

the validation test, was computed over each fold to manage the trade-off between precision and recall for the predicted values. The threshold selected was the one that maximized the F1 score for the predicted values.

The results are individually presented for one of the patients inside G1 and G2, with the purpose of highlighting the entire workflow executed for each specific patient based on their assigned group. This first approach allows a comprehensive understanding of the tailored ML-trained models, as well as the analysis conducted for patients within each distinct group.

3.4.1. Group 1

The first patient group, Group 1, that was evaluated with the designed pipeline was composed of 7 patients with 3 seizure recordings each. As the number of seizures was not enough to create a test set, only train and validation sets were configured. The train set contained 2 of the seizures, while the remaining one was on the validation set. For G1 the selected patient was P_5 The following tables (3.1,3.2,3.3) show the assessed metrics Accuracy, F1 score, Precision, Recall, Specificity, and G-mean, obtained for the training and validation set, obtained for P_5 for each of the evaluated models for the preictal window 40, 60, and 80 minutes.

Table 3.1, provides an overview of the five trained models and their performance on the train and validation sets, over a preictal window of 40 minutes. A prediction threshold of **0.52** was the best value selected that maximized the F1 score. Among all the evaluated models, RF, SVC, and LSTM an F1 score value higher than 60% was obtained for P_5. XGB obtained a 100% score, for all metrics on the training set, which suggested overfitting on the training set. For the 40-minute window, the best-performing model was SVC, with an F1-score of 71.37% on the validation set. The drop for this metric compared to the train set was not very high.

	Preictal window	40 min									
	Model	RF		SVC		XGB		CNN		LSTM	
		Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set
Metric	Accuracy	92.31%	80.97%	85.43%	83.61%	100%	70.14%	92.31%	73.75%	84.85%	74.58%
	F1 Score	90.20%	67.89%	81.97%	71.37%	100%	58.78%	85.95%	59.94%	79.49%	61.18%
	Precision	75.22%	36.07%	63.53%	40.26%	100%	30.16%	95.45%	23.87%	68.16%	24.17%
	Recall	99.49%	79.17%	91.41%	88.33%	100%	79.17%	71.21%	82.50%	77.78%	94.17%
	Specificity	90.15%	81.14%	83.64%	83.18%	100%	69.32%	98.64%	72.95%	86.97%	72.80%
	G-mean	94.71%	80.15%	87.44%	85.72%	100%	74.08%	83.81%	77.58%	82.24%	82.80%

Table 3.1: Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_5 for a 40-minute preictal window. In red, the best values for each metric in the validation set.

Table 3.2 summarizes the performance of all models obtained for patient P_5 for the second preictal window evaluated, 60-minute. A prediction threshold of **0.48** was the best value selected that maximized the F1 score. Upon the obtained results, once again, SVC presented a better trade-off between precision and recall, obtaining an F1 score of 73.31% also, no overfitting was shown for the training set. Models like CNN and XGB showed an increase in the F1 score with respect to the previous preictal window evaluated. Additionally, precision increased a 5% with respect to the previous 40-minute preictal window, indicating that the prediction of positive instances was better for the current 60-minute window.

	Preictal window	60 min									
	Model	RF		SVC		XGB		CNN		LSTM	
		Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set
Metric	Accuracy	94.74%	77.71%	85.99%	85.35%	99.18%	81.74%	95.83%	75.21%	84.22%	80.42%
	F1 Score	93.83%	69.09%	83.69%	73.31%	99.00%	68.17%	95.17%	66.61%	72.14%	64.64%
	Precision	85.77%	40.25%	71.87%	45.16%	98.44%	36.69%	89.48%	35.64%	59.10%	38.98%
	Recall	98.10%	83.89%	87.78%	77.22%	98.73%	69.44%	98.41%	90.56%	51.11%	68.89%
	Specificity	93.40%	76.83%	85.27%	86.51%	99.37%	83.49%	94.79%	73.02%	97.46%	82.06%
	G-mean	95.72%	80.28%	86.51%	81.73%	99.05%	76.14%	96.59%	81.31%	70.58%	75.19%

Table 3.2: Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_5 for 60-minute preictal window. In red, the best values for each metric in the validation set.

Table 3.3 shows the highest F1 score value obtained with the 80-minute preictal window. An F1 score of 75.5%, was obtained with an RF model, therefore this configuration was selected as the best-performing one for P_5. A prediction threshold of **0.53** was the best value selected that maximized the F1 score. The XGB model obtained an F1 score value greater than 70%, however, all training set metrics were greater than 99%, hence potential overfitting could be seen in the train set. For SVC, CNN, and LSTM models precision and recall decreased with respect to the 60-minute window.

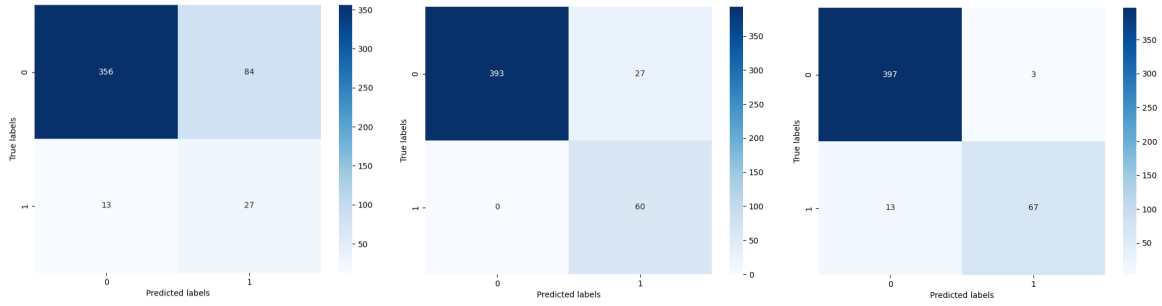
	Preictal window	80 min									
	Model	RF		SVC		XGB		CNN		LSTM	
		Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set
Metric	Accuracy	95.74%	81.94%	80.56%	75.76%	99.35%	81.32%	85.00%	74.72%	83.47%	75.35%
	F1 Score	95.34%	75.54%	76.47%	68.50%	99.28%	72.11%	78.76%	66.77%	79.28%	68.36%
	Precision	89.32%	56.69%	73.93%	41.22%	98.52%	47.77%	94.73%	38.17%	79.35%	39.56%
	Recall	99.17%	78.75%	74.03%	81.67%	99.58%	69.17%	60.00%	76.67%	64.17%	83.75%
	Specificity	94.03%	82.58%	83.82%	74.58%	99.24%	83.75%	97.50%	74.33%	93.13%	73.67%
	G-mean	96.56%	80.64%	78.77%	78.04%	99.41%	76.11%	76.49%	75.49%	77.30%	78.55%

Table 3.3: Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_5 for 60-minute preictal window. In red, the best values for each metric in the validation set.

Figure 3.4 shows the Confusion matrices for each of the evaluated windows (40, 60, and 80 minutes) for P_5, confirming RF as the best classifier for the 80-minute window. It can

be seen how the False Negative Rate decreases when the window is increased, as more samples of the preictal class are assigned to the window. Also, a low False Negative Rate (FNR) was achieved, indicating an optimal Recall value, among all the preictal windows.

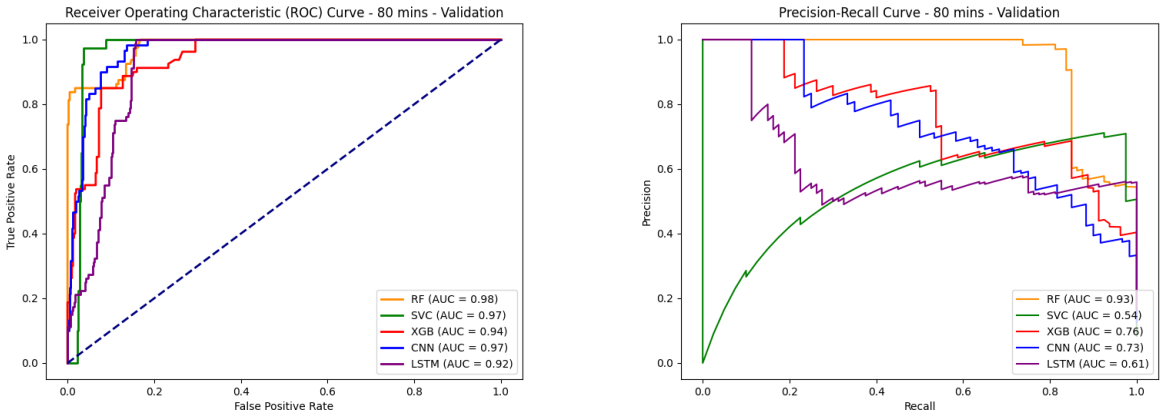
An additional control to select the best window and model was carried out by reviewing the ROC curve. As the number of negative instances significantly outweighed the positive instances, the ROC curve may give a falsely optimistic view of the model's performance. Thus, the Precision-Recall curve was also evaluated as it is less biased skewed class distribution. Figure 3.5 presents the ROC and PR curves for the RF model of the best-evaluated fold in the pipeline for the 80-minute preictal window.



