Computational Approaches to Epilepsy: Graph Networks for Localization and Multi-model Approach for Prediction

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Sai Sanjay Balaji

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Prof. Keshab K. Parhi, Advisor

December, 2025

Acknowledgements

My graduate school journey in a foreign country has been a marathon filled with hurdles, and crossing the finish line would have been impossible without the incredible support system that cheered me on at every stage.

I am profoundly grateful to my advisor, Prof. Keshab Parhi, for recognizing my potential, encouraging my transition to the PhD program, and for over five years of mentorship that guided me into a field I am now passionate about. This work was significantly refined by various committee members, Prof. Emad Ebbini, Prof. Alik Widge, Prof. Mos Kaveh, and Prof. Mehmet Akçakaya, and its real-world purpose was enhanced by the clinical insights of committee member and collaborator, Dr. Thomas Henry. I also extend a special thanks to Prof. Martina Cardone, whose teaching first sparked my interest toward a career in academia, which ultimately led me on this journey.

Within my lab, I owe a special thanks to Sandeep for his invaluable guidance; to my lab besties, Lulu and Sin-Wei, whose encouragement was my anchor to sanity; and to Wenrui, Joe, and Xingyi for their constant friendship. I am also grateful for the larger community of friends who made this graduate school journey a vibrant adventure, especially Ranjeet, Himanshu, Rithika, Manuraj, Yeena, Shashank, Yasir, Nandini, Faija, Evelyn, Aswin, and countless others. A special thanks to my "personal therapist", Dr. Taylor Swift, whose work provided a necessary emotional framework for navigating the life of a grad student.

Finally, to the people who have my whole heart: my parents, for their love, patience, and the hard work that raised me to be the man I am today; my feline son, Saffron, for providing the best stress relief and silent companionship; and my wonderful extended family, especially the Krishnas, for the unwavering encouragement.

Dedication

This dissertation is lovingly dedicated to the two leading ladies of my life - my mother, Girija Balaji, and my grandmother, Rajeswari Krishnamoorthy, who are more excited about me completing my Ph.D. than I can ever be.

Abstract

Epilepsy is one of the world's most common neurological diseases, affecting over 50 million people. Nearly a third of these individuals have drug-resistant epilepsy (DRE), a condition characterized by a lack of response to antiseizure medications. For these individuals, the primary remaining treatments are surgical intervention or neuromodulation. The success of these interventions depends mainly on solving two critical challenges: first, precisely localizing the seizure onset zone (SoZ) for resection or stimulation targets, and second, accurately predicting seizures for advisory or therapeutic devices. This dissertation addresses both challenges by developing and validating two distinct computational frameworks based on the retrospective analysis of intracranial EEG (iEEG) recordings.

First, this thesis establishes a generalizable, biomarker-driven framework for identifying the SoZ. It demonstrates that epileptic networks possess common, population-level signatures. By leveraging effective connectivity measures based on frequency-domain convergent cross-mapping (FD-CCM) and graph centrality, a robust model is developed. This approach is validated on challenging interictal recordings, demonstrating that integrating these novel graph features with established biomarkers in a hybrid classifier achieves state-of-the-art performance with 90% accuracy, providing a robust and generalizable tool for pre-surgical evaluation.

Second, the dissertation pivots to the challenge of seizure prediction, where it develops a patient-specific, multi-model framework to account for the high degree of intra-subject seizure heterogeneity. This approach uses unsupervised learning to cluster seizures based on their unique preictal signatures and trains an ensemble of specialized classifiers. The power and flexibility of this framework are demonstrated through two distinct feature engineering approaches. Its effectiveness was first validated using a comprehensive set of power spectral density (PSD) features, achieving a mean sensitivity of 98.54% and reducing the false positive rate by nearly 50% compared to the unclustered approach. In parallel, a novel biomarker, the Absolute Mean Instantaneous Frequency Difference (AMIFD), was introduced and also demonstrated strong predictive performance (92.08% sensitivity), confirming the framework's robustness.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 The Clinical Challenge of Epilepsy

Epilepsy stands as one of the world's most prevalent chronic neurological diseases, affecting over 50 million people globally. It is a condition characterized by an enduring predisposition of the brain to generate spontaneous, abnormal electrical events known as epileptic seizures. While a significant portion of this population can successfully manage their condition with a regimen of Anti-Epileptic Medications (AEDs), a devastatingly large subset cannot [2].

1.1.1 The Problem of Drug-Resistant Epilepsy (DRE)

Approximately one-third of all individuals with epilepsy suffer from Drug-Resistant Epilepsy (DRE), which is clinically defined by the failure of two or more appropriately chosen and tolerated AEDs to achieve sustained seizure freedom. For this population, numbering in the millions, the clinical, social, and personal burden is immense. The unpredictable, disruptive nature of seizures severely compromises quality of life, restricting fundamental activities such as driving, working, and social engagement. More critically, uncontrolled seizures are associated with significant comorbidities, including cognitive decline, psychiatric disorders, physical injury, and a tragically elevated risk of Sudden Unexpected Death in Epilepsy (SUDEP) [3, 4].

The failure of pharmacological treatment forces patients and clinicians to turn to more invasive, high-stakes interventions: namely, resective surgery and neuromodulation. The

success of these interventions, however, is not guaranteed. It hinges on our ability to solve two of the most complex computational and clinical challenges in epileptology.

1.1.2 Twin Challenges in DRE Management

For any given patient with DRE, the clinical pathway is determined by two fundamental questions: "Where in the brain is the seizure coming from?" and "When will the next seizure occur?" This dissertation develops advanced computational frameworks to address both of these critical questions, which are central to the two primary interventional strategies for DRE.

1.2 Intracranial EEG: The Gold Standard for Analysis

To address either the localization or prediction challenge, we require a signal that captures brain activity with the highest possible fidelity. While non-invasive scalp electroencephalography (EEG) is a crucial diagnostic tool, it suffers from a low signal-to-noise ratio and poor spatial resolution as the electrical signals are smeared and attenuated by the skull and scalp. For the high-stakes decisions involved in pre-surgical evaluation and the development of next-generation devices, the "gold standard" for data acquisition is intracranial EEG (iEEG) [5].

1.2.1 Data Acquisition and Properties

iEEG is an invasive procedure, typically performed on DRE patients being evaluated for surgery, which involves implanting electrodes directly in or on the brain. This is accomplished in one of two ways:

- Electrocorticography (ECoG): A craniotomy is performed to place grid or strip electrodes directly on the surface (cortex) of the brain.
- Stereoelectroencephalography (sEEG): Multiple depth electrodes are stereotactically inserted into deeper brain structures, allowing for the sampling of a three-dimensional volume.

While invasive, iEEG provides an unparalleled window into brain dynamics. The resulting signals have an extremely high spatiotemporal resolution (sampling activity from millimeter-scale neural populations at 1 kHz or higher) and a signal-to-noise ratio far superior to any non-invasive modality. It is this high-fidelity data, collected continuously over days or weeks, that forms the basis for all work in this dissertation.

1.2.2 Key Electrophysiological States

The analysis of long-term iEEG data is typically segmented by its temporal relationship to a seizure event. Understanding these distinct electrophysiological states is fundamental to this dissertation:

- Ictal: The period during which the seizure itself is occurring. It is visually and computationally distinct, often characterized by high-amplitude, hypersynchronous, and rhythmic discharges.
- Interictal: The "baseline" state of the brain. This refers to the long periods (hours, days, or even weeks) *between* seizure events.
- Preictal: The short, subtle, and transitional state that precedes a seizure. This period, which can last from seconds to hours, is theorized to contain the unique electrophysiological "fingerprint" or signature that signals an impending seizure.

A primary goal of modern computational epileptology is to identify robust, clinically-relevant biomarkers within the *interictal* and *preictal* states. Relying on the ictal state is insufficient; for localization, it can be misleading due to rapid seizure propagation, and for prediction, it is by definition too late.

1.3 Computational Gaps and Research Opportunities

Despite access to high-fidelity iEEG data, accurately localizing the SoZ and reliably predicting seizures remain largely unsolved problems. This is because traditional analysis methods are often insufficient to capture the profound complexity of the epileptic brain. The work in this dissertation is designed to address two specific, critical gaps in the literature.

1.3.1 Gap 1: Limitations in SoZ Localization

The current clinical gold standard for SoZ localization relies heavily on the visual inspection of iEEG recordings by expert epileptologists. This is an extraordinarily difficult, time-consuming, and subjective process that requires a clinician to manually review days or weeks of multi-channel data to identify the earliest electrographic changes of a seizure. This process is known to suffer from poor inter-rater reliability and lacks standardization.

To augment this visual analysis, researchers have proposed quantitative biomarkers. As noted in Paper 2 of this thesis, these include features such as the rate of interictal epileptiform discharges (IEDs), the prevalence of pathological high-frequency oscillations (HFOs), and changes in relative band power (RBP).

However, these biomarkers have significant limitations. HFOs, while a promising marker of pathology, are complex to detect reliably and automatically, and their precise relationship to the SoZ (as opposed to just irritated tissue) is still debated. More fundamentally, these biomarkers are all *univariate*. They analyze each electrode channel in isolation, ignoring the fact that epilepsy is now widely understood to be a *network disease*. The SoZ is not just an isolated "bad" electrode but rather a set of critical nodes within a larger, pathologically connected epileptogenic network. Methods that ignore the *interactions* between brain regions fail to capture the complex network dynamics that underlie seizure generation and propagation.

While some studies have begun to apply network analysis, they often rely on simple, linear, and non-directional measures of *functional* connectivity, such as cross-correlation. What is critically needed is a framework that models the *effective connectivity*—the directed, causal influence that one brain region exerts over another—and uses the graph-theoretic properties of this network to build a robust, *generalizable* biomarker for the SoZ. Furthermore, this biomarker must be effective on the more clinically convenient and readily available *interictal* data, not just on the ictal event itself.

1.3.2 Gap 2: Hurdles in Seizure Prediction

The challenge of seizure prediction is even more notorious. The field has a long history of studies reporting high accuracy, only to fail in rigorous validation or prospective trials. The reasons for this difficulty are manifold, but two stand out as the most significant hurdles,

which are directly addressed in Papers 3 and 4 of this thesis.

First is the problem of signal non-stationarity. The statistical properties of the iEEG signal (its mean, variance, and spectral content) are unstable; they drift over time due to patient state, medication, and sleep-wake cycles. A model trained on data from Monday may fail on data from Wednesday simply because the baseline brain activity has shifted.

Second, and arguably the single most significant gap in the literature, is the problem of intra-subject seizure variability. The vast majority of existing prediction algorithms are built on a "one-model-per-patient" assumption. This implicitly treats all of a patient's seizures as identical events, preceded by an identical preictal signature. Clinical observation and electrophysiological analysis prove this is false. A single patient may experience multiple distinct seizure *types*, with different clinical manifestations, propagation pathways, and, most importantly, different preictal electrical signatures.

A model trained to detect only one of these signatures will fail catastrophically when another seizure type occurs, resulting in a false negative. This failure to account for seizure *heterogeneity* is a primary reason why no prediction algorithm has successfully transitioned to widespread clinical use. The field does not need another feature set; it needs a new *framework* that abandons the single-model assumption and embraces a patient-specific, multi-model approach capable of automatically identifying and modeling the true diversity of a patient's seizures.

1.4 Thesis Statement and Objectives

This dissertation addresses these two fundamental challenges by developing, validating, and analyzing two distinct, high-performance computational frameworks tailored to the specific demands of localization and prediction.

The central thesis of this work is that (1) a generalizable, biomarker-driven framework built on novel graph-theoretic measures of effective connectivity can robustly identify the SoZ from intracranial iEEG, and (2) a patient-specific, multi-model framework is essential for reliable, long-term seizure prediction, as it can explicitly model the seizure heterogeneity present within a single individual.

To validate this thesis, the following objectives were set:

1. To investigate the capability of supervised machine learning models using graph centrality features, derived from model-free effective connectivity (EC) measures—specifically Directed Information (DI), Mutual Information-guided Granger Causality (MI-GCI), and Frequency-Domain Convergent Cross-Mapping (FD-CCM)—to identify the SoZ from peri-ictal iEEG recordings.

2. To develop a robust, generalizable SoZ localization framework that functions on more challenging interictal data by creating a hybrid multi-layer perceptron (MLP) model that integrates novel FD-CCM graph features with established biomarkers (IEDs, HFOs, and RBP).

3. To design and validate a patient-specific, multi-model seizure prediction framework that uses unsupervised feature selection (Minimum Uncertainty and Sample Elimination, or MUSE) to automatically cluster seizure-specific preictal patterns based on power spectral density (PSD) features, thereby accounting for intra-subject seizure heterogeneity.

4. To introduce a novel electrophysiological biomarker, the Absolute Mean Instantaneous Frequency Difference (AMIFD), and validate its effectiveness as a feature set within this multi-model prediction framework, demonstrating the framework's flexibility and robustness.

1.5 Dissertation Roadmap

This dissertation is structured in two main parts, corresponding to the twin challenges of localization and prediction, preceded by a review of the technical background and followed by a final synthesis.

- Chapter 2 provides a comprehensive review of the technical background and literature. This includes a deeper discussion of iEEG signal processing, the foundations of the effective connectivity measures and graph theory used, which form the basis of the subsequent chapters.

- Part I: Seizure Onset Zone Localization (Chapters 3 and 4) focuses on developing a generalizable model for SoZ identification.

  - Chapter 3 introduces the core methodology, demonstrating the capability of effective brain connectivity and graph centrality features to identify the SoZ in peri-ictal recordings, and identifies FD-CCM as a particularly potent measure.
  - Chapter 4 extends this work to develop a robust, generalizable model for identifying the SoZ from *interictal* data, thereby fulfilling Objective 2 by building and validating a hybrid model combining biomarkers.

- Part II: Patient-Specific Seizure Prediction (Chapters 5 and 6) pivots to the challenge of prediction, arguing for and developing a patient-specific, multi-model approach.
  - Chapter 5 introduces the patient-specific, multi-model classification framework. It validates this novel approach, fulfilling Objective 3 by using PSD features to cluster seizures and demonstrating significant improvements in sensitivity and false-positive rate over single-model baselines.
  - Chapter 6 presents the novel AMIFD biomarker. It fulfills Objective 4 by validating this feature's effectiveness within the multi-model framework, thereby establishing a flexible and powerful pipeline for personalized seizure prediction.