(a) 40-minute preictal window. (b) 60-minute preictal window. (c) 80-minute preictal window.

Figure 3.4: Confusion Matrices of RF best-performing model for P_5 across different preictal windows for the Validation set.

The ROC-AUC value obtained was greater than 95%. An AUC value for the PR curve of 0.93 was obtained for the RF model, which indicates strong performance in terms of precision and recall trade-offs. The model was effective at identifying positive instances while maintaining a low false positive rate.



(a) ROC-AUC curve.

(b) Precision-Recall curve.

Figure 3.5: ROC-AUC and PR curve for the best preictal window, 80-minute, for the *Validation set* for all the evaluated models of P_5.

Similarly, the model training and evaluation process was carried out for every patient in Group 1. Table 3.4 presents the performance metrics obtained for the best-performing preictal window and model per patient. It can be observed that 6/7 patients chose the 80-minute preictal window as the optimal one. The XGB was the best-performing model with the most incidences among patients.

				Metric					
Patient ID	Number of Seizures	Model	Preictal window [min]	Accuracy	**F1 score**	Precision	Recall	Specificity	**G-mean**
P_4	3	RF	80	79.79%	70.27%	45.83%	67.92%	82.17%	74.70%
P_5	3	SVC	80	81.94%	75.54%	56.69%	78.75%	82.58%	80.64%
P_6	3	LSTM	40	78.68%	67.69%	36.35%	88.33%	77.80%	82.90%
P_7	3	SVC	80	64.17%	47.70%	17.52%	30.42%	70.92%	46.44%
P_{12}	3	RF	80	60.42%	52.49%	28.07%	54.58%	61.58%	57.98%
P_{15}	3	LSTM	80	74.20%	66.70%	37.96%	78.33%	73.36%	75.81%
P_{17}	3	CNN	80	79.24%	62.16%	36.26%	37.50%	87.58%	50.31%

Table 3.4: Best-performing model and preictal window performance metrics for each patient in G1. In red, are the patients with F1 score greater than 65%.

3.4.2. Group 2

For the case of group 2, seizure recordings for 10 patients were available. It is very important to acknowledge the variation in the number of seizures across the different patients, which led to an inherent disparity in the amount of epileptic data available for training, validation, and test sets. The selected patient to illustrate the performance

evaluation of the proposed workflow evaluated for every patient in G2 was P_3, this patient had the highest number of seizure instances among the patients.

Table 3.5 presents a summary of the performance of the five trained models on both the training and validation sets within a preictal window of 40 minutes for P_3. The selected optimal prediction threshold for maximizing the F1 score was **0.46**. Notably, the SVC, XGB, and LSTM models achieved an F1 score of over 70%. Conversely, the RF and CNN models exhibited an F1 score greater than 65%. Among the models evaluated for the 40-minute window, LSTM emerged as the top performer, surpassing a 72% F1 score, and the decline of this metric compared to the training set was relatively moderate which discards a possible overfitting of the model.

	Preictal window	40 mins								
	Model	RF		SVC		XGB		CNN		LSTM	
		Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set
Metric	Accuracy	93.07%	82.52%	88.10%	85.92%	98.70%	85.89%	90.40%	82.73%	82.85%	86.56%
	F1 Score	91.10%	68.77%	84.71%	71.89%	98.23%	70.78%	78.61%	67.64%	79.70%	72.73%
	Precision	77.00%	32.59%	69.65%	37.28%	94.86%	35.39%	59.62%	31.20%	57.92%	38.76%
	Recall	99.81%	94.17%	88.70%	88.13%	100%	84.17%	66.48%	85.42%	94.13%	87.92%
	Specificity	91.04%	81.46%	87.92%	85.72%	98.31%	86.04%	97.58%	82.48%	79.47%	86.44%
	G-mean	95.33%	87.58%	88.31%	86.91%	99.15%	85.10%	80.54%	83.94%	86.49%	87.17%

Table 3.5: Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_3 for 40-minute preictal window. In red, the best values for each metric in the validation set.

Similarly, the performance evaluation of the models for a 60-minute preictal window was carried out. Table 3.6 displays the evaluation metrics obtained for each of the models for P_3. The best prediction threshold implemented was of **0.36**. The threshold selection allows to manage the trade-off between precision and recall for the predicted values. All models obtained an F1 score greater than 60% which indicates an acceptable trade-off between precision and recall. The models are achieving a moderate balance in minimizing both false positives and false negatives. The highest score was obtained by the SVC model, with an F1 value above 68% while.

	Preictal window	60 mins								
	Model	RF		SVC		XGB		CNN		LSTM	
		Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set
Metric	Accuracy	92.20%	77.57%	84.85%	80.26%	99.84%	76.04%	83.70%	78.94%	99.83%	82.31%
	F1 Score	91.07%	67.05%	82.02%	68.76%	99.81%	64.92%	75.26%	66.95%	99.79%	65.28%
	Precision	78.95%	34.80%	71.36%	36.86%	99.49%	32.54%	82.07%	34.39%	99.49%	37.16%
	Recall	99.10%	82.92%	80.00%	77.92%	99.96%	77.92%	55.87%	75.00%	99.92%	48.47%
	Specificity	89.41%	76.81%	86.79%	80.60%	99.79%	75.77%	94.83%	79.50%	99.79%	87.14%
	G-mean	94.16%	79.80%	83.33%	79.24%	99.88%	76.84%	72.79%	77.22%	99.86%	64.99%

Table 3.6: Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_3 for 60-minute preictal window. In red, the best values for each metric in the validation set.

Finally, the 80-minute window performance for all the models was evaluated. Table 3.6 reflects the performance metrics obtained for P_3, this gives information about the behavior of the ML and DL models in the preictal period characterization task. The best prediction threshold selected for the current window was of **0.42**. Again, all models obtained an F1 score that exceeded 60% displaying a reasonable performance in the classification task. The CNN model with an F1 score of 66.5% was the best-performing model trained for the longest preictal window evaluated in the present work.

	Preictal window	80 mins									
	Model	RF		SVC		XGB		CNN		LSTM	
		Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set	Train set	Validation set
Metric	Accuracy	88.74%	72.26%	81.45%	69.62%	99.63%	74.86%	84.02%	74.22%	99.50%	74.91%
	F1 Score	87.71%	64.46%	79.30%	63.21%	99.58%	65.83%	75.59%	66.55%	99.44%	61.48%
	Precision	78.65%	35.00%	71.22%	35.70%	99.14%	37.35%	91.64%	37.97%	98.70%	33.53%
	Recall	90.56%	76.46%	74.97%	80.73%	99.76%	70.52%	59.62%	77.40%	99.83%	48.65%
	Specificity	87.83%	71.42%	84.69%	67.40%	99.57%	75.73%	96.22%	73.58%	99.34%	80.17%
	G-mean	89.18%	73.89%	79.68%	73.76%	99.66%	73.08%	75.74%	75.47%	99.58%	62.45%

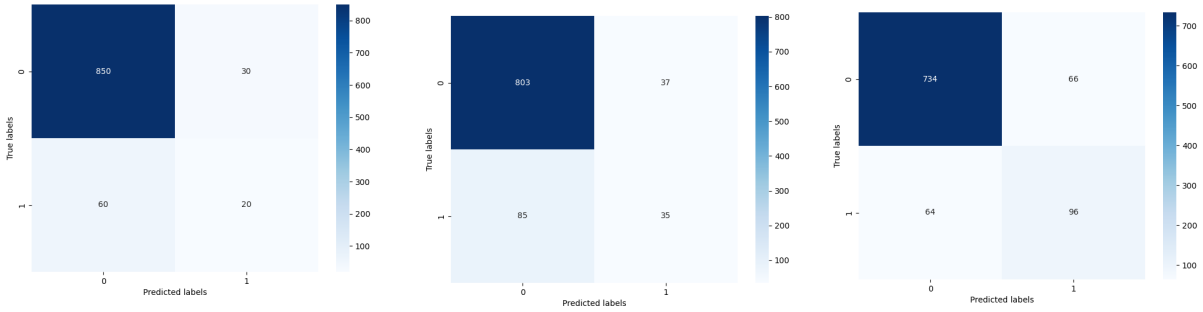
Table 3.7: Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_3 for 80-minute preictal window. In red, the best values for each metric in the validation set.