- Chapter 7 provides a final interpretation of these findings. It discusses the broader implications of the generalizable localization and patient-specific prediction frameworks, acknowledges the limitations of this body of work, and proposes exciting avenues for future research that build directly upon these contributions.

Chapter 2

Background and Literature Review

2.1 Fundamentals of iEEG Analysis and Brain Rhythms

Upon inventing the EEG to record the brain's electrical activities, Berger identified specific patterns of continuous recurring oscillations from the EEG recordings. He used the terms alpha and beta waves to describe these oscillatory *brain rhythms* in [6]. Since this discovery, researchers have documented numerous sustained oscillations across various mammalian brains, ranging from periods in minutes to frequencies over 600 Hz generated by different mechanisms. Furthermore, brain activity comprises multiple rhythms (frequencies) and varies over time [7]. Thus, time-frequency analyses of brain signals provide an appropriate measure of changes in frequency over time. Two standard methods for frequency decomposition are the short-term Fourier transform (STFT) and the wavelet transform [7]. STFT computes the Fourier transform over a time window that spans the time series. A signal is decomposed by wavelet transform onto a set of basis functions called wavelets. These are obtained from scaling and time shifts of a single template, called the mother wavelet [8].

Over time, scientists grouped these oscillations into different frequency bands. Although there is no exact limit for each band, the commonly categorized bands are delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), gamma (30-80 Hz), ripple (80-200 Hz), and fast ripple (200-600 Hz) [9]. Based on the sampling frequency, the ripple band can be band-limited to 80-250 Hz, and the fast ripple to 250-500 Hz. These frequency bands are shown in Table 2.1. iEEG oscillations above frequencies of 20 Hz are typically grouped

as fast oscillations (FOs). FOs include the beta-2 sub-band in 20 - 30 Hz, the gamma band, ripples, and fast ripples. High-frequency oscillations (HFOs), a subgroup of FOs, are characterized by at least four oscillations in the ripple or fast-ripple frequency bands (80 Hz and above). HFOs occur for a relatively short duration (10–100 ms) and have a higher amplitude (10–1000 μV) than background EEG/seizures [10].

Table 2.1: EEG frequency bands

Frequency Range	**Band Name**
1 – 4 Hz	Delta (δ)
4 – 8 Hz	Theta (θ)
8 – 13 Hz	Alpha (α)
13 – 30 Hz	Beta (β)
30 – 80 Hz	Gamma (γ)
80 – 250 Hz	Ripples
250 Hz and above	Fast ripples

2.2 Graph-based Brain Connectivity

One of the primary goals of cognitive neuroscience is to identify the neural interactions within the brain corresponding to a specific cognitive task. By comparing these interactions during seizures with a healthy baseline, we can identify the abnormalities. Researchers have used this approach to generate a brain map from iEEG data for the analysis of epileptic patients.

Brain connectivity, as described in [11], can be classified into three types: structural connectivity, functional connectivity, and effective connectivity. *Structural/ anatomical connectivity* refers to the physical interconnection of axons emerging from a brain region connecting to another. Any of the widely available non-invasive magnetic resonance imaging techniques can identify the structural connectivity. *Functional connectivity* is a measure of the correlation or the statistical dependency between the brain regions inferred from a

time-series signal. It helps group regions of the brain into an interactive network corresponding to a cognitive task. The Pearson's correlation coefficient is a commonly used measure of functional connectivity. *Effective connectivity* represents the *directional* data flow between the brain regions during a cognitive task. It identifies a causal brain network corresponding to the directed interactions among brain areas during a cognitive task [12]. There are other ways to describe and classify brain connectivity measures. For example, complex network models are used in [13] to define structural and functional connectivity employing small-world architectures, clustering, and quantifiable topological parameters, such as modularity, centrality, and hub distribution. In [14], functional connectivity is described as an observable phenomenon quantified with measures of statistical dependencies, and effective connectivity corresponds to the parameter of a model that tries to explain observed dependencies described by functional connectivity. Functional connectivity has been used to discriminate borderline personality disorder in [15].

The presence of hidden sources highly impacts the estimation of connectivity measures. For example, in functional connectivity estimation using Pearson's correlation, two independent processes can still be highly correlated. However, neither could be a causal factor for the other, and a hidden process could be driving both. Hidden factors can also lead to spurious connectivity estimates using Granger Causality measures if a common input node is not observed or the coupling strength varies between nodes [16, 17]. This, however, can be eliminated by observing all the nodes and using re-normalization. Other factors that affect the accurate estimation of connectivity measures are signal-to-noise ratio and unequal observation periods [16]. Thus, care must be taken to minimize the noise and avoid sampling bias.

Identifying a causal relation in a complex system such as the human brain can be complicated. Scientists have developed multiple methods over the years for estimating effective connectivity. However, it is often debated whether the effective connectivity measure is an actual representation of the causal interactions within the brain. This stems from the idea that a few hundred electrophysiological recordings from different regions may be insufficient to derive a brain's causal model consisting of billions of neurons, as the activities of the numerous unobserved neurons may likely have a more significant influence than the ones

recorded [18]. Perhaps effective connectivity can be better described as a directional measure inferred from the statistical dependence of the time-series of iEEG recordings across channels. Nonetheless, such effective connectivity measures have shown promising results in locating SoZ. For the context of this manuscript, any reference to "causal" networks refers to effective connectivity.

2.3 State-of-the-Art in SoZ Identification

The following sections are adapted from the author's review paper on SoZ identification [5] and provide a comprehensive overview of the field, from the definition of the epileptogenic zone to the various biomarkers that have been proposed to identify it.

2.3.1 Epilepsy and the Epileptogenic Zone

A seizure is a sudden unrestricted disruption of electrical activity across the neurons [19]. Usually caused by the local disturbances of the brain's electrical activity, seizures typically cause short-term abnormalities in muscular movements and consciousness. Although sometimes a known external reason, such as stroke or head injury, causes seizures, the cause remains unidentified in many cases. Physicians diagnose a person as epileptic when they encounter more than two unprovoked seizing episodes separated by at least one day. Epilepsy is one of the most diagnosed neurological disorders affecting around 65 million global citizens, with about 3.4 million of them in the United States [2]. Roughly half of them are diagnosed with a known cause. These can be genetic or caused by a structural change in the brain, a brain infection, or autoimmune disorder. Identification of the cause can help in curating a successful treatment strategy.

A group of features characterizes epileptic syndrome. Examples include seizure type, affected regions, cause, age of onset, severity, and attack frequency [2]. Seizures are broadly classified into three groups by the International League Against Epilepsy (ILAE) based on the onset, the patient's level of awareness during the seizure, and the presence/absence of motor symptoms exhibited by the patient [20]. The different groups are shown in Figure 2.1. Focal onset seizures, also called partial seizures, are localized, with the seizures starting in one part of the brain. Temporal lobe epilepsy (TLE) is one of the most common

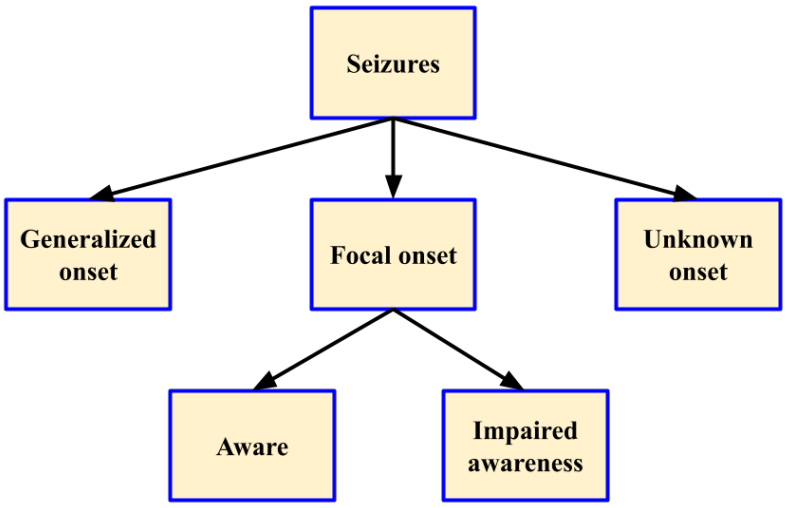

Figure 2.1: The ILAE classification of seizures.

forms of epilepsy. It is the most common focal onset epilepsy [21]. Focal onset seizures are further classified into focal onset awareness if the person is awake and knowledgeable, and focal onset impaired awareness when the seizure affects the patient's awareness. Generalized onset seizures, also called generalized tonic-clonic or grand mal seizures, affect both hemispheres of the brain simultaneously and frequently cause impaired cognition. When the origin of the seizure is undetermined, it's categorized as an unknown-onset seizure.

Note that the seizure types described in this thesis were defined according to the 1981 ILAE seizure classification [20], as portions of the underlying data were recorded and clinically annotated before the introduction of the revised 2017 ILAE classification [22, 23]. To maintain consistency with the original clinical labels and avoid retrospective reinterpretation, the historical classification framework was retained throughout the analysis.

Localizing epilepsy has been one of the crucial tasks since its diagnosis. Some of the early work from [24] defined the concept of epileptogenic lesion responsible for epilepsy as structurally and functionally distributed tissue lesions, but are operationally abutting grey matter. Based on this, Jasper [25] defined the lesion area, epileptogenic focus, and the epileptogenic spikes as three concentric circles with different properties that are evident on EEG analysis. This idea has evolved over the years. From a surgical perspective, SoZ

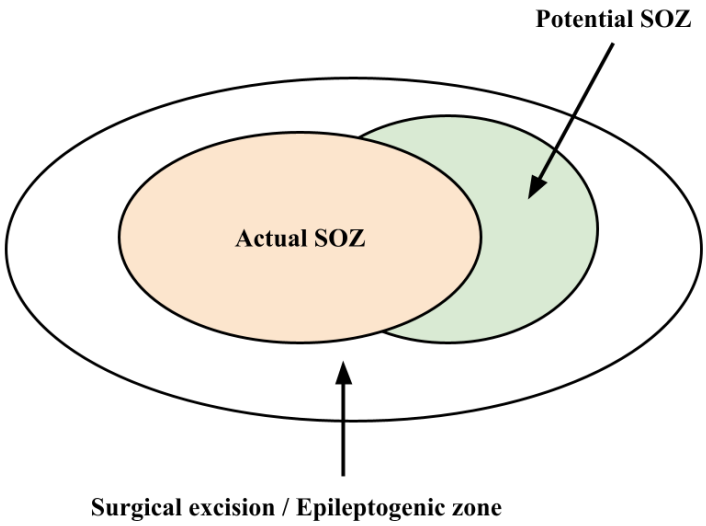

Figure 2.2: The epileptogenic zone (EZ) [1].

is the cortical region where seizures originate. The EZ, which refers to the crucial areas for generating seizures, is equivalent to a subsection of cortical regions that include SoZ and potential SoZ regions inferred from *irritative zone* - the cortical regions that generate interictal spikes [26]. Resection of EZ should render patients free from seizures. Figure 2.2 shows the pictorial representation of the EZ encompassing the actual SoZ and potential SoZ [1]. As the location of EZ cannot be deduced directly, it is inferred mainly from SoZ. Thus, the two terms are used interchangeably [26].

Currently, SoZ localization is performed by epileptologists using visual inspection. First, magnetic resonance imaging (MRI) of the brain is used to locate focal lesions. Then, the ictal EEG or iEEG recordings of individual electrodes are inspected to identify signatures of SoZ, such as low-voltage fast activity and periodic spiking [27]. EEG recorded during sedated sleep, and sleep deprivation has shown to improve the diagnosis of epilepsy by locating sites of epileptiform discharges that were not seen in routine EEG [28]. Thus, an epileptologist may also choose to record EEG (or iEEG) from patients under sedated sleep or sleep-deprived states.

2.3.2 Biomarker I: High-frequency Oscillations (HFOs)

HFOs are widely studied biomarkers for SoZ identification. HFO occurrence rate is frequently used to identify SoZs. In [29], iEEG from ten patients with intractable epilepsy showed significantly higher HFO rates in SoZ channels (24.3 ± 32.4) than in non-SoZ channels (1.9 ± 4.7), particularly for fast ripples. HFOs occurred independently of epileptic spikes, and HFO rate outperformed spikes with 95% specificity. In [30], HFOs were identified in ten presurgical patients using visual inspection and fast Fourier transform. Similar findings appear in [31, 32, 33]. Additional features, such as amplitude, duration, and peak frequency, complement HFO rate for seizure onset prediction [34, 35, 36]. Resection of regions with high interictal HFO activity correlates with improved seizure outcomes in [37, 38, 39], supporting presurgical utility.

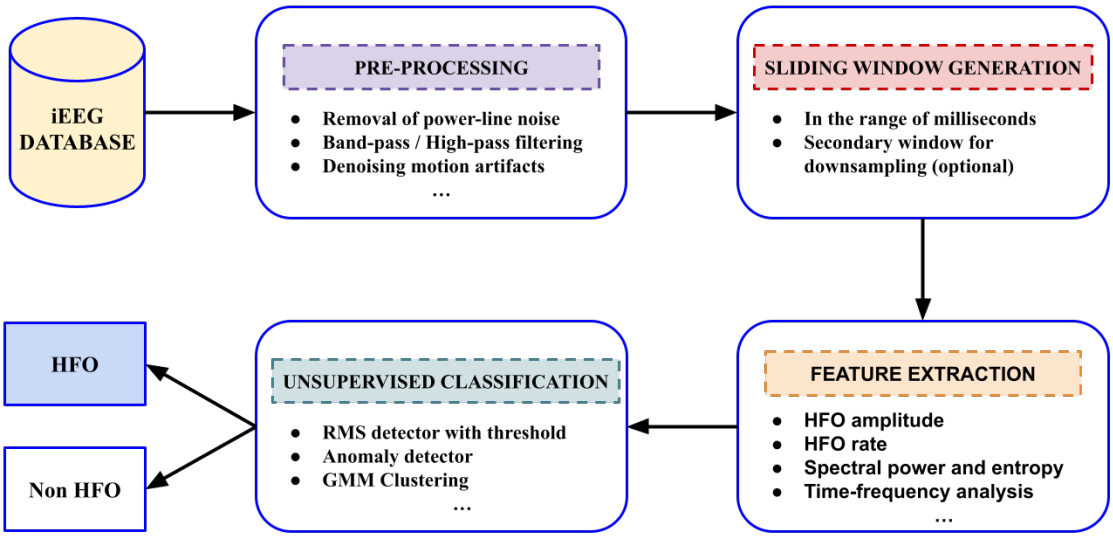

Figure 2.3: Automatic detection of High-frequency Oscillations.