The LSTM model achieved the highest F1 score in the 40-minute window, leading to the selection of this configuration as the best-performing model for patient P_3. It is important to remark that in order to keep the coherence and avoid any possible data leakage, the same preprocessing pipeline (i.e., the coefficients for outlier removal, standardization, and missing values imputation) applied to the training set was fitted to the test set, before model evaluation. Finally, the selected model was evaluated on this preprocessed test set. To further evaluate the accuracy of the selected model in the present classification task, the model was evaluated over the 40, 60, and 80-minute preictal window for every patient using the test set. Table 3.8 presents the obtained metrics.

	Model	LSTM		
	Preictal Window	40 min	60 min	80 min
Metric	Accuracy	90.63%	87.29%	79.58%
	F1 Score	62.87%	64.70%	61.49%
	Precision	40.00%	48.61%	37.32%
	Recall	25.00%	29.17%	33.13%
	Specificity	96.59%	95.60%	88.88%
	G-mean	49.14%	52.80%	54.26%

Table 3.8: Test set performance metrics LSTM model to classify pre-ictal states for 40, 60, and 80-minute preictal window for P_3. In red, are the best values for each metric in the test set.

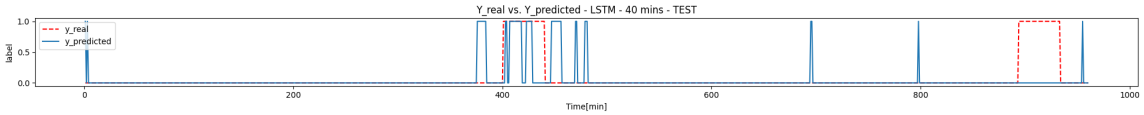
Figure 3.4 displays the Confusion matrices for the assessed windows (40, 60, and 80 minutes) for P_3, validating LSTM as the best-performing classifier specifically for the 80-minute window.



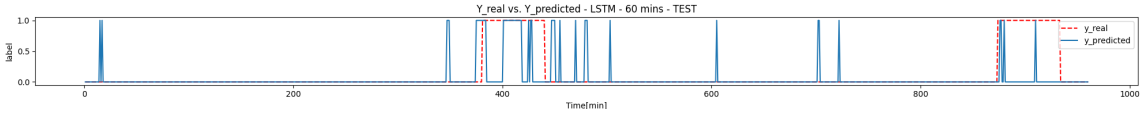
(a) 40-minute preictal window. (b) 60-minute preictal window. (c) 80-minute preictal window.

Figure 3.6: Confusion Matrices of LSTM best-performing model for P_3 across different preictal windows for the Test set.

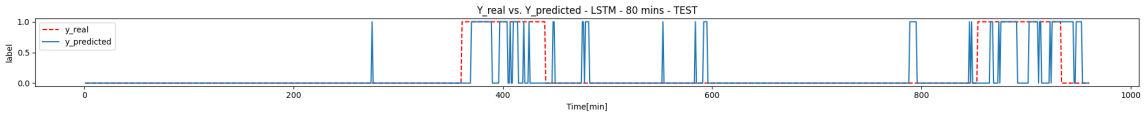
An additional evaluation of the specificity and recall of the LSTM model to characterize preictal states in epileptic patients for every time window was carried out by plotting the vector of predictions against the real values. This approach was able to complement the quantitative metrics and provides valuable insights for model refinement and decision-making about the obtained results, allowing an overview of the overall trend or pattern in the prediction for each of the evaluated windows, as shown in Figure 3.7. Additionally, systematic errors can be identified, such as the ones seen in the first minutes of the interictal period, misclassified points.



(a) 40-minute preictal window.



(b) 60-minute preictal window.



(c) 80-minute preictal window.

Figure 3.7: Prediction vector evaluation of LSTM best-performing model for P_3 across different preictal windows for the Test set.

Finally, G2 implemented the same additional method to determine the optimal window and model by examining the ROC curve and the PR curve. The ROC-AUC obtained was 0.92. Furthermore, an AUC value of 0.61 was achieved for the PR curve with the LSTM model.

In the same way, the training and evaluation process of each of the models was performed for every patient among G2. Table 3.9 presents the performance metrics achieved for the best-performing model and the best preictal window identified for each patient after performing the model's evaluation on each test set. It can be observed that P_3 despite being trained on a dataset containing the highest number of seizure instances, the model did not achieve optimal metrics. SVC model obtained the highest F1 score, 80.10%, and it was also the best-performing model preferred by 40% of the patients. LSTM model was selected by 3/10 patients, also with short preictal windows, 40 and 60-minute. In general, the optimal window selection was highly variable among patients due to the seizure instances available for each subject.

				Metric					
Patient ID	Number of Seizures	Model	Preictal window [min]	Accuracy	**F1 score**	Precision	Recall	Specificity	**G-mean**
P_1	4	RF	80	66.67%	54.79%	24.03%	46.25%	70.75%	57.20%
P_2	7	SVC	80	78.75%	69.93%	42.14%	73.75%	79.75%	76.69%
P_3	14	LSTM	60	87.29%	64.70%	48.61%	29.17%	95.60%	52.80%
P_8	5	SVC	40	82.92%	59.80%	22.37%	42.50%	86.59%	60.66%
P_9	5	RF	80	85.21%	74.97%	54.84%	63.75%	89.50%	75.54%
P_{10}	8	LSTM	40	93.75%	70.57%	85.71%	30.00%	99.55%	54.65%
P_{11}	5	LSTM	60	78.17%	60.84%	27.72%	46.67%	82.66%	62.11%
P_{13}	4	SVC	80	89.17%	80.10%	68.42%	65.00%	94.00%	78.17%
P_{14}	7	SVC	80	66.67%	53.98%	22.97%	42.50%	71.50%	55.17%
P_{16}	6	CNN	60	71.27%	56.44%	22.83%	48.33%	74.81%	60.13%

Table 3.9: Best-performing model and preictal window performance metrics for each patient in G2. In red, are the patients with F1 score greater than 60%.

4 | Discussion

Pattern recognition in Brain Networks to characterize preictal states plays a crucial role in the seizure prediction field. In the medical field, missing the detection of the preictal state, which is crucial for anticipating a potential epileptic seizure, could have significant and severe consequences. Predicting the occurrence of epileptic seizures by means of electroencephalographic activity (EEG) could establish new therapeutic strategies for those patients with insufficiently controlled seizures improving the patient's quality of life. The present study proposed a novel ML pipeline that aims to classify interictal and preictal states across different preictal time windows. The most relevant Brain Network Features that contributed to the identification of patterns, and so a better classification of the epileptic states were identified, such as <ec> for γH band and <C> for δ band were the most significant features extracted with the feature selection pipeline proposed. As reviewed in the literature, the γ band oscillations may be associated with the abnormal synchronization of neuronal networks, which is a key feature for seizure onset identification. The following section discusses the key findings derived from the present investigation for each of the proposed groups, G1 and G2.

Figure 2.1 shows the ML pipeline designed in the present study. It is composed of 5 main blocks, Dataset Selection, Signal Preprocessing, Functional Connectivity Analysis, Machine Learning Classification, and a final Statistical Validation. The detailed pipeline proposed is presented in Appendix A.1.

Regarding the Functional Connectivity Analysis performed over G1 and G2, as shown in Figures 3.1a and 3.1b, the dynamic behavior of the Brain Network features analyzed was very different between patients, which confirmed the need to perform a patient-specific model. For example, while $< ec >$ for γH increased its magnitude value on P_5 when approaching the seizure, for P_3, the same feature maintained a constant low value when approaching the SOZ activation time.

Additionally, for both patients, $< C >$ and $< S >$ for γH seem to have the same behavior as the seizure evolves in time. The linkage between these features in a brain network relates to the network's overall organization and functional segregation/integration. In the case

of P_5 a high $< C >$ and $< S >$ value, before the seizure, reflected a balance between specialized local processing (i.e., high clustering) and good global communication (i.e., high node strength). While *sp* and $< B >$ features behaved in an opposite way before the seizure in both patients.

By examining the feature significance (see Figure 3.3, it was encountered that for G1 and G2, the most relevant Network characteristics were $< ec >$ and σ_{ec} forγH band. Furthermore, the δ band showed a high relevance for epileptic states differentiation with the features $< C >$ and σ_C. <B> was also part of the Top 5 significant features selected.

Eigenvector Centrality (<ec>) represents the level of existing correlation between nodes in a Brain Network. Lohmann et al [77], demonstrated the computational efficiency of this Brain Network feature as it has a great capability to capture intrinsic neural architecture on a node level. Conversely, Rubinov et al [36], described the Clustering Coefficient as another relevant feature to characterize the behavior of brain networks. In fact, this property is able to reflect, on average, the prevalence of clustered connectivity around the individual nodes in the network. The important information given by these two features was fundamental to distinguish the variations in the behavior of interictal and preictal states.

In total, five ML and DL models were evaluated on each patient for 3 different preictal time windows (40, 60, and 80 minutes), as described in section 2.6. The ML pipeline designed in the present study proposes an analysis based on two group configurations due to the limitation in the number of seizure instances available for some patients, G1 and G2. The first one contains all patients with at least 3 EEG seizure recordings, allowing only the evaluation of the models with a train validation set. The latest group contains patients with more than 3 EEG seizure recordings, allowing the configuration of a test set to further check the generalization capabilities of the models. Tables 3.4 and 3.9 present the results obtained for G1 and G2, respectively.

In G1, P_5, obtained the best F1 score, 75.54%, among all patients with the RF model and a preictal window of 80 minutes, which outstands the models' ability to handle imbalanced data and capture complex relationships and patterns between features. A precision of greater than 50% was obtained, indicating that instances predicted as TP (preictal) were correctly identified by the model. In contrast with some studies, a specificity of 82.58% was accomplished, which is only 15% less than the one obtained in [65].

Moreover, the 80-minute preictal window was the most selected optimal one which allowed the best-performance of the selected model for most patients. Nevertheless, statistical tests should be applied to evaluate if the mean performance across patients for the same

window was significant. This could mean that due to the lack of information available, a bigger window is needed in order to capture more information that helps to distinguish interictal and preictal states and characterize them in a correct manner the preictal state. XGB was the most best-performing model selected for the patients in G1, with a recurrence of 43%, this model is well-suited for capturing non-linear relationships in data. In fact, P_4 obtained an F1 score only 5% lower than P_5, the model accurately identified preictal class in 45% of cases. Broadly speaking, 5 patients out of 7 in G1 obtained an F1 score over 60%. It is certain that one limitation for G1 is the lack of at least an additional seizure to perform a better assessment of the model's generalization capacity.

For G2, the best-performing model and preictal window were selected according to the maximum F1 score obtained in the validation step. Then, the selected model was evaluated in each patient-specific Test set. The main purpose of implementing a test set is to evaluate the extent to which the LSTM-trained model can apply its learned knowledge for the characterization of preictal states to unfamiliar seizure data. P_{10} and P_{11} also had as best model LSTM, the first one obtaining an F1 score of 70% and the latest a score 5% lower than P_3. The highest precision among patients, 85.71%, was achieved by P_{10}.

Likewise, Bongiorni et al [27] observed that an LSTM model was able to classify ictal regions with up to 61% sensitivity and 99% specificity, with analysis window between 60s and 240s, which confirmed the capacity of RNNs of LSTM type to assimilate temporal patterns in EEG data. P_{10} achieved a recall of 30% and specificity of 99.55%, while P_{11} achieved a recall only 15% lower than the proposed study and specificity of 82.66%.

SVC was selected in 40% of the cases as the best-performing model among the patients, all of them with a preictal analysis window of 80 minutes. An F1 score over 80% was obtained by P_{13} and a recall of 65%. P_2 in contrast, achieved an F1 score of 70% and a recall greater than 70%. Compared with the literature, Shiao et al. [28], SVM-based seizure forecasting model achieved an average sensitivity of 95%, with a window analysis ranging from 20s to 1 hour. The study suggested a minimum amount of seizures, between 5 and 7, in the training data to guarantee a good performance classifying interictal and preictal states.

Finally, the RF model was selected by 20% of the cases and the CNN model only by 1 out of 10 patients. Another outcome in G2 was the fact that for 50% of the cases, the 80-minute preictal window was selected as it allowed the best model's performance.

The diverse distribution of seizure frequencies within the patients contributed to an uneven distribution of seizure instances for each dataset. This inequality means that models trained with a higher amount of seizure instances might exhibit better predictive patterns

as more information is available for the training set compared to those trained on subjects with fewer seizure instances. Thus, the model's generalization capability to predict *preictal* patterns might be biased due to the data imbalance.

5 | Conclusions and Future Developments

In the present work, a novel ML pipeline has been designed and evaluated to classify interictal and preictal states in epileptic patients across three different preictal time windows, 40, 60, and 80 minutes. The approach implements a Functional Connectivity Analysis using Brain Networks to characterize preictal states. A workflow comprising 5 main blocks, **i)**Dataset Selection, **ii)**Signal Preprocessing, **iii)**Functional Connectivity Analysis, **iv)**Machine Learning Classification, and a final **v)**Statistical Validation, was proposed in order to achieve the intended goal. In total, five ML and DL models were evaluated, RF, SVC, XGB, CNN, and LSTM.

Due to the brain's dynamicity and so EEG signal's variability, the activity that occurs during seizures can manifest differently for each patient. Identifying patterns in brain networks to characterize preictal states in epileptic patients is a very challenging task because of the different behaviors encountered between the networks. Consequently, *patient-specific* models are highly suggested to classify interictal and preictal states.

The challenges encountered in the development of the pipeline included limitations such as individual variability, clinical heterogeneity, data imbalance, and sensitivity and specificity trade-offs. Firstly, as seizures manifest in a unique manner for each patient according to their state, knowing the current patient condition when a seizure manifests is helpful when analyzing brain network behavior. Conversely, assuring homogeneity in a controlled environment will give the same conditions for the patients and could lead to better results.

Another constraint was Data Imbalance, 7/17 patients only had 3 EEG seizure recordings available which implied only having a train and validation set, eliminating the possibility to further evaluate the model's performance on unseen data. This also affects the training process as only a few seizure instances are available. It is suggested a minimum of 3-5 seizures as training data. Also, the reduced amount of preictal samples leads to an imbalanced dataset, larger preictal windows could be analyzed, under the constraint that more information previous to the seizure must be available.

Lastly, sensitive and specificity trade-offs affected the analysis especially when the test set was evaluated, maintaining a balance between these two metrics allows the control between TP and FN instances. Custom thresholds to perform the post-hoc model evaluation could lead to a better evaluation according to the goal.

A total of 17 patients were evaluated with the proposed pipeline. Due to the imbalance in the seizure instances of the dataset used in the present problem, the patients were analyzed in two main groups, G1 and G2. G1 contained all patients with at least 3 seizures. In 57% of the patients, the F1 score was over 65%, with one of them reaching 75.54%. In 6/7 patients, the optimal preictal window was the 80-minute window. The use of a larger preictal window allows a more comprehensive understanding of the brain network dynamics prior to a seizure event.

In contrast, G2 contained the patients with more than 3 seizures. In 6/10 patients, an F1 score surpasses 60% in the Test set, with one of them reaching up to 80%. Results comparable with those found in the literature. Due to the imbalance in the number of seizures present for the train, validation, and test set, the preictal window selected was highly variable across patients. Nevertheless, 50% of the patients selected the longest preictal window, an 80-minute window, as the optimal one.

Another relevant outcome regards the feature selection process. The most frequent features, in G1 and G2, were $< ec >$ and σ_{ec} for γH band, which thanks to its association with multiple cognitive functions any change or disruption in neural activity over this band could precede a seizure onset. $< C >$ and σ_C network characteristics for the δ band presented a high significance level to differentiate interictal and preictal pattern distributions as a result of the P_rms increase in the preictal period, according to literature.

The current analysis put in evidence how the variability in the brain network dynamics, not only across patients but also between seizure instances of the same patient, biased the selection of one model over another. Additionally, more than 60% of the patients in total achieved the highest performance with an 80-minute preictal window.

Future work could be directed towards selecting the best model one among the 5 proposed and test the same for all patients, further analysis evaluating the mean and standard deviation of the metrics achieved by every patient across each of the proposed windows should be carried out. Furthermore, special attention should be directed towards the improvement of the imbalance of the classes, using other ensemble methods or implementing techniques such as Adaptative Synthetic Sampling (ADASYN), to reduce the false positive predictions and improve the precision.

Predicting the occurrence of epileptic seizures by means of electroencephalographic activity (EEG) and the application of Functional Analysis based on Brain Networks could establish new therapeutic strategies for those patients with insufficiently controlled seizures improving the patient's quality of life.

Bibliography

[1] K. L. Skjei and D. J. Dlugos, "The evaluation of treatment-resistant epilepsy," *Seminars in Pediatric Neurology*, vol. 18, pp. 150–170, 9 2011.