HFOs also occur in non-SoZ regions, prompting classification of pathological versus physiological events. In [35], 35 patients' fast ripples were classified into Type I (IED-coupled) and Type II. Type I ripples had higher frequency ($p = 0.019$), shorter duration ($p = 0.000$), and greater amplitude ($p = 0.000$). HFO power often rises sharply at seizure onset or minutes prior. In [40], power doubled in 40–50 Hz and increased fivefold in 80–120

Hz. Very fast oscillations seconds before seizure onset were observed in [41]. HFOs localized to seizure-onset electrodes for all patients in [42], with high-frequency activity rising up to 20 minutes before seizures. In [43], HFOs above 100 Hz were localized at onset electrodes, with significant power increase 8 seconds before onset.

Automated HFO detection typically uses anomaly detection to identify high-frequency events. Figure 2.3 shows a generalized block diagram. In [44], a moving average RMS amplitude detector in 100–500 Hz with a 3 ms sliding window selected candidate windows exceeding five standard deviations, with at least six rectified peaks required. In [45], similar candidate selection was followed by unsupervised classification using spectral features. In [46], dynamic time warping measured window distances, and clustering separated background from epileptogenic activity.

Interictal spikes (IES), brief EEG events (<250 ms), are sensitive spatial biomarkers but lack specificity [47, 48]. Combining IES with HFOs improves onset zone identification. In [49], a kernelized SVM trained on eight patients and simulated data achieved 95–100% accuracy and 96–100% specificity. IES alone performed poorly, while combining IES with HFOs improved predictions. Sensitivity decreased in noisy simulated data.

iEEG recordings often contain high-frequency artifacts from muscle movements, which can be mistaken for HFOs [50]. Semi-supervised learning or tailored filtering improves detection. In [32], time-frequency denoising removed spikes, followed by clustering based on spectral features to isolate HFOs.

Despite their promise, HFOs vary across brain regions and depend on sleep state and recording conditions. In [51], HFO rate during non-REM sleep was no better than spikes for SoZ localization. In [52], group-level correlations between HFO removal and seizure freedom weakened at the individual level; some patients became seizure-free without resecting most HFO-generating regions. These limitations highlight the need for improved detection methods, discrimination of pathological versus physiological HFOs, and complementary biomarkers to enhance SoZ identification.

2.3.3 Biomarker II: Effective Brain Connectivity Measures

This section reviews two main measures of effective connectivity used to localize seizures: family of Granger Causality (GC) measures and Directed Information (DI). In both cases,

connectivity is represented using edges of a directed graph with channels as nodes. Graph features such as node degree and centrality help distinguish SoZ from non-SoZ channels.

Granger Causality (GC) [53] is a popular method in neuroscience to infer causality between two time series. For two AR time series X and Y, X Granger causes Y if the predictability of Y improves when its past values and X are used. Mathematically, for $\mathbf{X}^N = [x_1, \ldots, x_N]$ and $\mathbf{Y}^N = [y_1, \ldots, y_N]$, consider:

$$y_i = \sum_{j=1}^{L} a_j y_{i-j} + e_i = \sum_{j=1}^{L} (b_j y_{i-j} + c_j x_{i-j}) + \tilde{e}_i$$

Here, e_i is the error predicting Y from its own past, and $\tilde{e}_i$ is the error using both X and Y. X Granger causes Y if the mean squared error $\tilde{e}_i$ is substantially lower than e_i.

GC can be extended to multivariate AR models. Frequency-domain measures such as Directed Transfer Function (DTF) [54] and Partial Directed Coherence (PDC) [55] are useful for identifying interactions in specific frequency bands. In [56], Fourier and wavelet-based GC spectra showed high-frequency causal outflow 2–12 s before visible ictal onset across eight patients, with net causal outflow exceeding three SDs. In [57], 25 patients' interictal iEEG analyzed with GCCA [58] revealed "causal nodes" concentrated around ictally active electrodes, with a chance probability below 10^{-20}.

GC supports the network view of focal seizures. In [59], SoZ is identified as the crucial node whose removal stops seizures. Graph measures such as node degree and centrality help locate these nodes. DTF applied to ictal and interictal windows identified the top 5% strongest causal connections, and k-means clustering based on betweenness centrality separated active regions, correlating with resected areas in seizure-free patients [54, 60]. In [61], time-varying PDC revealed high net outflow regions concordant with clinically estimated EZ in 16 TLE patients.

GC's limitations include linearity assumptions and stationarity requirements, making it less effective for nonlinear brain interactions [62]. Nonlinear GC variants [63, 64, 65, 66] and adaptive approaches [67, 68, 69, 70] address these issues using time-varying MVAR models and adaptive Kalman filters. Adaptive DTF was also used to predict seizure onset and guide resections, with net outflow and out-degree measures corresponding to epileptologists' identified regions [71, 72].

Directed Information (DI) [73] is an information-theoretic causal measure. For N

samples of two channels X and Y, DI quantifies the bits of uncertainty in Y causally explained by X [74, 75]:

$$I(\mathbf{X}^N \to \mathbf{Y}^N) = \sum_{n=1}^{N} I(\mathbf{X}^N; y_n | \mathbf{Y}^{n-1}).$$

DI is model-agnostic and applicable to EEG and ECoG.

In [74], a data-driven likelihood estimator computed DI between electrodes. Net outward flow identified regions with significant causal influence, matching visual expert analysis, with false positives in only two patients. Areas were ranked by clinical relevance based on net outflow.

Combining DI and GC enhances SoZ localization. In [76], an automated algorithm using IEEG-Portal data [77] estimated pairwise causal influence via DI and GC, then applied a PageRank-based scoring [78]. In 17 of 19 patients, the method correctly identified over 50% of onset electrodes or immediate neighbors.

2.3.4 Biomarker III: Phase-Amplitude Coupling (PAC)

Neuronal synchrony can be interpreted as interactions between oscillations at different frequencies. Lower frequencies coordinate long-range interactions, while high-frequency bands mediate short-term local interactions [79]. Oscillations across frequency bands often influence each other, a phenomenon known as *cross-frequency coupling* (CFC) [80]. A major form of CFC is *phase-amplitude coupling* (PAC), where the phase of low-frequency oscillations modulates the amplitude of high-frequency activity. In [81, 82], CFC was shown to reflect ongoing brain activity and vary with cognitive tasks. Ictal iEEG studies [83, 84, 85] demonstrate that PAC between low-frequency phase and high-frequency gamma amplitude effectively characterizes seizures. These observations have motivated studies on PAC's predictive value for SoZ identification [86, 87, 88].

Several methods exist to quantify PAC [89, 90, 91]. A widely used approach is the *modulation index* (MI) [92], based on a normalized entropy measure [93]. The Hilbert transform extracts the high-frequency envelope and low-frequency phase. Phases are binned against the mean amplitude of the high-frequency signal, and the normalized entropy calculates MI. In [86], MI was calculated across all low- and high-frequency pairs during the first four minutes of each sleep stage (N1, N2, N3, REM). Channels were classified as SoZ, exclusively

irritative zone (EIZ), or normal zone (NoZ). MI was higher in SoZ channels than EIZ or NoZ across sleep stages, peaking during deep sleep (N3). Prior work [44, 31] also links sleep modulation of HFOs with PAC. However, regional PAC differences and its role in guiding resections remain unexplored. In [87], phase-locking value analysis similarly showed higher PAC near seizure onset electrodes. In [88], ripples (80–150 Hz) and EEG spikes in mesial temporal lobe epilepsy patients were analyzed during sleep. PAC was assessed via phasor transformation and spectral averaging, showing stronger coupling inside SoZ than outside.

Supervised machine learning supports PAC as a SoZ biomarker. In [94], a logistic regression classifier using PLV features of high-gamma amplitude and low-frequency phase accurately identified SoZ in ten patients. The number of non-resected SoZ electrodes predicted seizure outcome in non-seizure-free patients.

Despite promising results, PAC findings are based on small studies. Their generalizability is uncertain, and the quantitative relationship between PAC and SoZ remains to be fully established. Further work is needed to validate PAC as a reliable biomarker for seizure localization.

2.3.5 Role of Machine Learning in SoZ Identification

The growing availability of data has enabled extensive use of machine learning (ML) algorithms for predictive tasks, including in biological datasets. Many studies have applied ML to brain data for seizure detection and prediction with high accuracy [95, 96, 97, 98, 99]. Seizure localization is more complex, as ground truth from visual inspection is subjective; one way to validate is by assessing post-surgical seizure freedom after resection of the identified regions.

As noted in section 2.3.2, unsupervised ML methods such as clustering and anomaly detection are commonly used for automated HFO identification. These methods accurately locate SoZ, corresponding to resected regions in seizure-free patients. Similarly, graph features from connectivity measures (section 2.3.3) can effectively distinguish SoZ from non-SoZ channels.

Supervised ML methods rely on class labels provided by epileptologists, with features mainly extracted from the discussed biomarkers. Each biomarker alone has limitations;

combining multiple biomarkers can improve predictive power. For example, [100] developed an SVM classifier using local HFO, interictal epileptiform discharge (IED), and PAC features. The combined classifier outperformed unsupervised and single-biomarker approaches, achieving an AUC of 0.79.

The availability of large datasets has also fostered deep neural network (DNN) approaches for SoZ localization. DNNs can better identify viable biomarkers compared to conventional ML. In [101], a convolutional neural network (CNN) using time-frequency maps detected HFOs with higher specificity than standard ML methods. In [102], long short-term memory (LSTM) models generalized across twelve patients from two independent datasets for HFO detection. Recent works also apply DNNs to non-invasive scalp EEG. For instance, the SZLoc architecture [103] combines local and global CNN encodings via a transformer layer, achieving mean accuracy of 71.1% and generalizing across datasets. However, performance is lower than ML approaches using iEEG recordings. These results suggest that DNNs can serve as efficient presurgical tools, enabling rapid analysis of large datasets while improving the identification of potential SoZ regions.

2.3.6 SoZ Review Summary and Identified Gaps

Localizing seizure onset from iEEG can be posed as a classification task to separate electrodes into SoZ and non-SoZ. Various phenomena that are characteristic of the iEEG data from the onset zone have been discussed in this review. One of the most extensively studied electrophysiological biomarkers is HFOs, often identified using unsupervised anomaly detection. An alternate approach is the evaluation of brain networks using graph theory and information-theoretic methods. Measures of synchrony and PAC have also shown increased coupling of specific frequency bands, which can be used to localize SoZ.

However, the true success of an algorithm is measured by post-surgical seizure freedom. Most discussed methods are either retrospective or performed in conjunction with a physician's analysis. The final decision for resective surgeries is still performed by visual inspection, which can result in unsatisfactory surgical outcomes [104].

This review highlights a critical gap, which is also noted in Chapter 4 of this thesis: traditional biomarkers like HFOs and IEDs often fail to capture the complex *network dynamics* underlying seizure generation. While graph-based methods are promising, their

application, especially using advanced non-linear *effective connectivity* measures, remains a key area for development. Part I of this thesis is designed to fill this gap by developing a robust, generalizable framework based on these advanced network features.

2.4 State-of-the-Art in Seizure Prediction: A Review

2.4.1 The Clinical Need and Core Challenges

Beyond localization, the second major goal of computational epileptology is seizure prediction. The persistent unpredictability of seizures poses significant long-term risks and psychological distress for patients with DRE [3, 4]. This has motivated the pursuit of seizure prediction systems designed for *continuous, long-term* monitoring, offering patients improved safety and autonomy [105, 106, 96, 107]. Long-term intracranial EEG (iEEG) monitoring enables real-time seizure forecasting, which can reduce injury risk, improve patient independence, and alleviate anxiety [108, 109, 110, 111, 112, 113, 114].

However, the clinical viability of such systems is fundamentally challenged by the difficulty of maintaining robust performance over extended durations. A primary issue is *alarm fatigue*, where even a low false positive rate (FPR) becomes an unsustainable burden over months, causing patients to lose trust in the device [106, 97, 115, 116, 117]. This challenge is rooted in two core technical problems: the *non-stationarity* of brain signals, which degrades model performance over time [118], and *intra-subject heterogeneity*, a concept that will be explored in detail.

2.4.2 Evolution of Preictal Features and Models

Early efforts in seizure prediction often framed the problem as a binary classification task, distinguishing preictal segments (typically 30 to 60 minutes pre-seizure) from interictal activity [96, 97, 119, 120]. This approach, while an oversimplification, facilitated initial algorithm development.

A vast array of analysis techniques have been applied to identify these discriminating preictal patterns. Initial work often employed nonlinear dynamical features, such as Lyapunov exponents and correlation integrals, to capture changes in brain state preceding a seizure [108, 109, 121, 122, 123]. Later efforts expanded to a wide range of features,

including spectral power, coherence, phase synchronization, and entropy, coupled with various classifiers like Support Vector Machines (SVMs) and artificial neural networks [124, 125, 126, 127, 42, 30]. Peripheral physiological signals, such as heart rate, have also been proposed as complementary predictors [128, 129].

More recently, deep learning architectures such as Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks have been applied to scalp and sub-scalp EEG to leverage spatio-temporal patterns [110, 113, 130, 131, 132, 120, 133, 134], though often at the cost of high computational complexity or reduced interpretability.

A significant leap in understanding came from implanted advisory systems, which revealed that seizure likelihood is not binary but fluctuates over time, often following circadian and multiday cycles [135, 136]. This shift in perspective has driven efforts to move beyond binary labels and toward continuous seizure risk scoring [137, 138, 139, 140].

2.4.3 Real-World Systems and Benchmarking

Real-world datasets from implanted systems have been invaluable for both therapy and research. The NeuroPace RNS trial, for instance, demonstrated a significant median seizure reduction with its brain-responsive stimulation system [141, 142, 143]. Similarly, the NeuroVista advisory system, which used real-time iEEG-based forecasting, achieved high sensitivities (65–100%) in a cohort of 15 patients, validating the potential of real-time forecasting [135]. These systems not only reduce seizure burden but also generate the long-term electrophysiological data critical for algorithm development [144, 145].

In parallel, crowd-sourced initiatives, such as the American Epilepsy Society Seizure Prediction Challenge, have enabled the global research community to test algorithms on publicly available long-term datasets [146, 147]. These efforts have helped benchmark progress and improve reproducibility. The top-performing models in these challenges achieved an average AUC in the range 0.79–0.84 [146]. However, this performance often drops when evaluated on ultra-long iEEG recordings [147], highlighting the challenge of maintaining robust performance over extended periods.

2.4.4 The Unaddressed Gap: Intra-Subject Seizure Heterogeneity

Despite these advances, a fundamental limitation persists. Current approaches, including those in benchmark challenges and clinical trials, often rely on a *single model per patient*. This approach may be insufficient to capture the breadth of seizure variability [148, 149].

The same patient may exhibit different seizure types over time, varying in location, duration, frequency content, and electrophysiological progression [150, 151, 152]. This diversity is not merely inter-patient, but deeply intra-patient. A dynamical systems framework has even been proposed to capture the range of seizure onset and offset behaviors, further supporting the idea of seizure diversity [153].

This failure to account for seizure heterogeneity is a primary reason for poor long-term performance. A model trained on all seizures as if they were identical, or trained on only one subtype, will fail when a different seizure type with a distinct preictal signature occurs. Addressing this diversity requires more granular modeling strategies, such as the personalized signature patterns that have proven effective in seizure detection [154, 155]. This review identifies a critical gap in the field: the need for a seizure prediction framework that explicitly and automatically models this intra-subject seizure heterogeneity.

Chapter 3

SoZ Identification using Effective Brain Connectivity Measures

3.1 Chapter Overview

As established in Chapter 1, accurately localizing the SoZ is a critical, unresolved challenge in pre-surgical planning for drug-resistant epilepsy. Chapter 2 provided the technical background for analyzing brain networks, highlighting the limitations of traditional, univariate biomarkers and the theoretical advantages of using effective connectivity (EC) and graph theory.

This chapter presents the first application of this network-based framework. The objective is to demonstrate the capability of utilizing graph feature-based supervised algorithms on iEEG recordings for the identification of SoZs in individuals with drug-resistant epilepsy. This study aims to apply novel data-driven techniques for estimating the directed propagation of signals from iEEG recordings to precisely localize the SoZ.

To this end, we extract EC graphs from *ictal* iEEG recordings using three emerging methods: directed information (DI), mutual information-guided Granger causality index (MI-GCI), and frequency-domain convergent cross-mapping (FD-CCM). Graph centrality measures at different sparsity levels are then utilized as features for the classifier and analyzed through interpretable machine learning models. **A key innovation of this work is the novel combination of these advanced EC algorithms with topological graph**

features to improve automated SoZ mapping. We utilize an open-source dataset of drug-resistant epilepsy patients from OpenNeuro (ds004100) to enable reproducibility.

The content of this chapter is adapted from [156]. The subsequent sections will demonstrate that this approach is highly effective, with centrality features achieving high accuracies exceeding 90% in distinguishing SoZ electrodes from non-SoZ electrodes. We will demonstrate that FD-CCM centrality measures are particularly significant, achieving a mean AUC of 0.93 and outperforming the prior literature. This research highlights the potential of specific EC measures in enhancing discriminative power within the context of epilepsy research.

3.2 Dataset Description

Our investigation utilized de-identified iEEG data obtained from 58 individuals undergoing surgical treatment for drug-resistant epilepsy at the Hospital of the University of Pennsylvania (HUP) [157]. Every participant involved in this research provided written informed consent in alignment with the Institutional Review Board at the University of Pennsylvania. Subjects underwent intracranial EEG using subdural grids, strips, and depth electrodes (ECoG) or stereotactically placed depth electrodes (SEEG), followed by surgical resection or laser ablation. The sampling frequency (F_s) and the number of electrodes varied based on the subject and the type of electrode used. The dataset comprises iEEG recordings during interictal and ictal periods, as well as electrode localizations in ICBM152 MNI space [158]. Additionally, it includes bad channel information, clinically identified seizure onset channels, channels intersecting with the resection/ablation zone, and the surgical outcome, as per the Engel Epilepsy Surgery Outcome Scale [159], for each subject.

Given the subjective nature of determining the actual SoZ channels, contingent upon seizure freedom after resection or ablation, our analysis focused solely on patients achieving seizure-free status (Engel Class I). Subsequently, the investigation is confined to seizure-free subjects with complete electrode location data, resulting in a subset of 28 individuals from the initial 58. Including detailed electrode location information is crucial for addressing the high correlation between neighboring electrodes and SoZ electrode activity, necessitating precise considerations in algorithm development for distinguishing SoZ from non-SoZ

electrodes. Finally, as data-driven methods are used to estimate EC, which requires a sufficient sample size, we excluded recordings from a single subject whose iEEG was sampled at 256 Hz. A summary of the iEEG recordings from these subjects, including details such as iEEG type, number of channels, gender, and epilepsy type, is presented in Fig. 3.1. Patient details, including epilepsy type, iEEG type, and SoZ location for all subjects from the dataset, are included in Appendix A.1.

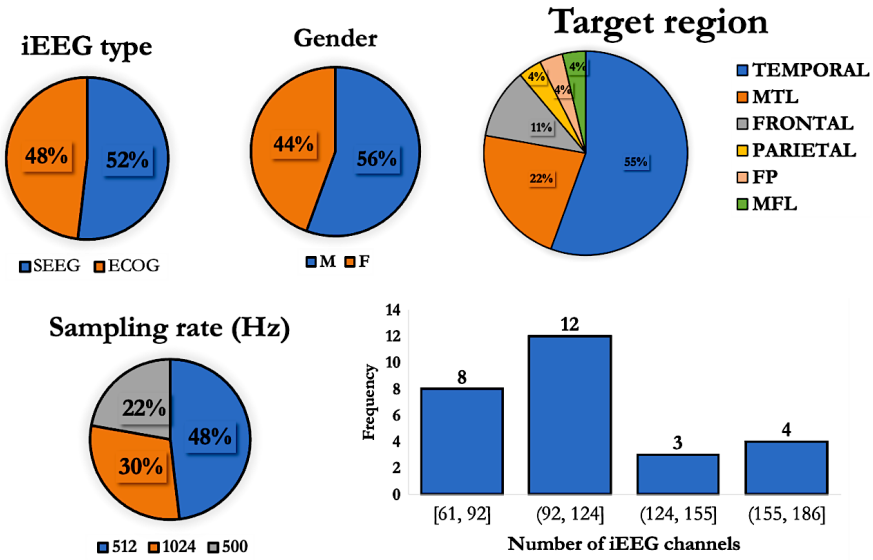

Figure 3.1: Summary of Demographics and iEEG Recording Characteristics for the 27 Subjects Analyzed in This Study.

To select the best time frame for constructing the classifier, features were extracted from preictal, ictal, and interictal time windows. We then compared discriminability using Welch's two-sample test, assessing both the number of features with statistically significant mean differences ($p < 0.05$) and the t-statistic value for measuring the magnitude of feature distinctions between the groups. Analysis revealed that the 30-second window centered around the seizure onset displayed a higher average t-value for the ten features of the two classes – SoZ vs. non-SoZ (see A.2). Consequently, the graph model is derived from the 30-second interval centered around the onset of each ictal recording. The preprocessing steps executed before generating the graphs include eliminating iEEG data from unreliable

electrodes identified by the experts during data curation, employing a 6th-order notch filter centered at 60 Hz to eliminate power noise, applying a 6th-order high-pass filter with a passband greater than 0.25 Hz to eliminate drift, and conducting average referencing of signals from reliable electrodes.

3.3 Methodology

Figure 3.2 offers an overview of the steps involved in the classification process at a high level. Each of these steps is elucidated in detail within this section.

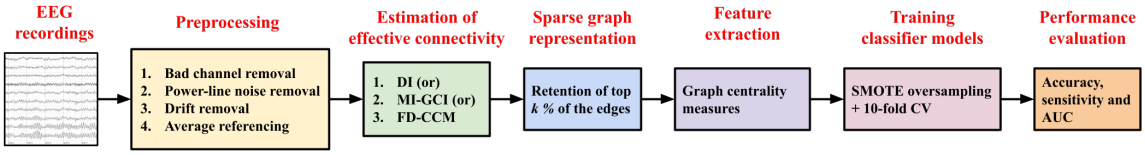

Figure 3.2: Methodology overview: Identifying SoZ from iEEG recordings through the application of graph centrality measures for the classification of electrodes into SoZ or non-SoZ types.

3.3.1 Effective Connectivity (EC) Measures

The iEEG recordings during seizure onset are converted from time-series to graph signals. For this, three model-free brain EC measures are assessed. These three measures are described below:

1. **Directed information (DI)**: This is a measure rooted in information theory that quantifies the impact of one process on inferring causality in another process [73]. With reference to [74], the DI from channel *'x'* to *'y'*, denoted by $I(\mathbf{X}^N \rightarrow \mathbf{Y}^N)$ is expressed mathematically as

$$\hat{I}(\mathbf{X}^N \rightarrow \mathbf{Y}^N) = \hat{h}(\mathbf{Y}^N) - \hat{h}(\mathbf{Y}^N||\mathbf{X}^N) \tag{3.1}$$

where $\hat{h}(\mathbf{Y}^N)$ and $\hat{h}(\mathbf{Y}^N||\mathbf{X}^N)$, respectively, denote the differential entropy of $\mathbf{Y}^N$ and the differential entropy of $\mathbf{Y}^N$ causally conditioned by $\mathbf{X}^N$ [74]. Data-driven

likelihood estimator adapted from [12] is used to obtain the differential entropy estimates. The mathematical expressions are provided by the equations (3.2) and (3.3), where J_{yy} and K_{yx} are the Markovian parameters that correspond to the number of past samples of $\mathbf{X}^N$ and $\mathbf{Y}^N$ that influence the present sample of $\mathbf{Y}^N$, respectively.

$$\hat{h}(\mathbf{Y}^N) = \frac{1}{N}\sum_{n=1}^{N}\{-logP(y_n)\} \tag{3.2}$$

$$\hat{h}(\mathbf{Y}^N||\mathbf{X}^N) = \frac{1}{N}\sum_{n=1}^{N}\{-logP(y_n|\mathbf{Y}_{n-J_{yy}}^{n-1}, \mathbf{X}_{n-(K_{yx}-1)}^{n})\} \tag{3.3}$$

2. **Mutual information-guided Granger causality index (MI-GCI)**: This is a data-driven causality index that uses mutual information (MI) to overcome the limitations of GC, i.e., linearity and stationarity assumptions. In the context of GC, MI assesses the predictability of one variable based on the past values of another variable beyond what can be explained by the past values of the variable itself. The GC index using MI is estimated by calculating the Kullback–Leibler (KL) divergence [160] between the probability distributions $p(\mathbf{Y})$ and $p(\mathbf{Y}|\mathbf{X})$ [161]. Note that the two distributions are equal when $\mathbf{X}^N$ and $\mathbf{Y}^N$ are independent. The KL divergence is given by

$$F_{X\to Y} \equiv I(y_{i+1}; \mathbf{X}^i|\mathbf{Y}^i) = h(y_{i+1}|\mathbf{Y}^i) - h(y_{i+1}|\mathbf{X}^i\mathbf{Y}^i) \tag{3.4}$$

where $I(;)$ and $h()$ represent mutual information and differential entropy, respectively. The entropies are estimated using covariance-based approximations discussed in [161].

3. **Frequency-domain convergent cross-mapping (FD-CCM)**: This is a novel measure that infers causality in the frequency domain [162] (`https://github.com/parhi/Freq-domain-CCM`), inspired by the time domain causality index of convergent cross-mapping [163]. The motive behind this measure is that if a time-series $\mathbf{X}^N$ has a causal influence on the time-series $\mathbf{Y}^N$, then $\mathbf{X}^N$ will influence the frequency dynamics of $\mathbf{Y}^N$. Hence, it becomes feasible to determine the frequency dynamics of $\mathbf{X}^N$ from those of Y. This involves initially projecting the time-series into a higher-dimensional space derived from the time-frequency mapping of the two

series. Subsequently, the embedding of $\mathbf{X}^N$ (M_X) is estimated based on the embeddings of $\mathbf{Y}^N$ (M_Y) using a straightforward projection estimate. The *cross-mapping score* (CMS), represented by the absolute correlation coefficient between the original signal $\mathbf{X}^N$ and its reconstructed version from $\mathbf{Y}^N$, serves as a metric for causal coupling.

$$\text{CMS } = \rho_{\hat{\mathbf{X}}|M_Y} = \text{corr}(\mathbf{X}^N, \hat{\mathbf{X}}^N) \tag{3.5}$$

Here $\hat{\mathbf{X}}^N$ is the reconstructed version of $\mathbf{X}^N$ obtained from the weighted estimates of the projections from M_Y. A higher CMS signifies a strong correlation between the reconstructed and original signals, indicating a strong causal effect of $\mathbf{X}^N$ on $\mathbf{Y}^N$.

Table 3.1 describes the underlying assumptions, crucial parameters, and the EC estimation process for the three EC measures.

3.3.2 Extraction of Graph Centrality Features

Graph *centrality* refers to mathematical measures that assign each node in a network a quantitative score based on its position and connectivity patterns within the broader topology. These measures characterize the structural importance of nodes in facilitating communication across the graph. Key local centrality metrics such as *degree centrality* and *eigenvector centrality* reward connections to highly influential nodes. Global metrics like *betweenness centrality* instead assess integral roles in mediating flows between distal regions.

Brain graphs created from electrode connections based on electrical signal correlations have been extensively studied to understand healthy cognition and neurological disorders. Nodes with high centrality serve as critical points for information transfer or computational hubs within brain systems. When centrality measures are applied to EC measures, they can capture directional influences that shape brain-wide dynamics more accurately. In epilepsy, changes in network communication patterns are thought to trigger transitions into hypersynchronous ictal states. Analyzing centrality metrics in epileptogenic networks could pinpoint regions impacted by altered functional roles and interactions.