[2] A. Sharmila and P. Geethanjali, "A review on the pattern detection methods for epilepsy seizure detection from eeg signals," *Biomedical Engineering / Biomedizinische Technik*, vol. 64, no. 5, pp. 507–517, 2019. [Online]. Available: https://doi.org/10.1515/bmt-2017-0233

[3] W. H. O. R. O. for the Eastern Mediterranean, "Epilepsy and seizures," World Health Organization. Regional Office for the Eastern Mediterranean, Technical documents, 2019-11.

[4] M. S. Farooq, A. Zulfiqar, and S. Riaz, "Epileptic seizure detection using machine learning: Taxonomy, opportunities, and challenges," *Diagnostics*, vol. 13, no. 6, 2023. [Online]. Available: https://www.mdpi.com/2075-4418/13/6/1058

[5] S. Sanei and J. Chambers, *Fundamentals of EEG Signal Processing*. John Wiley 'I&' Sons, Ltd, 2007. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119386957.ch4

[6] T. Alotaiby, F. E. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for eeg signal processing," *Eurasip Journal on Advances in Signal Processing*, vol. 2015, pp. 1–21, 12 2015. [Online]. Available: https://asp-eurasipjournals.springeropen.com/articles/10.1186/s13634-015-0251-9

[7] Satyender, S. K. Dhull, and K. K. Singh, "A review on automatic epilepsy detection from eeg signals," *Springer*, pp. 1441–1454, 2021.

[8] Y. Paul, "Various epileptic seizure detection techniques using biomedical signals: a review," *Brain Informatics*, vol. 5, pp. 1–19, 12 2018. [Online]. Available: https://braininformatics.springeropen.com/articles/10.1186/s40708-018-0084-z

[9] M. Le Van Quyen, J. Soss, V. Navarro, R. Robertson, M. Chavez, M. Baulac, and J. Martinerie, "Preictal state identification by synchronization changes in

long-term intracranial eeg recordings," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 559–568, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245704004626

[10] E. J. Ngamga, S. Bialonski, N. Marwan, J. Kurths, C. Geier, and K. Lehnertz, "Evaluation of selected recurrence measures in discriminating pre-ictal and inter-ictal periods from epileptic eeg data," *Physics Letters A*, vol. 380, no. 16, pp. 1419–1425, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0375960116001614

[11] J. Martinerie, C. Adam, M. L. V. Quyen, M. Baulac, S. Clemenceau, B. Renault, and F. J. Varela, "Epileptic seizures can be anticipated by non-linear analysis," *Nature Medicine 1998 4:10*, vol. 4, pp. 1173–1176, 10 1998. [Online]. Available: https://www.nature.com/articles/nm1098_1173

[12] M. Le Van Quyen, J. Martinerie, V. Navarro, P. Boon, M. D'Havé, C. Adam, B. Renault, F. Varela, and M. Baulac, "Anticipation of epileptic seizures from standard eeg recordings," *The Lancet*, vol. 357, no. 9251, pp. 183–188, 2001. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0140673600035911

[13] F. Li, Y. Liang, L. Zhang, C. Yi, Y. Liao, Y. Jiang, Y. Si, Y. Zhang, D. Yao, L. Yu, and P. Xu, "Transition of brain networks from an interictal to a preictal state preceding a seizure revealed by scalp eeg network analysis," *Cognitive Neurodynamics*, vol. 13, 04 2019.

[14] J. Gotman and P. Gloor, "Automatic recognition and quantification of interictal epileptic activity in the human scalp eeg," *Electroencephalography and Clinical Neurophysiology*, vol. 41, no. 5, pp. 513–529, 1976. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0013469476900638

[15] J. Gotman, "Automatic recognition of epileptic seizures in the eeg," *Electroencephalography and Clinical Neurophysiology*, vol. 54, no. 5, pp. 530–540, 1982. [Online]. Available: https://www.sciencedirect.com/science/article/pii/0013469482900384

[16] A. M. Murro, D. W. King, J. R. Smith, B. B. Gallagher, H. F. Flanigin, and K. Meador, "Computerized seizure detection of complex partial seizures," *Electroencephalography and Clinical Neurophysiology*, vol. 79, no. 4, pp. 330–333, 1991. [Online]. Available: https://www.sciencedirect.com/science/article/pii/001346949190128Q

[17] I. Osorio, M. G. Frei, and S. B. Wilkinson, "Real-time automated detection

and quantitative analysis of seizures and short-term prediction of clinical onset," *Epilepsia*, vol. 39, no. 6, pp. 615–627, 1998. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1528-1157.1998.tb01430.x

[18] S. Grewal and J. Gotman, "An automatic warning system for epileptic seizures recorded on intracerebral eegs," *Clinical Neurophysiology*, vol. 116, no. 10, pp. 2460–2472, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245705002671

[19] M. Bandarabadi, C. A. Teixeira, J. Rasekhi, and A. Dourado, "Epileptic seizure prediction using relative spectral power features," *Clinical Neurophysiology*, vol. 126, no. 2, pp. 237–248, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245714002971

[20] H. Dickten, S. Porz, C. E. Elger, and K. Lehnertz, "Weighted and directed interactions in evolving large-scale epileptic brain networks," *Scientific Reports*, vol. 6, no. 1, p. 34824, Oct 2016. [Online]. Available: https://doi.org/10.1038/srep34824

[21] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: Structure and dynamics," *Physics Reports*, vol. 424, no. 4, pp. 175–308, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S037015730500462X

[22] V. L. Morgan, B. Abou-Khalil, and B. P. Rogers, "Evolution of functional connectivity of brain networks and their dynamic interaction in temporal lobe epilepsy," *Brain Connectivity*, vol. 5, no. 1, pp. 35–44, 2015, pMID: 24901036. [Online]. Available: https://doi.org/10.1089/brain.2014.0251

[23] R. Fruengel, T. Bröhl, T. Rings, and K. Lehnertz, "Reconfiguration of human evolving large-scale epileptic brain networks prior to seizures: an evaluation with node centralities," *Scientific Reports*, vol. 10, no. 1, p. 21921, Dec 2020. [Online]. Available: https://doi.org/10.1038/s41598-020-78899-7

[24] T. Rings, R. von Wrede, and K. Lehnertz, "Precursors of seizures due to specific spatial-temporal modifications of evolving large-scale epileptic brain networks," *Scientific Reports*, vol. 9, no. 1, p. 10623, Jul 2019. [Online]. Available: https://doi.org/10.1038/s41598-019-47092-w

[25] N.-F. Chang, T.-C. Chen, C.-Y. Chiang, and L.-G. Chen, "Channel selection for epilepsy seizure prediction method based on machine learning," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 5162–5165.

[26] V. Pezoulas, kostas Tsiouris, M. Zervakis, S. Konitsiotis, D. Koutsouris, and D. Fotiadis, "A long short-term memory deep learning network for the prediction of epileptic seizures using eeg signals," *Computers in Biology and Medicine*, vol. 99, pp. 24–37, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S001048251830132X

[27] L. Bongiorni and A. Balbinot, "Evaluation of recurrent neural networks as epileptic seizure predictor," *Array*, vol. 8, p. 100038, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590005620300230

[28] H.-T. Shiao, V. Cherkassky, J. Lee, B. Veber, E. E. Patterson, B. H. Brinkmann, and G. A. Worrell, "Svm-based system for prediction of epileptic seizures from ieeg signal," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 5, pp. 1011–1022, 2017.

[29] T. Blauwblomme, P. Jiruska, and G. Huberfeld, "Chapter seven - mechanisms of ictogenesis," in *Modern Concepts of Focal Epileptic Networks*, ser. International Review of Neurobiology, P. Jiruska, M. de Curtis, and J. G. Jefferys, Eds. Academic Press, 2014, vol. 114, pp. 155–185. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780124186934000078

[30] M. A. Kramer and S. S. Cash, "Epilepsy as a disorder of cortical network organization," *The Neuroscientist*, vol. 18, no. 4, pp. 360–372, 2012, pMID: 22235060. [Online]. Available: https://doi.org/10.1177/1073858411422754

[31] L. Kuhlmann, K. Lehnertz, M. P. Richardson, B. Schelter, and H. P. Zaveri, "Seizure prediction - ready for a new era," *Nat Rev Neurol*, vol. 14, no. 10, pp. 618–630, Oct. 2018.