Sparsifying a graph when extracting centrality measures provides practical benefits such as computational efficiency, improved interpretability, noise reduction, and a focus on

Table 3.1: Key Parameters and Steps Used in Estimating Different EC Measures

EC measure	Assumptions, key parameters, and estimation method
DI	• The 30-sec iEEG recording is resampled to 100 Hz and is split into 1-sec windows. • The stationary assumption confirmed using the augmented Dickey-Fuller (ADF) test [164] and the optimal lag/Markovian parameters are determined using average mutual information (AMI) [165] to be $J_{yy} = K_{yx} = 5$ samples. • The likelihood or probability density function is estimated using Gaussian kernel function-based estimators.
MI-GCI	• The estimation of MI-GCI and suitable windows are adapted from the Frites python toolbox [166, 167]. • The iEEG recording is resampled to 500 Hz to maintain uniformity across the different subjects. • The 30-second recording is divided into smaller windows of 1 sec with a high overlap of 0.8 sec to avoid the effects of windowing. • Using an optimal lag of 10 samples determined using AMI, the MI-GCI value is estimated for each time window and is averaged across the windows to obtain the mean MI-GCI estimate.
FD-CCM	• The 500 Hz resampled recording is split into 500 ms windows with 95% overlap, as suggested in [162]. • The higher-dimensional embedding process included partitioning the signal into distinct frequency bands. We divided the signal evenly in the logspace. This approach leads to more bands at lower frequencies, where most spectral information is concentrated, and fewer bands at higher frequencies, which contain more sparsely distributed spectral information. • The CMS is evaluated using the coefficient of determination (R^2 value) averaged across the different time windows.

key features, making the analysis more manageable and insightful. Hence, we examined centrality measures across various sparsity levels, considering only the top k % of directed edges. Ten different centrality measures are used as input features to the classifier. They are extracted for the three graph types generated by the three EC measures, and their predictive performance is compared across different sparsity levels. The ten centrality features include indegree and outdegree (degree centrality), in-closeness and out-closeness centrality [168], local clustering coefficient [169], local efficiency [170], betweenness centrality [171], PageRank centrality [172], and hub and authority scores computed using the HITS algorithm [173].

3.3.3 Classifier Model

The ten centrality values derived from an EC measure serve as input features for the classifier. Given the low dimensionality, no additional dimensionality reduction techniques are included. The group-level mean classifier performance is evaluated using a 10-fold cross-validation scheme. Ensuring that each subject appears only once in a fold enhances model generalization to diverse data and mitigates biases or information leakage specific to individual subjects. Thus, the frequency of seizures of a subject does not influence the identification of the SoZ. Even if one subject experiences more seizures, those data points are exclusively used to train the model tested on another subject. Due to a significant class imbalance, with approximately 11 times more data points for non-SoZ channels than SoZ channels, the training data is oversampled using the synthetic minority oversampling technique (SMOTE) [174] before model training.

By employing data-driven feature extraction guided by domain knowledge and focusing on a limited set of pertinent features, it becomes feasible to leverage simple machine learning models for classification instead of opting for deep neural networks. A range of techniques, such as support vector machines (SVM), ensemble methods, multilayer perceptrons (MLP), boosting strategies, and decision trees, are tested for classification. This approach provides flexibility in model selection, catering to the specific characteristics of the problem while still preserving interpretability.

The electrodes are classified into SoZ and non-SoZ based on expert-marked locations.

However, it is essential to acknowledge that the electrical activity of electrodes neighboring SoZ can be highly correlated with the activity of SoZ electrodes. Therefore, additional steps are required to address neighboring electrodes, and various methods can be employed. For instance, in [76], the unsupervised algorithm's success is defined by achieving at least a 50% overlap between algorithm-identified SoZ channels and the ground truth. Meanwhile, in [175], a comparative analysis involves nullifying the activity of k nearest electrodes neighboring the k SoZ electrodes. In this study, though the class labels for training rely solely on expert identification, the loss computation is modified to avoid penalizing misclassifications of neighboring electrodes during training, i.e., the loss associated with the neighboring electrodes is not included in the loss function. This involves calculating the Euclidean distance between each pair of electrodes and identifying the k neighbors nearest to SoZ, where k corresponds to the number of expert-identified SoZ electrodes for a specific subject. An example showing the identified neighbors for one patient (HUP065) is shown in Figure 3.3. The open-source Python packages sci-kit-learn [176] and sci-kit-optimize [177] are used to build the classifiers and optimize the loss function.

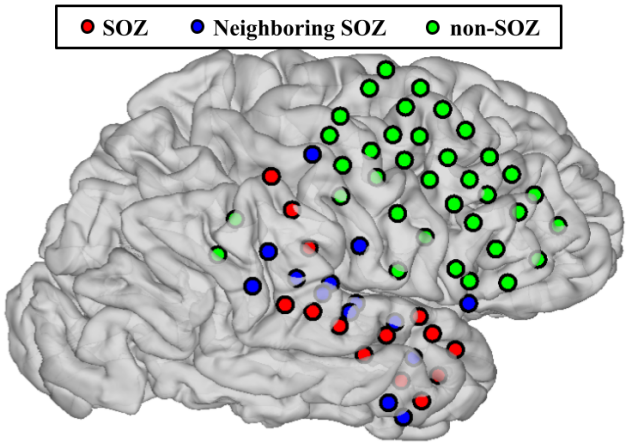

Figure 3.3: Schematic of the cerebral cortex with grid electrode locations of the good electrodes for the subject HUP065. The loss function of the classifier omits any misclassification of the neighboring electrodes (in blue) closest to the SoZ electrodes (red). The non-SoZ electrodes, further away from the SoZ electrodes, are marked in green.

3.4 Results and Discussion

3.4.1 Statistical Analysis of Graph Centrality Measures

The predictability of a feature is defined by the notable variation in its value among distinct classes. Hence, we compared ten centrality measures derived from each EC measure at different sparsities. The goal is to identify the optimal sparsity level capable of effectively distinguishing between the two classes - SoZ and non-SoZ electrodes. This analysis uses a two-means t-test with a significance level α set at 0.05. In a two-means t-test, the t-value measures the difference between the means of two groups, considering the variability within each group. Meanwhile, the p-value reflects the likelihood of observing such a difference under the assumption that no genuine disparities exist between the groups. If the p-value is below the designated significance level (0.05 in this case), the observed difference is statistically significant. A preference for a higher t-value arises in classification, as it indicates greater differentiation between the groups. This becomes particularly crucial for classification, as a more pronounced disparity facilitates the establishment of meaningful boundaries. Higher t-values contribute to better discrimination, aiding in identifying features or variables relevant for classification.

Figure 3.4 and Figure 3.5 provide an overview of the t-test outcomes for the three measures across various sparsity levels. The sparsity level, spanning from 10% to 100%, indicates the top k% of edges preserved from the initial fully connected graph rather than the proportion of all edges present in the SoZ. Figure 3.4 emphasizes the count of centrality features among the ten extracted that exhibit a significant difference in mean values between the two classes with statistical significance, considering different sparsity levels. Figure 3.5, on the other hand, portrays the fluctuation in mean t-values for the statistically significant features. As previously mentioned, the preference is for more discriminative features with a higher average t-value. Hence, the sparsity that results in the highest product of the mean t-value and the number of significant features is chosen as the optimal choice for a given EC measure. Following this criterion, the optimal sparsity, measured by the % of retained edges, is 10% for centrality measures derived from DI and FD-CCM and 90% for MI-GCI.

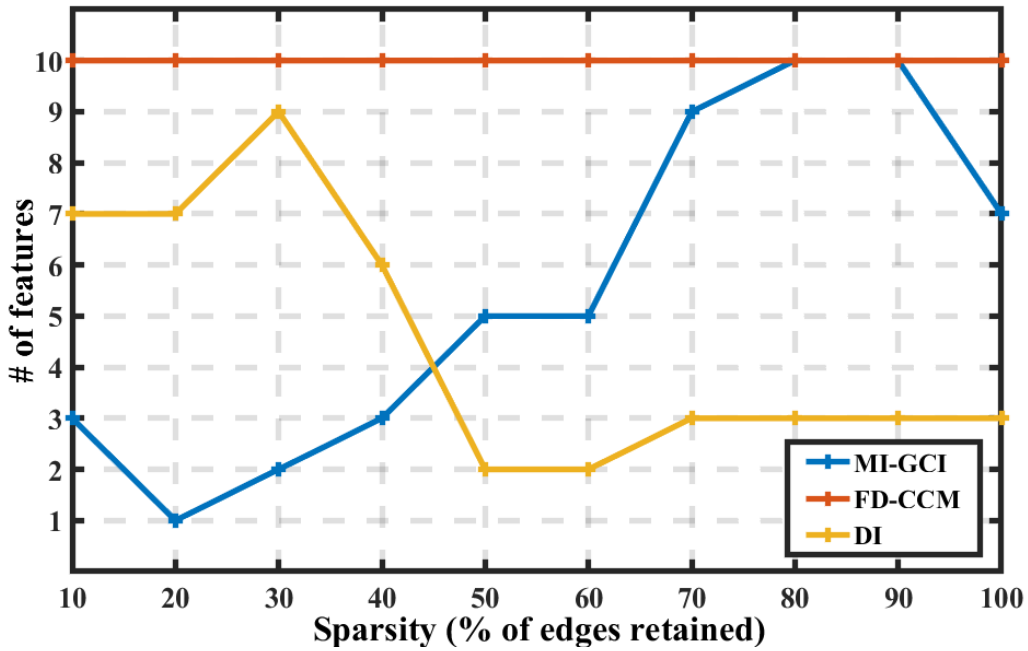

Figure 3.4: SoZ vs. non-SoZ electrodes: Number of statistically significant ($\alpha = 0.05$) graph centrality features extracted from the three EC measures for different levels of sparsity.

3.4.2 Classifier Results

The classification performance of graph centrality extracted from each of the three effective connectivity (EC) measures is summarized in Table 3.2. Due to the unbalanced nature of the data, in addition to overall accuracy, mean sensitivity, specificity, and AUC values were computed for each case to understand class-level accuracy measures. Furthermore, the table also includes information on the classifier model that yielded the best performance and the sparsity level used for each EC measure while generating the centrality features. All three EC measures demonstrated strong performance in the classification task, achieving a mean accuracy exceeding 90%.

The centrality features derived from DI required a sparse graph model, retaining only 10% of the top edges. The applied multi-layer perceptron (MLP) model, comprising two layers with ten neurons in the first layer and 20 neurons in the second layer and sigmoid activation at the output layer, yielded a mean accuracy of 92.12%. This model exhibited a

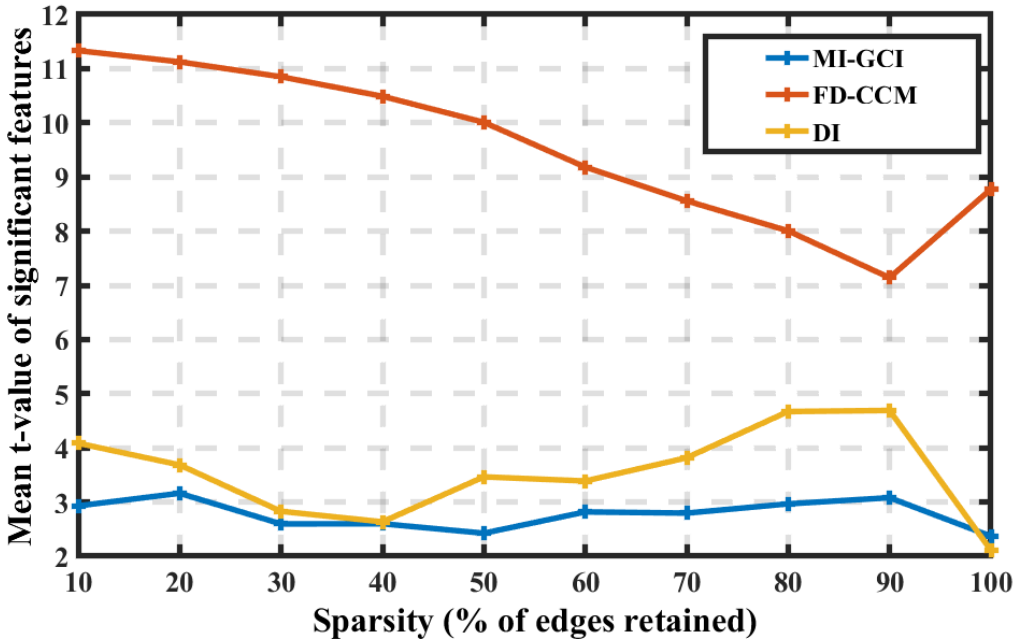

Figure 3.5: SoZ vs. non-SoZ electrodes: Mean *t-value* of statistically significant ($\alpha = 0.05$) graph centrality features extracted from the three EC measures for different levels of sparsity.

mean AUC value of 0.89, a sensitivity of 85.3% for the SoZ class, and a specificity of 92.8% for the non-SoZ class. In contrast, the MI-GCI-based features necessitated a denser graph model with 90% of the edges to achieve comparable performance. Employing a support vector machine (SVM) classifier with radial basis function (RBF) kernels, this feature set achieved the highest mean accuracy among the three EC measures, reaching 96.32%. The superior accuracy was accompanied by a sensitivity of 89.74%, which is lower than that of the FD-CCM-based model described below.

Like DI, the FD-CCM-based centrality measures required a sparse graph model with the top 10% of the edges to attain competitive performance using a standard SVM classifier. Notably, this feature set exhibited reduced bias in the trained model compared to the MI-GCI-based model, with a mean sensitivity of 92.3%, specificity of 94.25%, and the highest mean AUC value of 0.93.

Table 3.2: SoZ vs. non-SoZ: Mean 10-fold cross-validation performance of graph centrality features extracted from the three EC measures from 27 subjects. The mean performance is measured by sensitivity (SS), specificity (SP), accuracy (ACC), and AUC. Additionally, the F1-score and positive predicted value (PPV) are shown.

EC	**Sparsity**	**Classifier**	**SS(%)**	**SP(%)**	**ACC(%)**	**PPV(%)**	**F1**	**AUC**
DI	10 %	2-layer MLP (10 units + 20 units)	85.30	92.8	92.12	54.22	0.66	0.89
MI-GCI	90 %	SVM (RBF kernel)	89.74	96.32	95.72	70.92	0.79	0.90
FD-CCM	10 %	SVM (RBF kernel)	92.30	94.25	94.10	61.6	0.75	0.93

While all three EC measures demonstrated strong classification performance, each exhibited unique characteristics. DI-based features, implemented in a sparse graph model and an MLP architecture, yielded a balanced accuracy profile. MI-GCI features achieved the highest mean accuracy, albeit with a trade-off in sensitivity. With a sparse graph model, FD-CCM features demonstrated superior discrimination ability, boasting the highest mean AUC value. For this classification scheme, we prioritized higher sensitivity to ensure the identification of all SoZ electrodes. However, this choice leads to increased false positives (or lower positive predicted values (PPV)), especially in heavily imbalanced test data, exacerbating the issue of false positives. These results emphasize the need to customize connectivity measures and models for specific neuroscientific applications, carefully balancing accuracy, sensitivity, and specificity.

3.4.3 Testing Pre-trained Models on Subjects with Seizure Recurrence

The classifier models, trained using recordings from 27 subjects, are evaluated on subjects from the dataset who did not achieve seizure freedom post-surgery (Engel classes II, III, and IV). Since most electrodes in a given recording are labeled as non-SoZ, the models are

anticipated to exhibit higher specificity. To ascertain the actual agreement between SoZ electrodes identified by epileptologists and those identified by the algorithm, the sensitivity values (accuracy of SoZ classification) for the three EC measures are examined. It is important to note that channel location data is unavailable for all subjects. Thus, we only applied the relaxation of misclassification for neighboring electrodes to recordings with channel location information. The sensitivity, i.e., the overlap percentage of algorithm-identified SoZ with that identified by the experts, is depicted through a boxplot for the three EC measures in figure 3.6. Here, each data point on the boxplot represents the sensitivity of the model tested on a seizure episode.

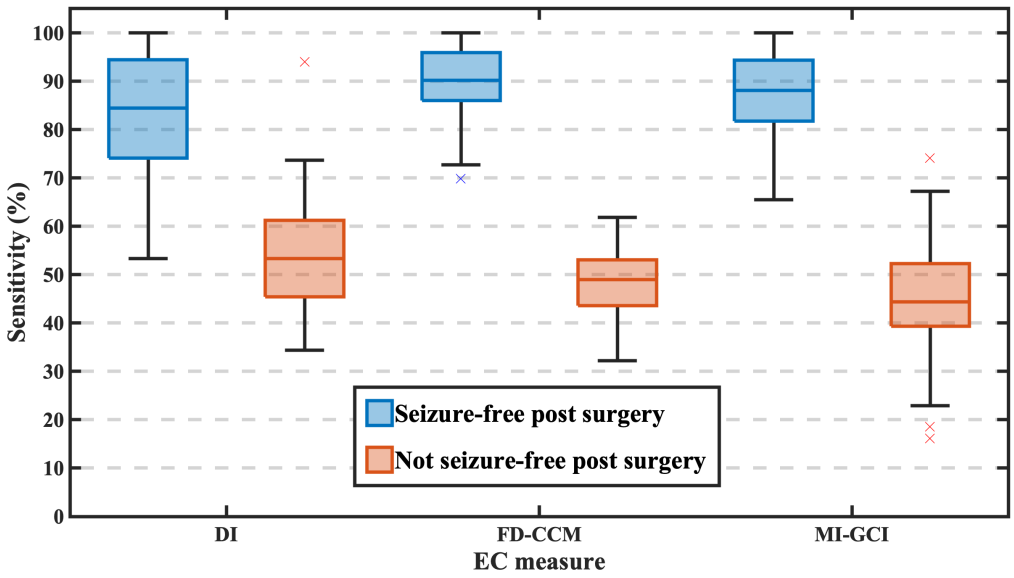

Figure 3.6: Comparison of sensitivity for the three EC measures of models pre-trained on recordings from the 27 Engel I subjects and tested on the 22 subjects with Engel II-IV surgical outcomes. Additionally, the sensitivity of seizure episodes from the 27 Engel I subjects is provided for comparison.

The figure illustrates that the median sensitivity ranges from 45% to 50% for the three measures across subjects belonging to Engel classes II to IV. The lower percentages indicate a weaker agreement between algorithm-identified and epileptologist-identified SoZ electrodes. This observation reveals a correlation trend between a lower sensitivity and

the absence of seizure freedom post-surgery. Figures 3.7 and 3.8 depict specific examples of model prediction employing FD-CCM-based graphs for subjects HUP065 (Engel I) and HUP075 (Engel IV), respectively, to aid in visualizing the results. Note that these two figures do not show non-SoZ electrodes.

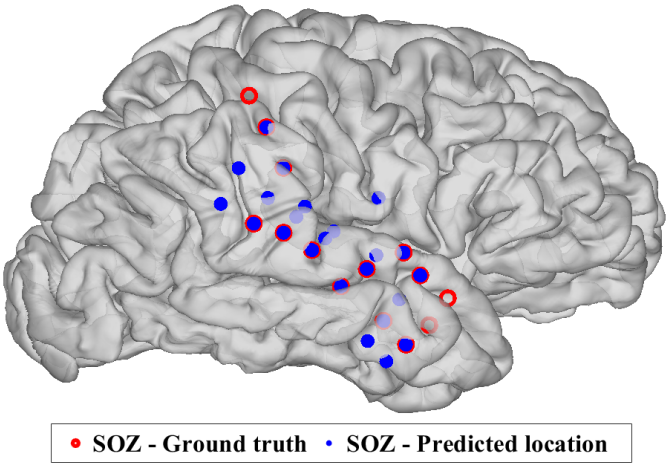

Figure 3.7: Schematic of the brain of subject HUP065 (Engel I) with electrode locations showing the ground truth of SoZ identified by experts and the algorithm-identified SoZ. Non-SoZ electrodes are not shown in this figure.

3.4.4 Feature Ranking

Performing feature ranking after classification is an indispensable step with multiple benefits. It facilitates the identification of the most informative features, shedding light on the key factors influencing the classification outcomes. Additionally, it enhances model interpretability, enabling researchers to discern the variables driving the classification results. This process also guides researchers in focusing on critical variables, optimizing resource allocation, and refining data collection strategies for future studies.

The minimum redundancy maximum relevance (mRMR) technique is employed to rank the ten input features for the three trained models. By assessing MI scores, the algorithm prioritizes features that exhibit high relevance to the target variable while minimizing redundancy among the selected set [178]. The superior ranking of features obtained through

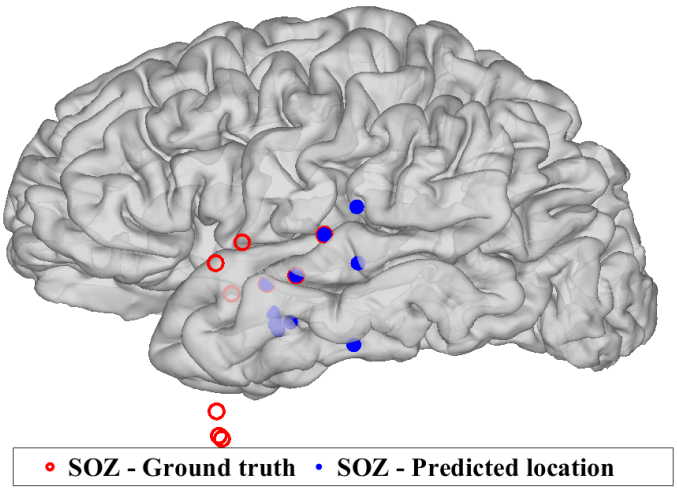

Figure 3.8: Schematic of the brain of subject HUP075 (Engel IV) with electrode locations showing the ground truth of SoZ identified by experts and the algorithm-identified SoZ. Electrodes on the inferior surface are spatially projected outward for visualization. Non-SoZ electrodes are not shown in this figure.

mRMR enhances the performance of machine learning models [179].

Figure 3.9 displays the feature ranking based on mRMR. Notably, the importance scores of FD-CCM measures stand out significantly compared to the other two EC measures. This outcome aligns with expectations, considering the higher separability of features, as indicated by the elevated t-values observed when comparing centrality measures between the two classes. Across both MI-GCI and FD-CCM-based models, the consistently top-ranked features include Authority, Betweenness Centrality (BC), and Out-Closeness, underlining their pivotal roles in accurate classification. Intriguingly, the DI-based model relies more heavily on features such as Indegree and Hubs for effective classification, illustrating nuanced preferences for specific centrality measures across different EC methods.

The discerning outcomes of the feature ranking underscore the nuanced significance of centrality measures derived from diverse EC methods. With a specific emphasis on the distinctive contributions of FD-CCM measures, our subsequent exploration delves into a thorough examination of their feature characteristics. This entails a focused investigation

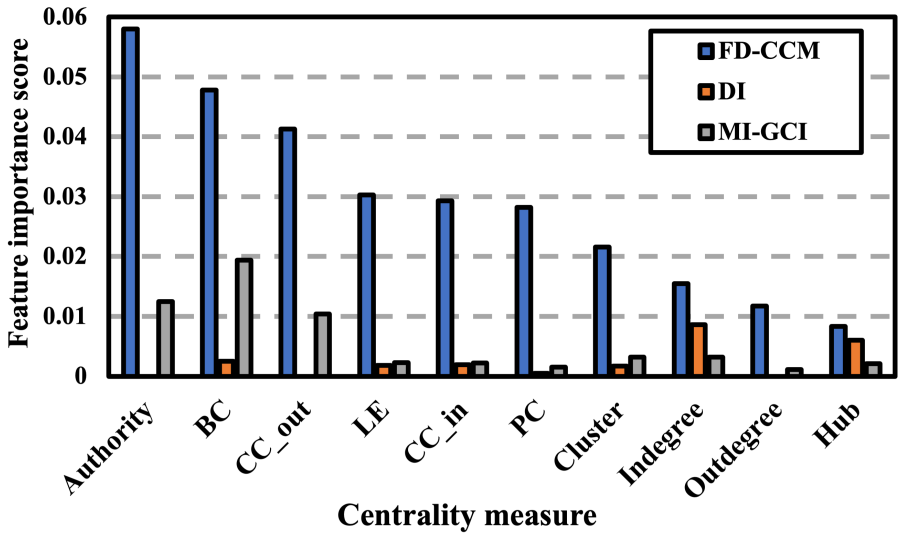

Figure 3.9: Ranking of graph centrality features for 10% sparse DI graph, 90% sparse MI-GCI graph, and 10% sparse FD-CCM graph.

into the discriminative power and potential variations within FD-CCM features, primarily through analyzing the distribution of centrality measures across the two classes. Figure 3.10 portrays this distribution using boxplots representing different features of FD-CCM measures. This thorough examination aims to clarify how specific centrality measures contribute to the classification process and to pinpoint any unique patterns or characteristics that may further refine our understanding of these measures in the context of effective connectivity.

All ten centrality measures consistently demonstrated elevated values in the SoZ electrodes compared to the non-SoZ electrodes, and this difference was found to be statistically significant ($p < 0.05$). The heightened centrality values across these measures suggest that the nodes corresponding to SoZ electrodes hold increased significance or influence within the network. In the specific context of seizure onset, the nodes exhibiting heightened centrality are pivotal in initiating or facilitating the spread of seizures. Consequently, the algorithm rightfully identifies these nodes as SoZ electrodes, underlining their critical role in the epileptic network dynamics.

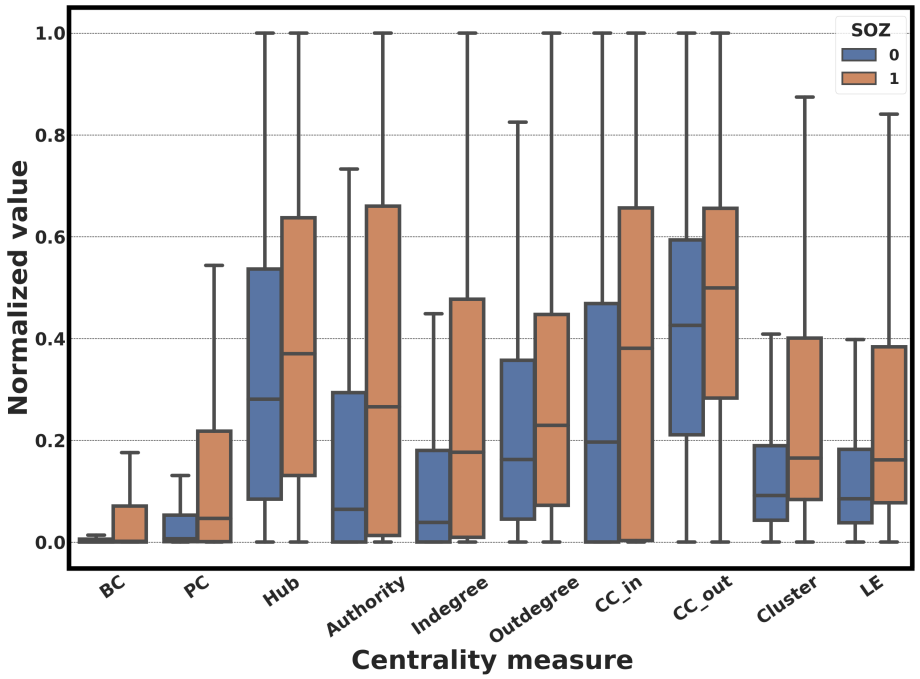

Figure 3.10: The distribution of the ten graph centrality measures extracted from the FD-CCM-based graph at 10% sparsity. SoZ = 1 and SoZ = 0 correspond to the SoZ and non-SoZ electrodes, respectively.

3.4.5 Performance Comparison with Previously Established Methods

The performance of EC-based graph centrality aligns with established state-of-the-art methodologies in identifying SoZ [74, 94, 76, 180, 100, 103, 181], and particularly the FD-CCM-based measure outperforms the prior literature. The primary impediment to direct comparisons lies in the limited availability of public datasets. Additionally, the recording stages of iEEG vary across different studies. Findings from the above-mentioned methods suggest that ictal recordings offer superior performance in localizing SoZ. However, ensuring electrode implantation and signal recording during seizures is only sometimes feasible. This renders interictal recordings more advantageous as they capture data during the subject's resting state. Furthermore, the inherently invasive nature of iEEG, requiring surgical electrode implantation, prompts consideration of scalp EEG as an alternative. Nonetheless,

scalp EEG tends to be noisier due to artifacts from muscle and eye movements, necessitating a delicate balance between invasiveness for signal acquisition, the recording phase, and modeling performance. Table 4.4 offers a comprehensive comparison of our work with select prior literature. We note that the proposed approach achieves the best accuracy and sensitivity among all prior approaches despite the heterogeneity of data and methods. When evaluating performance, we encourage readers to take into account the trade-off considerations highlighted earlier.

Table 3.3: Performance Comparison of FD-CCM-based Graph Centrality Measures with Prior Work on Identifying SoZ.