[32] O. Sporns, "Graph theory methods: applications in brain networks," *Dialogues in Clinical Neuroscience*, vol. 20, no. 2, pp. 111–121, 2018, pMID: 30250388. [Online]. Available: https://doi.org/10.31887/DCNS.2018.20.2/osporns

[33] S. L. Bressler and V. Menon, "Large-scale brain networks in cognition: emerging methods and principles," *Trends in Cognitive Sciences*, vol. 14, no. 6, pp. 277–290, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364661310000896

[34] O. Sporns, D. R. Chialvo, M. Kaiser, and C. C. Hilgetag, "Organization, development and function of complex brain networks," *Trends in Cognitive Sciences*, vol. 8, no. 9, pp. 418–425, 2004. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1364661304001901

[35] O. Sporns, "Structure and function of complex brain networks," *Dialogues in clinical neuroscience*, vol. 15, pp. 247–62, 09 2013.

[36] M. Rubinov and O. Sporns, "Complex network measures of brain connectivity: Uses and interpretations," *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, 2010, computational Models of the Brain. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S105381190901074X

[37] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems (nature reviews neuroscience (2009) 10, (186-198))," *Nature Rew. Neurosc.*, vol. 10, pp. 186–198, 03 2009.

[38] M. P. van den Heuvel and O. Sporns, "Rich-club organization of the human connectome," *Journal of Neuroscience*, vol. 31, no. 44, pp. 15 775–15 786, 2011. [Online]. Available: https://www.jneurosci.org/content/31/44/15775

[39] F. Bartolomei, I. Bosma, M. Klein, J. C. Baayen, J. C. Reijneveld, T. J. Postma, J. J. Heimans, B. W. van Dijk, J. C. de Munck, A. de Jongh, K. S. Cover, and C. J. Stam, "Disturbed functional connectivity in brain tumour patients: Evaluation by graph analysis of synchronization matrices," *Clinical Neurophysiology*, vol. 117, no. 9, pp. 2039–2049, 2006. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245706002240

[40] A. M. Bastos and J.-M. Schoffelen, "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls," *Frontiers in Systems Neuroscience*, vol. 9, 2016. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnsys.2015.00175

[41] B. Chai, D. Walther, D. Beck, and L. Fei-fei, "Exploring functional connectivities of the human brain using multivariate information analysis," in *Advances in Neural Information Processing Systems*, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds., vol. 22. Curran Associates, Inc., 2009. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2009/file/8248a99e81e752cb9b41da3fc43fbe7f-Paper.pdf

[42] P. Babaeeghazvini, L. M. Rueda-Delgado, J. Gooijers, S. P. Swinnen, and A. Daffertshofer, "Brain structural and functional connectivity: A review of combined works of diffusion magnetic resonance imaging and electroencephalography," *Frontiers in Human Neuroscience*, vol. 15, 2021. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2021.721206

[43] Z. Šverko, M. Vrankić, S. Vlahinić, and P. Rogelj, "Complex pearson correlation coefficient for EEG connectivity analysis," *Sensors (Basel)*, vol. 22, no. 4, Feb. 2022.

[44] F. De Vico Fallani, J. Richiardi, M. Chavez, and S. Achard, "Graph analysis of functional brain networks: Practical issues in translational neuroscience," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 369, 10 2014.

[45] S. Spencer, "Neural networks in human epilepsy: Evidence of and implications for treatment," *Epilepsia*, vol. 43, no. 3, pp. 219–227, 2002. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1528-1157.2002.26901.x

[46] J. Engel, Jr, P. M. Thompson, J. M. Stern, R. J. Staba, A. Bragin, and I. Mody, "Connectomics and epilepsy," *Curr Opin Neurol*, vol. 26, no. 2, pp. 186–194, Apr. 2013.

[47] M. Centeno and D. W. Carmichael, "Network connectivity in epilepsy: Resting state fmri and eeg–fmri contributions," *Frontiers in Neurology*, vol. 5, 2014. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fneur.2014.00093

[48] E. van Diessen, S. J. H. Diederen, K. P. J. Braun, F. E. Jansen, and C. J. Stam, "Functional and structural brain networks in epilepsy: What have we learned?" *Epilepsia*, vol. 54, no. 11, pp. 1855–1865, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.12350

[49] E. J. Pegg, J. R. Taylor, P. Laiou, M. Richardson, and R. Mohanraj, "Interictal electroencephalographic functional network topology in drug-resistant and well-controlled idiopathic generalized epilepsy," *Epilepsia*, vol. 62, no. 2, pp. 492–503, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.16811

[50] G. Slinger, W. M. Otte, K. P. Braun, and E. van Diessen, "An updated systematic review and meta-analysis of brain network organization in focal epilepsy: Looking back and forth," *Neuroscience and Biobehavioral Reviews*, vol. 132, pp. 211–223, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0149763421005212

[51] W. Liao, Z. Zhang, Z. Pan, D. Mantini, J. Ding, X. Duan, C. Luo, G. Lu, and H. Chen, "Altered functional connectivity and small-world in mesial temporal lobe epilepsy," *PLOS ONE*, vol. 5, no. 1, pp. 1–11, 01 2010. [Online]. Available: https://doi.org/10.1371/journal.pone.0008525

[52] L. Bonilha, T. Nesland, G. U. Martz, J. E. Joseph, M. V. Spampinato, J. C.

Edwards, and A. Tabesh, "Medial temporal lobe epilepsy is associated with neuronal fibre loss and paradoxical increase in structural connectivity of limbic structures," *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 83, no. 9, pp. 903–909, 2012. [Online]. Available: https://jnnp.bmj.com/content/83/9/903

[53] E. van Diessen, W. J. E. M. Zweiphenning, F. E. Jansen, C. J. Stam, K. P. J. Braun, and W. M. Otte, "Brain network organization in focal epilepsy: A systematic review and meta-analysis," *PLOS ONE*, vol. 9, no. 12, pp. 1–21, 12 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0114606

[54] F. Pittau, C. Grova, F. Moeller, F. Dubeau, and J. Gotman, "Patterns of altered functional connectivity in mesial temporal lobe epilepsy," *Epilepsia*, vol. 53, no. 6, pp. 1013–1023, 2012. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1528-1167.2012.03464.x

[55] V. L. Morgan, B. Abou-Khalil, and B. P. Rogers, "Evolution of functional connectivity of brain networks and their dynamic interaction in temporal lobe epilepsy," *Brain Connect*, vol. 5, no. 1, pp. 35–44, Jul. 2014.

[56] W. Löscher, H. Potschka, S. M. Sisodiya, and A. Vezzani, "Drug resistance in epilepsy: Clinical impact, potential mechanisms, and new innovative treatment options," *Pharmacol Rev*, vol. 72, no. 3, pp. 606–638, Jul. 2020.

[57] F. Mormann, R. Andrzejak, C. Elger, and K. Lehnertz, "Seizure prediction: The long and winding road," *Brain : a journal of neurology*, vol. 130, pp. 314–33, 03 2007.

[58] S. Li, W. Zhou, Q. Yuan, and Y. Liu, "Seizure prediction using spike rate of intracranial eeg," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 21, no. 6, pp. 880–886, 2013.

[59] M. Le Van Quyen, J. Martinerie, V. Navarro, P. Boon, M. D'Havé, C. Adam, B. Renault, F. Varela, and M. Baulac, "Anticipation of epileptic seizures from standard eeg recordings," *Lancet*, vol. 357, pp. 183–8, 02 2001.

[60] M. Le Van Quyen, J. Soss, V. Navarro, R. Robertson, M. Chavez, M. Baulac, and J. Martinerie, "Preictal state identification by synchronization changes in long-term intracranial eeg recordings," *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 116, pp. 559–68, 04 2005.

[61] M. Mursalin, Y. Zhang, Y. Chen, and N. V. Chawla, "Automated epileptic seizure detection using improved correlation-based feature selection with random

forest classifier," *Neurocomputing*, vol. 241, pp. 204–214, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0925231217303442

[62] U. R. Acharya, Y. Hagiwara, and H. Adeli, "Automated seizure prediction," *Epilepsy and Behavior*, vol. 88, pp. 251–261, 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1525505018305791

[63] F. Song, Z. Guo, and D. Mei, "Feature selection using principal component analysis," in *2010 International Conference on System Science, Engineering Design and Manufacturing Informatization*, vol. 1, 2010, pp. 27–30.

[64] S. Wang, W. A. Chaovalitwongse, and S. Wong, "Online seizure prediction using an adaptive learning approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 12, pp. 2854–2866, 2013.

[65] M. Mursalin, Y. Zhang, Y. Chen, and N. Chawla, "Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier," *Neurocomputing*, vol. 241, 02 2017.