Method (year)	**Dataset**	**Algorithm type and description**	**Measure of success**	**Performance metric**
Malladi *et al.*[74] (2016)	ECoG sampled at 1 kHz from the study group of 5 patients during the ictal stage.	Unsupervised: Nodes with significant net outward flow of information ($> 5\%$) extracted from DI assigned as SoZ electrodes	Identification of electrodes assigned as SoZ by ground truth (sensitivity)	80% (successful identification in 4 out of 5 subjects).
Elahian *et al.*[94] (2017)	Study group of 10 subjects with subdural ECoG recorded at 1 kHz during the preictal and ictal stage.	Supervised: Logistic regression classifier using features extracted from the phase-locking value	Accuracy (Acc), AUC, and precision	Acc = 83%, AUC = 0.69, and precision = 90%

Method (year)	**Dataset**	**Algorithm type and description**	**Measure of success**	**Performance metric**
Murin *et al.*[76] (2018)	ECoG recorded from 19 subjects with F_s ranging from 500 Hz to 5 kHz [77].	Unsupervised: The pairwise causal influence is first estimated using DI and GC methods. Then a variant of the PageRank algorithm followed by post-processing scoring is used for SoZ inference.	$\geq$ 50% overlap with ground truth SoZ electrodes or its immediate neighbor	Success rate = 89.5% with 0.03 false positive rate (FPR).
Quitadamo *et al.* [180] (2018)	SEEG/SEEG + ECoG iEEG recorded from a study group of 12 subjects at 2kHz	Unsupervised: SoZ is identified by identifying high-frequency oscillations (HFOs), a well-known biomarker of SoZ.	Mean SS and SP	Mean SS = 70.93% and mean SP = 93.12%
Varatharajah *et al.* [100] (2018)	2-hr long interictal iEEG data recorded at 32 kHz from the study group of 82 subjects	Supervised: RBF-SVM classifier with features extracted from a combination of SoZ biomarkers - rates of HFOs, IEDs, and PAC	Acc, SS, SP, and AUC	Acc = 74.22%, SS = 70.36%, SP = 75.09%, and AUC = 0.79
Craley *et al.* [103] (2022)	Preictal and ictal stage scalp EEG recorded from 34 subjects with focal epilepsy	Weakly supervised: End-to-end deep neural network using multi-resolution output for automated feature learning at the electrode-level and global-level.	Accuracy	Acc = 71.1% (24.2 ± 1.0 of 34 subjects)
Miao *et al.*[181] (2022)	1-hr interictal ECoG data from 7 subjects recorded at 2 kHz	Supervised: RBF-SVM classifier with PAC features quantified by mean vector length modulation index	AUC	AUC = 0.915

Method (year)	**Dataset**	**Algorithm type and description**	**Measure of success**	**Performance metric**
Proposed method	ECoG and SEEG recordings during seizure onset from 27 subjects with F_s ranging from 500 Hz to 1024 Hz	Supervised: RBF-SVM classifier with graph centrality measures extracted from the FD-CCM EC measure	Acc, SS, SP, and AUC	**Acc = 94.1%, SS = 92.3%, SP = 94.25%, and AUC = 0.93**

3.5 Conclusion

This study demonstrates the effectiveness of graph centrality measures derived from model-free EC measures for classifying brain regions as SoZ or non-SoZ. Among the three EC measures examined, FD-CCM exhibited superior classification performance, minimized bias in the trained model, and revealed higher mean centrality measures of SoZ electrodes relative to non-SoZ electrodes during ictogenesis. Furthermore, the analysis of trained models on non-seizure-free subjects revealed a significantly lower overlap between the algorithm-identified SoZ with that identified by experts. These outcomes emphasize the importance of delving into directional brain connectivity within the frequency domain for epilepsy analysis using data-driven automated algorithms to address neurological disorders arising from disruptions in brain network connectivity.

Nevertheless, it is crucial to acknowledge the limitations of this analysis. The study is confined to a smaller cohort of 27 subjects, necessitating further validation on a larger dataset to establish the statistical significance of the proposed model for seizure localization. Additionally, the EC measures utilized in this study were derived from ictal iEEG recordings. Given the challenge of consistently recording during seizures, further exploration is warranted to extend these findings to feature extraction during the interictal phase, providing insights into the normal brain function of epileptic patients when not experiencing seizures.

Chapter 4

SoZ Identification from Interictal Intracranial Electroencephalogram

4.1 Chapter Overview

The previous chapter established that supervised machine learning models, when combined with topological graph features, can successfully identify the SoZ. That analysis compared three different effective connectivity measures and identified that graph centrality features derived from *frequency-domain convergent cross-mapping (FD-CCM)* were the most significant, achieving a mean AUC of 0.93 and outperforming prior literature.

However, that foundational analysis was performed on ictal data. For a localization tool to be maximally effective and convenient for pre-surgical evaluation, it must be validated on the more challenging and readily available **interictal** recordings, which represent the brain's baseline pathological state between seizures. This chapter, therefore, extends the analysis of FD-CCM-derived graph centrality measures to the interictal phase.

Furthermore, we recognize that traditional SoZ biomarkers, such as interictal epileptiform discharges (IEDs), high-frequency oscillations (HFOs), relative band power (RBP), coherence, and phase-amplitude coupling (PAC), often fail to fully capture the complex network dynamics underlying seizure generation, but they still provide valuable and complementary information.

Therefore, the objectives of this chapter are twofold. First, to analyze the effectiveness

of our FD-CCM-derived graph centrality measures as a standalone biomarker in interictal data. Second, to develop and validate a hybrid classifier model that integrates these graph-based centrality features with the existing biomarkers from the literature (IED, HFO, RBP, PAC, etc.).

The content of this chapter is adapted from [182]. The subsequent sections will demonstrate that in the interictal state, SoZ regions exhibit significantly lower centrality values compared to non-SoZ areas. We demonstrate that a combined approach, utilizing features ranked by the Minimum Redundancy Maximum Relevance (mRMR) algorithm, achieves a mean performance of 89% accuracy and 82% sensitivity. This result not only outperforms individual feature sets but, most notably, demonstrates superior sensitivity compared to all prior works using interictal recordings. These findings demonstrate the importance of graph-based centrality measures in capturing interictal dynamics and highlight the value of integrating diverse biomarkers for improved discriminability and generalizability.

4.2 Dataset Description

This interictal analysis utilizes the same primary dataset [157] detailed in Chapter 3. Feature extraction and classifier models were developed using five-minute interictal recordings. All epochs selected for this study were confirmed to represent periods of awake brain activity, as validated by a custom non-REM sleep detector applied to daytime recordings [183].

A secondary dataset is used for validating the results. This publicly available dataset [184] consists of de-identified interictal iEEG (SEEG) recordings collected by three centers - The National Institute of Health (NIH), Johns Hopkins Hospital (JHH), and the University of Miami, Florida, Jackson Memorial Hospital (UMF) [185, 186]. The collection of this dataset was approved by the IRB at each clinical institution, with the recording experiments performed in accordance with relevant guidelines and regulations, and all patients provided informed consent prior to enrollment. 17 Engel-1 subjects were identified using the same selection criteria as the primary dataset. The iEEG recordings for both datasets underwent the same preprocessing steps described in Chapter 3.

4.3 Methodology

4.3.1 SoZ biomarkers: Feature extraction

1. **IED spikes** (3 features): The detection of IEDs is performed using a validated algorithm, following the methodology originally established by [187]. In this approach, IEDs are identified based on three key electrophysiological criteria: a peak amplitude exceeding a background-derived threshold, a characteristic sharp-wave component with a duration of 30–200 ms, and a subsequent after-going slow wave lasting 100–500 ms. This method was later adapted by [183] for the specific dataset used in this research. The publicly available detection code and pre-computed IED spike rate feature from the supplementary materials of [183] were directly utilized. In addition to the spike rate, the sharpness and peak amplitude of the IED events detected by the threshold of four standard deviations are also computed and used as features.

2. **Relative band power (RBP) features** (6 features): They describe the relative distribution of EEG signal power across standard frequency bands. The iEEG data is divided into overlapping segments, and Welch's method is applied with a window length of 0.25 seconds, 50% overlap, and a minimum FFT size of 256 points. Power spectral density (PSD) is calculated, and power is computed for six frequency bands: δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), γ (30–80 Hz), and high-frequency activities (HFA) (¿80 Hz). Band-specific power is derived by integrating the PSD over the respective frequency ranges. The band-specific power is divided by the total power across all bands to provide a normalized representation of the spectral characteristics of the EEG.

3. **Statistical features of HFOs** (3 features): HFO events are detected using the validated MNI detector[188] implemented using the PyHFO package [189]. In this method, a baseline of non-oscillatory activity is identified from 125-ms iEEG segments using wavelet entropy. The energy threshold is then set adaptively: if more than 5 seconds of baseline per minute is available, the threshold is the 99.9999th percentile of baseline energy; otherwise, it is set iteratively at the 95th percentile of the signal's energy. Finally, all detected events with a duration greater than 6 ms are selected as HFOs. Given the resampled frequency of 512 Hz, only ripples (80–250 Hz) are

analyzed, as per the Nyquist criterion. The extracted HFO features include the rate of detected events (events per second), the mean amplitude of these events, and the standard deviation of the amplitude, all calculated per channel. These features provide detailed metrics on the occurrence and intensity of HFOs.

4. **PAC measures** (8 features): These quantify the relationship between the phase of low-frequency oscillations and the amplitude of high-frequency oscillations. The EEG data is bandpass filtered into low-frequency bands (δ, θ, α, and β) and high-frequency bands (γ and HFA). The low-frequency phase and high-frequency amplitude are extracted using the Hilbert transform. The high-frequency amplitudes are grouped into eighteen phase bins based on the phase of the low-frequency signal. A phase-amplitude distribution is then generated, and the modulation Index (MI) is computed to quantify coupling strength. The MI is calculated based on the entropy of the amplitude distribution across phase bins, which measures how strongly low-frequency phases influence high-frequency amplitudes [190]. For a more robust PAC estimation, a variable window length tailored to the specific low-frequency band providing the phase is employed. This ensures that the window is sufficiently long to capture multiple cycles of the modulating frequency, which is critical for a reliable phase estimation. In this case, for PAC calculations where the phase-modulating frequency was in the δ, θ, α, and β bands, window lengths of 5 seconds, 1 second, 500 ms, and 250 ms were used, respectively.

5. **Coherence features** (6 features): Measures functional connectivity between the iEEG channels by analyzing frequency-domain relationships. Coherence is computed for each pair of channels using cross-power spectral density, with parameters including a variable window length (same as that used in PAC estimation), 50% overlap, and an FFT size matched to the window length. The coherence values are calculated for predefined frequency bands: δ, θ, α, β, γ, and HFA. For each channel, the 75th percentile of coherence magnitudes across all its connections is extracted for each frequency band, summarizing connectivity strength within that band, as suggested in [183].

6. **Graph-based centrality measures** (10 features): FD-CCM is utilized for the

graph construction method due to its superior performance compared to other data-driven graph measures, as demonstrated in Chapter 3 [156]. Preprocessing includes splitting iEEG recordings into 500 ms windows with 95% overlap, and the frequency bands are evenly partitioned into 40 bands. The directed graphs generated using FD-CCM enable detailed network analysis through ten centrality measures: degree (indegree and outdegree), closeness (in-closeness and out-closeness), betweenness centrality, clustering coefficient, local efficiency, PageRank, and hubs and authorities. These measures capture key characteristics of brain networks, such as identifying critical hubs for information flow, measuring efficiency, and detecting bottlenecks.

4.3.2 Classifier

The classifier employed in this study is a PyTorch-based multi-layer perceptron (MLP) designed for binary classification tasks, focusing on SoZ identification. The architecture consists of sequential linear layers, each followed by a ReLU activation function to introduce non-linearity. The number of layers and their respective sizes are flexible, determined through a grid search over predefined configurations: [(16, 8), (32, 16), (64, 32)]. These configurations allow the model to adapt its capacity to the complexity of the task. The final layer outputs probabilities for the two classes (SoZ and non-SoZ) used for binary classification. Training hyperparameters, such as learning rate ([0.0001, 0.001, 0.01, 0.1]), regularization strength (L2 penalty, [0.0001, 0.001]), and the maximum number of iterations (1000), are also optimized through a grid search to ensure that the classifier is well-suited for capturing the relevant patterns in the data.

Dynamic class weights are computed during training to address the inherent class imbalance in SoZ localization. The class weights are inversely proportional to the number of samples in each class. These weights mitigate the risk of bias. Labels are binarized for simplicity: SoZ and its neighboring regions are labeled positive (1), while non-SoZ regions are labeled negative (0). During performance evaluation, a spatial tolerance mask is applied over a target region that includes immediate neighbors of SoZs. The final sensitivity score is calculated solely based on the successful localization of the ground truth SoZ. The classification of a neighbor as non-SoZ is treated as a neutral event; it is neither penalized nor does it factor into the sensitivity or false negative calculations, which are exclusively

focused on detecting the ground truth SoZ itself. This step further refines the classifier's focus and improves its interpretability in the context of SoZ localization, making it more applicable and reliable.

The classifier's performance is evaluated using a 10-fold group cross-validation (CV) scheme to ensure subject independence across folds. In this scheme, the dataset is split into ten folds, each containing data from a unique set of subjects. This ensures that no samples from the same subject appear in both the training and test sets during a single fold, preventing data leakage and overfitting to subject-specific patterns. Each fold uses 90% of the subjects for training and reserves the remaining 10% for testing, with all subjects included in the test set precisely once over the ten iterations.

Similar to the ictal analysis in Chapter 3, the training process for each fold involves the dynamic calculation of class weights and optimization of model parameters using the grid search-defined hyperparameter space. Once trained, the model is evaluated on the independent test fold, and key classification performance metrics are recorded. The cross-validation process provides a robust estimate of the model's performance by averaging the metrics across all ten folds. The reported mean performance would lie within the limits of the classifier's capability to generalize unseen subjects. This is important in any clinical application where models must be applied to completely new patients.

4.4 Results

4.4.1 Performance Comparison

The performance of features from different biomarker families for SoZ localization is evaluated using sensitivity, specificity, area under the ROC curve (AUC), and area under the precision-recall curve (AU-PRC) metrics. Table 4.1 summarizes the results for univariate biomarkers, bivariate biomarkers, and FDCCM graph-based centrality measures.