[66] X. Wang, G. Gong, N. Li, and S. Qiu, "Detection analysis of epileptic eeg using a novel random forest model combined with grid search optimization," *Frontiers in Human Neuroscience*, vol. 13, 2019. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fnhum.2019.00052

[67] P. Vanabelle, P. De Handschutter, R. Tahry, M. Benjelloun, and M. Boukhebouze, "Epileptic seizure detection using eeg signals and extreme gradient boosting," *Epileptic seizure detection using EEG signals and extreme gradient boosting. Journal of biomedical research*, vol. 34(3), p. 228–239., 04 2019.

[68] A. Gupta, P. Singh, and M. Karlekar, "A novel signal modeling approach for classification of seizure and seizure-free eeg signals," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 5, pp. 925–935, 2018.

[69] H. Khan, L. Marcuse, M. Fields, K. Swann, and B. Yener, "Focal onset seizure prediction using convolutional networks," *IEEE Transactions on Biomedical Engineering*, vol. PP, pp. 1–1, 12 2017.

[70] Y. Gao, B. Gao, Q. Chen, J. Liu, and Y. Zhang, "Deep convolutional neural network-based epileptic electroencephalogram (eeg) signal classification," *Frontiers in Neurology*, vol. 11, 2020. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fneur.2020.00375

[71] M. Ihle, H. Feldwisch-Drentrup, C. A. Teixeira, A. Witon, B. Schelter, J. Timmer,

and A. Schulze-Bonhage, "Epilepsiae – a european epilepsy database," *Computer Methods and Programs in Biomedicine*, vol. 106, no. 3, pp. 127–138, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169260710002221

[72] M. A. C. Garcia, "Evolution of network organization to characterize pre-ictal states," Master's thesis, TPontificia Universidad Javeriana, 6 2021.

[73] W. Kirch, Ed., *Pearson's Correlation Coefficient*. Dordrecht: Springer Netherlands, 2008, pp. 1090–1091. [Online]. Available: https://doi.org/10.1007/978-1-4020-5614-7_2569

[74] N. Masuda, M. Sakaki, T. Ezaki, and T. Watanabe, "Clustering coefficients for correlation networks," *Frontiers in Neuroinformatics*, vol. 12, 2018. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fninf.2018.00007

[75] E. W. Dijkstra, "A note on two problems in connection with graphs," *Numerische Mathematik*, vol. 1, no. 1, pp. 269–271, Dec 1959. [Online]. Available: https://doi.org/10.1007/BF01386390

[76] S. Segarra and A. Ribeiro, "Stability and continuity of centrality measures in weighted graphs," *CoRR*, vol. abs/1410.5119, 2014. [Online]. Available: http://arxiv.org/abs/1410.5119

[77] G. Lohmann, D. Margulies, A. Horstmann, B. Pleger, J. Lepsien, D. Goldhahn, H. Schloegl, M. Stumvoll, A. Villringer, and R. Turner, "Eigenvector centrality mapping for analyzing connectivity patterns in fmri data of the human brain," *PloS one*, vol. 5, p. e10232, 04 2010.

[78] V. Aggarwal, V. Gupta, P. Singh, K. Sharma, and N. Sharma, "Detection of spatial outlier by using improved z-score test," in *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2019, pp. 788–790.

[79] L. Beretta and A. Santaniello, "Nearest neighbor imputation algorithms: A critical evaluation," *BMC Medical Informatics and Decision Making*, vol. 16, 07 2016.

[80] A. Mumuni and F. Mumuni, "Data augmentation: A comprehensive survey of modern approaches," *Array*, vol. 16, p. 100258, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590005622000911

[81] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: synthetic minority over-sampling technique," *Journal of artificial intelligence research*, vol. 16, pp. 321–357, 2002.

[82] N. Nachar, "The mann-whitney u: A test for assessing whether two independent samples come from the same distribution," *Tutorials in Quantitative Methods for Psychology*, vol. 4, 03 2008.

[83] C. Vercellis, "Business intelligence: Data mining and optimization for decision making," *Business Intelligence: Data Mining and Optimization for Decision Making*, 03 2009.

[84] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[85] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," *arXiv preprint arXiv:1511.08458*, 2015.

[86] A. Wibawa, A. B. Putra Utama, H. Elmunsyah, U. Pujianto, F. Dwiyanto, and L. Hernandez, "Time-series analysis with smoothed convolutional neural network," *Journal of Big Data*, vol. 9, p. 44, 04 2022.

[87] A. M. Abdelhameed, H. G. Daoud, and M. Bayoumi, "Deep convolutional bidirectional lstm recurrent neural network for epileptic seizure detection," in *2018 16th IEEE International New Circuits and Systems Conference (NEWCAS)*, 2018, pp. 139–143.

[88] R. C. Staudemeyer and E. R. Morris, "Understanding lstm–a tutorial into long short-term memory recurrent neural networks," *arXiv preprint arXiv:1909.09586*, 2019.

[89] S. Stević, *Geometric Mean*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 608–609. [Online]. Available: https://doi.org/10.1007/978-3-642-04898-2_644

[90] P. A. Flach, *ROC Analysis*. Boston, MA: Springer US, 2010, pp. 869–875. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_733

[91] A. A. Kuznetsova, *Statistical Precision-Recall Curves for Object Detection Algorithms Performance Measurement*. Cham: Springer International Publishing, 2021, pp. 335–348. [Online]. Available: https://doi.org/10.1007/978-3-030-66077-2_27

A | Appendix A

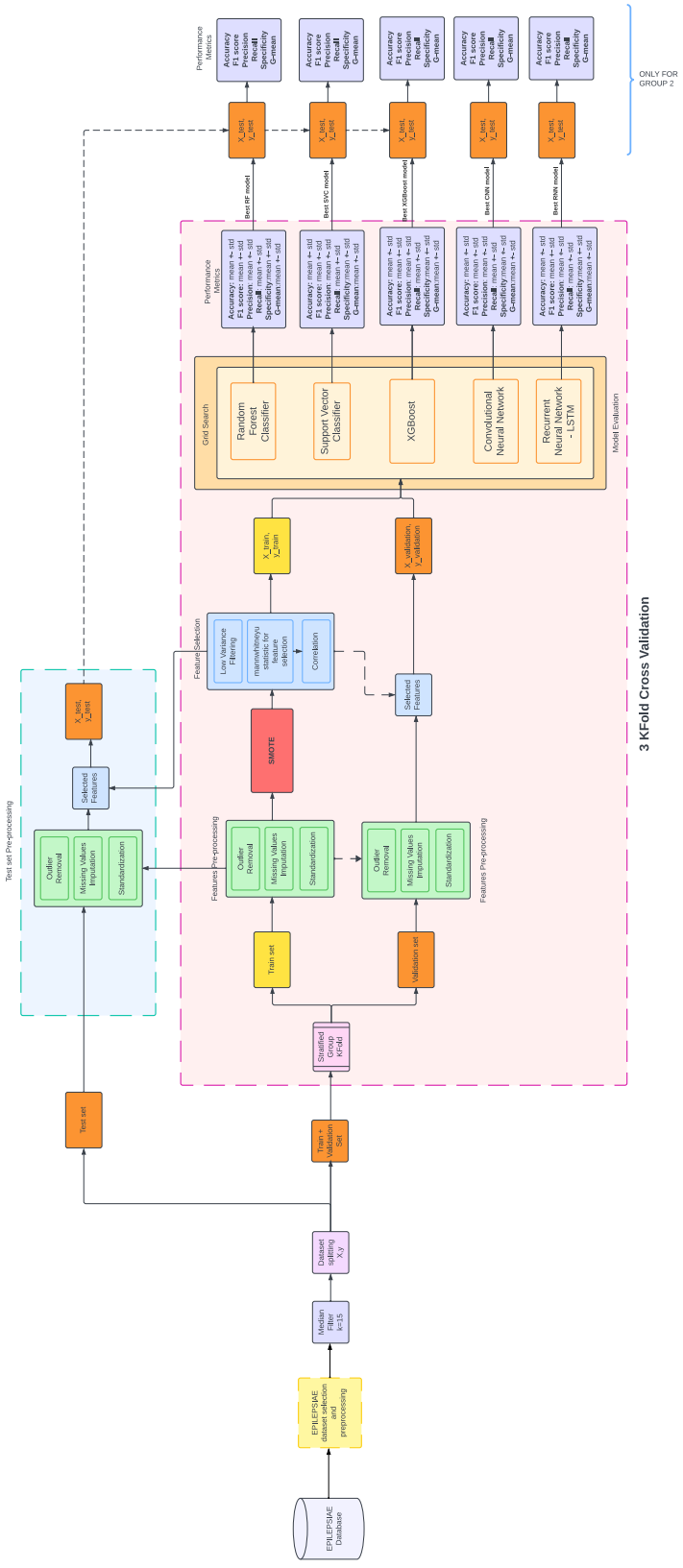


Figure A.1: Interictal and Preictal Brain Network Classification Pipeline

B | Appendix B

The following file contains all the metrics obtained for the two groups assessed within the present project: 2023_12_*Melgarejo_Quinones_Laura_Daniela_All_Patients_Data_3.pdf*

List of Figures

1.1 Measures of network topology. The left figure represents an undirected network. In the right figure, a directed network is shown. In green, are the shortest path lengths. Hub nodes, in black, lie on a high number of shortest paths. Taken from [31] 6
1.2 Structural brain network pipeline. Taken from [37] 7
1.3 Functional Brain Networks modeling and analysis pipeline. Taken from [44] 9
1.4 Seizure prediction system. **a)** Typical seizure prediction system scheme. **b)** Open-loop seizure prediction warning system **c)** Seizure prediction in practice. Taken from [31] 12

2.1 Preictal state characterization pipeline. 17
2.2 Number of seizures available per patient 19
2.3 PB1 Electrode behavior for P_3 after band filtering and recording concatenation. In red, the SOZ marker. 22
2.4 Connectivity Matrix for δ band for an interictal (*minute 50*) and preictal (*minute 450*) window for patient P_{16} 23
2.5 Local Patient Dataset Columns 26
2.6 Boxplot of the feature behavior before and after outlier removal for δ Band 27
2.7 Example of synthetic data points generation using SMOTE 29
2.8 SVM Classifier Hyperplane and Support Vectors. Kernel="rbf" 34
2.9 Activation functions used in the CNN implementation 36
2.10 CNN proposed structured for preictal state characterization 37
2.11 Standard LSTM memory block. Taken from [88] 38
2.12 LSTM-RNN proposed structured for preictal state characterization 39
2.13 Group splitting of the available patients 40
2.14 Dataset splitting for each group 41
2.15 Confusion Matrix for Interictal (0) and Preictal (1) classification periods . 42

3.1 Network feature evolution in time in γH band. 47

3.2 SMOTE and Undersampling sample behavior in Train set for patients in G1, with Majority (0) and Minority (1) Class. 48
3.3 Top 5 - Feature frequency selection in training set over the 3-GroupKFold for G1 and G2. 49
3.4 Confusion Matrices of RF best-performing model for P_5 across different preictal windows for the Validation set. 52
3.5 ROC-AUC and PR curve for the best preictal window, 80-minute, for the *Validation set* for all the evaluated models of P_5. 53
3.6 Confusion Matrices of LSTM best-performing model for P_3 across different preictal windows for the Test set. 56
3.7 Prediction vector evaluation of LSTM best-performing model for P_3 across different preictal windows for the Test set. 57

A.1 Interictal and Preictal Brain Network Classification Pipeline 78

List of Tables

2.1 Seizure patient information and Patient ID assigned for the present work . 20
2.2 Frequency bands and cutoff frequency range selected to characterize the EEG signal . . . 21
2.3 Features extracted for each band . . . 26
2.4 Random Forest Hyperparameter definition . . . 33
2.5 SVM Classifier Hyperparameter definition . . . 34
2.6 XGBoost Hyperparameter definiton . . . 35
2.7 CNN Hyperparameter definiton . . . 37
2.8 Long Short-Term Memory RNN's Hyperparameter definiton . . . 39
2.9 Classification metrics computed based on the Confusion Matrix . . . 43

3.1 Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_5 for a 40-minute preictal window. In red, the best values for each metric in the validation set. . . . 50
3.2 Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_5 for 60-minute preictal window. In red, the best values for each metric in the validation set. . . . 51
3.3 Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_5 for 60-minute preictal window. In red, the best values for each metric in the validation set. . . . 51
3.4 Best-performing model and preictal window performance metrics for each patient in G1. In red, are the patients with F1 score greater than 65%. . . . 53
3.5 Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_3 for 40-minute preictal window. In red, the best values for each metric in the validation set. . . . 54
3.6 Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_3 for 60-minute preictal window. In red, the best values for each metric in the validation set. . . . 54

3.7 Train and Validation set performance metrics to classify pre-ictal periods in different windows for P_3 for 80-minute preictal window. In red, the best values for each metric in the validation set. 55
3.8 Test set performance metrics LSTM model to classify pre-ictal states for 40, 60, and 80-minute preictal window for P_3. In red, are the best values for each metric in the test set. 56
3.9 Best-performing model and preictal window performance metrics for each patient in G2. In red, are the patients with F1 score greater than 60%. . . 58

Acknowledgements

First of all, I want to express my deepest gratitude to my advisor Prof. Luca Mainardi for entrusting me with this thesis project, his invaluable help and support in the making of this research were fundamental. The knowledge he always transmitted to students was very inspiring to me. I also want to thank my advisor Prof. Catalina Alvarado for her unwavering support and mentorship throughout my academic journey. I am immensely grateful to her for showing me the amazing world of signals and inspiring me to put hands-on with real-world problems, as it is Epilepsy, to help change lives. Thank you for letting me return to my *alma mater* and let me assist to a real Epilepsy surgery, it was a unique experience that I will never forget. Special thanks to my co-advisors Lupe and Andrea, for their assessment and support, especially for their patience. During the toughest moments, you always encouraged me to continue with my research even if sometimes I didn't like the outcome, thank you for always pushing me to go the extra mile.

To my Amplifon colleagues, thank you so much for joining me throughout this year and being witnesses to one of the greatest chapters of my life. As I was starting a big challenge, each one of you always gave me words of encouragement whenever I needed it, this truly helped me carry on. Thanks for all the laughs, the morning coffees, the Italian lectures, the adventures, and so much more. In particular, I want to thank Luca and Gio for believing in me since day one and inspiring me to always give my best. It was a pleasure working with all of you. Thank you for letting me be part of the "Dream team".

I want to give a special thanks to my friends, who turned out to be more than a family to me during these two years in Milano. Juli and Isa, I am more than sure that you were the best friends one could have ever asked for to share this amazing adventure, you guys were always beside me during some of the most difficult moments of my life and for that, I will always be grateful. Juli, my roomie, and classmate, thank you for being almost like a brother to me, for all the adventures, laughs, and fights, and for teaching me the real meaning of friendship by always being there and helping me to be a better person. Isa, you are a sunshine, a person with an amazing heart, and always ready to help, you always stood by me at my worst and your loyalty is a treasure beyond measure, and for that, I am more than grateful. Thank you to Sofi, Juliana, and Diana, you girls have a beautiful

heart from which I carry valuable lessons. Thank you to my international family Pedro, Felipe, and Julie, my favorite Brazilians, to Lana Maria, Liu, Ali, Jess, Gaia, Benja, and all the "kidz", you made every moment of this experience unforgettable, thank you for all the amazing memories, you guys taught me that friendship transcends borders.

To my best friends in Colombia, Anto, Aleja, and Arjo, despite the thousands of miles away, I know I can always count on each of you, and for that, I am more than grateful. Your unconditional support since the first moment I knew I would come to Italy to fulfill my dream was truly invaluable. You know I love you with all my heart. Special thanks to Peter, for being my best friend since day one of Uni, for all the laughs and funny stories, for making the best lemon pie, for his loyalty, and for being always there whenever I needed it.

Lastly, I want to thank my family, who are my biggest support in life and without whom I would not be the person I am today. I will never have enough words to thank you for everything you have done for me. Fernando (*Piquirris*), thank you for being an inspiration to me *amigo cuadrito*, for showing me the wonders of creating new things, for helping me understand that the world is huge and that the only limit is set by each one of us. Thank you for teaching me that the only difficult thing in life is what you don't understand and that's why we should always be in search of new solutions. Amparo (*Amiguis*), thank you for your unconditional love and support, for teaching me to dream big and never give up despite the challenges, for being the best mom and friend in the whole world, you always make me feel close to home even miles apart. Cami (*Sis*), your dedication in everything you do has been a true inspiration to me, your nobility with others and your simplicity is something I have always admired. Thank you for allowing me to be your role model and for listening to me when I need it, you are the best sister in the world. Finally, I extend my gratitude to my grandparents Victor, Flor, Hernando, and my *abuelita* Berenice, thank you for your unconditional love and support, and for helping me be who I am today.

I am filled with gratitude for the shared moments and the invaluable knowledge I acquired in the making of this project. I will carry with me all your inspiration as I embrace new horizons. Thank you for being part of this amazing adventure!