The performance of the *univariate biomarkers* (IEDs, RBP, and HFOs) was highly dependent on the recording modality. For instance, HFOs analyzed on ECoG data yielded the group's highest sensitivity at 0.85 ± 0.06, while IEDs on SEEG data achieved the highest specificity at 0.85 ± 0.05. The precision-recall performance for this group was more modest, though HFOs on ECoG data still achieved the best lift of 3.25 ± 0.36 times the baseline.

Table 4.1: Performance Evaluation of Individual SoZ Biomarker Features: Number of features used in each biomarker set, sensitivity, specificity, accuracy, area under the ROC curve (AUC), area under the precision-recall curve (AU-PRC), and the AU-PRC lift, which measures how many times model performance exceeds the random baseline. All performance metrics are reported as mean ± standard deviation.

SoZ Biomarker	**# feats.**	**Modality**	**SS (%)**	**SP (%)**	**ACC (%)**	**AU-ROC**	**AU-PRC**	**AU-PRC: Lift over baseline**
IED spikes	3	SEEG only	0.50 ± 0.08	0.85 ± 0.05	0.68 ± 0.04	0.70 ± 0.07	0.21 ± 0.02	2.53 ± 0.24
		ECoG only	0.62 ± 0.07	0.80 ± 0.06	0.72 ± 0.05	0.75 ± 0.06	0.26 ± 0.03	3.14 ± 0.36
		Combined	0.56 ± 0.06	0.74 ± 0.07	0.73 ± 0.05	0.65 ± 0.06	0.20 ± 0.02	2.41 ± 0.24
RBP	6	SEEG only	0.58 ± 0.06	0.82 ± 0.05	0.72 ± 0.04	0.68 ± 0.05	0.22 ± 0.02	2.65 ± 0.24
		ECoG only	0.68 ± 0.05	0.78 ± 0.04	0.77 ± 0.03	0.73 ± 0.05	0.27 ± 0.03	3.25 ± 0.36
		Combined	0.63 ± 0.07	0.77 ± 0.06	0.76 ± 0.04	0.70 ± 0.06	0.24 ± 0.02	2.89 ± 0.24
Coherence	6	SEEG only	0.55 ± 0.07	0.83 ± 0.06	0.75 ± 0.05	0.66 ± 0.06	0.21 ± 0.02	2.53 ± 0.24
		ECoG only	0.60 ± 0.06	0.79 ± 0.07	0.78 ± 0.04	0.72 ± 0.05	0.24 ± 0.02	2.89 ± 0.24
		Combined	0.58 ± 0.06	0.80 ± 0.05	0.78 ± 0.05	0.69 ± 0.06	0.23 ± 0.02	2.77 ± 0.24
HFO	3	SEEG only	0.70 ± 0.07	0.60 ± 0.06	0.65 ± 0.05	0.62 ± 0.06	0.19 ± 0.02	2.29 ± 0.24
		ECoG only	0.85 ± 0.06	0.75 ± 0.05	0.74 ± 0.05	0.70 ± 0.07	0.27 ± 0.03	3.25 ± 0.36
		Combined	0.77 ± 0.07	0.68 ± 0.06	0.69 ± 0.05	0.66 ± 0.06	0.24 ± 0.02	2.89 ± 0.24
PAC (MI)	8	SEEG only	0.70 ± 0.06	0.75 ± 0.07	0.75 ± 0.04	0.74 ± 0.05	0.26 ± 0.03	3.13 ± 0.36
		ECoG only	0.78 ± 0.05	0.85 ± 0.04	0.83 ± 0.03	0.84 ± 0.04	0.32 ± 0.04	3.86 ± 0.48
		Combined	0.74 ± 0.06	0.80 ± 0.05	0.80 ± 0.04	0.79 ± 0.05	0.29 ± 0.03	3.49 ± 0.36
FD-CCM-based graph	10	SEEG only	0.75 ± 0.06	0.80 ± 0.07	0.81 ± 0.04	0.78 ± 0.05	0.29 ± 0.03	3.49 ± 0.36
		ECoG only	0.83 ± 0.04	0.86 ± 0.03	0.85 ± 0.03	0.84 ± 0.04	0.34 ± 0.04	4.10 ± 0.48
		Combined	0.79 ± 0.05	0.83 ± 0.04	0.83 ± 0.04	0.81 ± 0.05	0.30 ± 0.03	3.61 ± 0.36

For the *bivariate biomarkers*, PAC demonstrated strong potential, achieving the highest AU-ROC of any single biomarker (0.84 ± 0.04) when applied to ECoG recordings. This strong performance was mirrored in its AU-PRC, which reached a peak lift of 3.86 ± 0.48 times the baseline, also on ECoG data. In contrast, the performance of coherence-based features showed comparatively lower performance.

The *graph-based centrality measures* derived from FD-CCM consistently provided the most robust and balanced results. When modalities were combined, the FD-CCM graph provided a strong balance of metrics, including an AU-ROC of 0.81 ± 0.05 and an AU-PRC lift of 3.61 ± 0.36. The approach was particularly effective on ECoG data, where it produced the study's highest specificity (0.86 ± 0.03), accuracy (0.85 ± 0.03), and AU-PRC lift (4.10 ± 0.48 times the baseline).

4.4.2 Feature Ranking using mRMR

The mRMR (minimum Redundancy Maximum Relevance) algorithm [178] is employed to evaluate the contribution of different biomarkers to SoZ identification. The mRMR scores are aggregated across ten cross-validation folds to ensure robustness and generalizability. The scores, normalized to a scale of 0 to 1, highlight the relative importance of each feature in distinguishing the SoZ.

The mRMR ranking, as depicted in Fig. 4.1, underscores the dominance of graph-based features in the rankings. Indegree, in particular, emerged as the most important feature among all those tested. Other graph-based hubs, clusters, and BC features also ranked highly, highlighting the crucial role of network-level metrics in our study. The RBP metrics of alpha, HFA, and gamma bands stood out among other non-network features, playing a significant role in localizing SoZ. The PAC MI values of alpha-HFA, delta-gamma, and delta-HFA consistently ranked high. The RBP of HFA appeared to be more informative in classification than statistical or distribution-based features of HFOs. A closer examination of the rankings reveals that advanced biomarkers consistently outperform simpler ones. In total, 15 of the top 25 features (60%) originated from either the FD-CCM (8 features) or PAC (7 features) families. This pattern is particularly clear in the top 10, which is composed of 4 features from FD-CCM and 4 features from PAC.

It is important to interpret these rankings in the context of the mRMR algorithm.

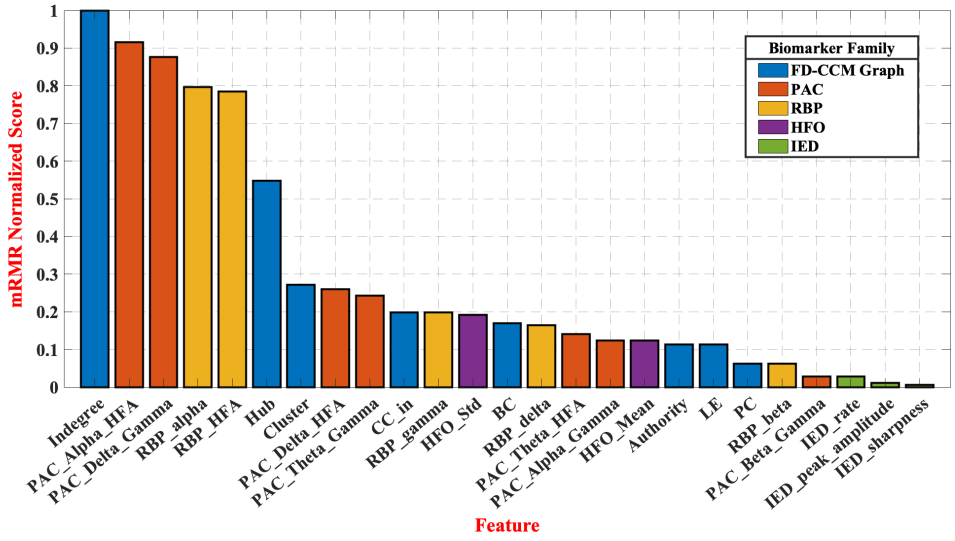

Figure 4.1: Feature ranking for the combined approach: Normalized mRMR Scores of Features (0–1 Scale) Aggregated Across 10 Cross-Validation Folds.

The method ranks features that have the strongest correlation with the clinical outcome (maximum relevance) but simultaneously penalizes features that are highly correlated with already selected, higher-ranking ones (minimum redundancy). Thus, some biomarkers that show good discriminability on their own might be ranked lower or omitted if their predictive information is already captured by a more prominent, less redundant feature.

4.4.3 Classification Using Combined Biomarker Features

Features are added sequentially based on their mRMR rankings to assess the combined performance. The performance metrics (AUC, sensitivity, specificity, and accuracy) are evaluated for each incremental addition. Figure 4.2 illustrates the performance progression as features are incrementally added based on their mRMR ranking for the data using both SEEG and ECoG modalities. The results demonstrated a steady improvement in performance as more features were incorporated, with all metrics continuing to increase up to the inclusion of 19 features. Beyond this point, performance plateaued, indicating redundancy in additional features without improving SoZ localization.

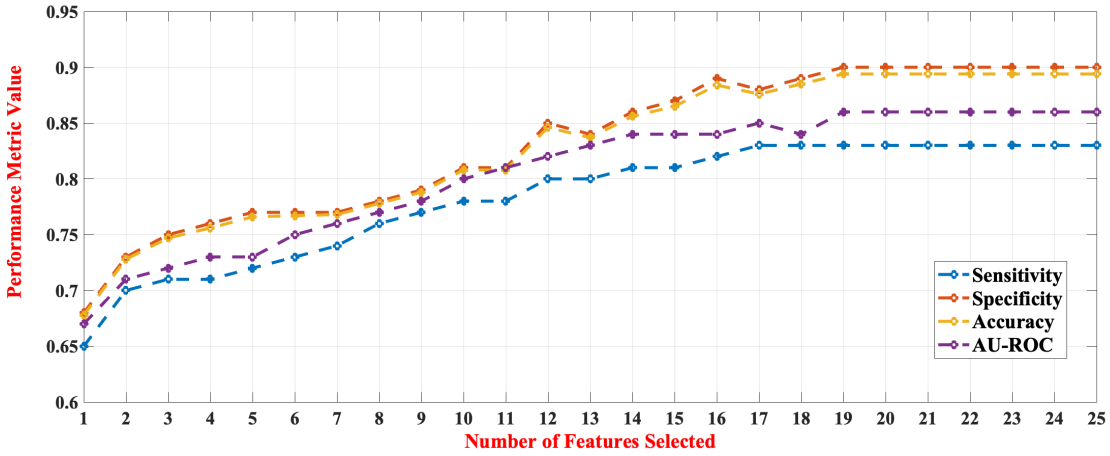

Figure 4.2: Progression of mean performance (AUC, accuracy, sensitivity, and specificity) with the incremental addition of features ranked by the mRMR algorithm.

Table 4.2 summarizes the performance of the hybrid model using the optimal subset of 19 features, trained exclusively on Engel I subjects. The model demonstrated higher and stable performance during cross-validation. When tested on held-out Engel I patients, the combined modality achieved a sensitivity of 0.83±0.04, accuracy of 0.89±0.04, and an AU-ROC of 0.86±0.06. The performance was robust on ECoG recordings, which yielded the highest AU-ROC (0.87±0.06) and AU-PRC Lift (5.90±0.42). As expected, when this pre-trained model was applied to patients with seizure recurrence (Engel II-IV), its performance was reduced and more variable. For the combined modality in this group, the AU-ROC dropped to 0.71±0.14, with a corresponding lift of 3.98±0.58, highlighting the challenge of generalizing from good-outcome to poor-outcome patient data.

4.4.4 Statistical Significance

Fig. 4.3 depicts the absolute t-statistic plot that provides insights into the discriminative power of features extracted from various biomarkers for differentiating SoZ and non-SoZ regions. Higher t-values indicate excellent statistical separability between the two classes, making these features more relevant for SoZ classification. Some of the key observations

Table 4.2: Performance (mean ± standard deviation) of the hybrid model, which was trained exclusively on Engel I subjects

Patient Outcome	**Modality**	**SS (%)**	**SP (%)**	**Acc (%)**	**AU-ROC**	**AU-PRC**	**AU-PRC Lift (×-Baseline)**
Engel I	SEEG only	0.77 ± 0.06	0.86 ± 0.05	0.85 ± 0.05	0.80 ± 0.07	0.41 ± 0.08	4.94 ± 0.36
	ECoG only	0.85 ± 0.05	0.93 ± 0.03	0.92 ± 0.04	0.87 ± 0.06	0.49 ± 0.07	5.90 ± 0.42
	Combined	0.83 ± 0.04	0.90 ± 0.02	0.89 ± 0.04	0.86 ± 0.06	0.48 ± 0.07	5.78 ± 0.23
Engel II-IV	SEEG only	0.69 ± 0.15	0.86 ± 0.10	0.85 ± 0.15	0.64 ± 0.13	0.22 ± 0.10	2.65 ± 0.95
	ECoG only	0.72 ± 0.14	0.86 ± 0.08	0.85 ± 0.15	0.79 ± 0.12	0.34 ± 0.12	4.10 ± 0.76
	Combined	0.72 ± 0.16	0.88 ± 0.09	0.87 ± 0.17	0.71 ± 0.14	0.33 ± 0.09	3.98 ± 0.58

Note: The table presents two distinct evaluations: the top rows show the aggregated cross-validation performance on held-out Engel I patients, while the bottom rows show the generalization performance of the pre-trained model when applied to patients with seizure recurrence (Engel II-IV).

are discussed below. These results emphasize the statistical significance of advanced features such as PAC and graph centralities, which provide more excellent separability than traditional univariate metrics.

1. *High-amplitude HFOs at SoZ*: The mean amplitude of ripples and fast ripples shows higher values in SoZ regions with moderate t-values ($\sim$ 3–4). These features demonstrate localized activity differences, contributing to moderate separability between SoZ and non-SoZ regions.
2. *Low β and γ power at SoZ*: PSD features in the β and γ bands are consistently lower in SoZ regions compared to non-SoZ regions, with moderate t-values ($\sim$ 3–4).
3. *Elevated PAC measures at SoZ*: PAC features, particularly the coupling of high-frequency amplitudes with the phase of δ, θ, and α bands, exhibit higher values in SoZ regions, with strong statistical significance (t-values $\sim$ 3-7). These measures are

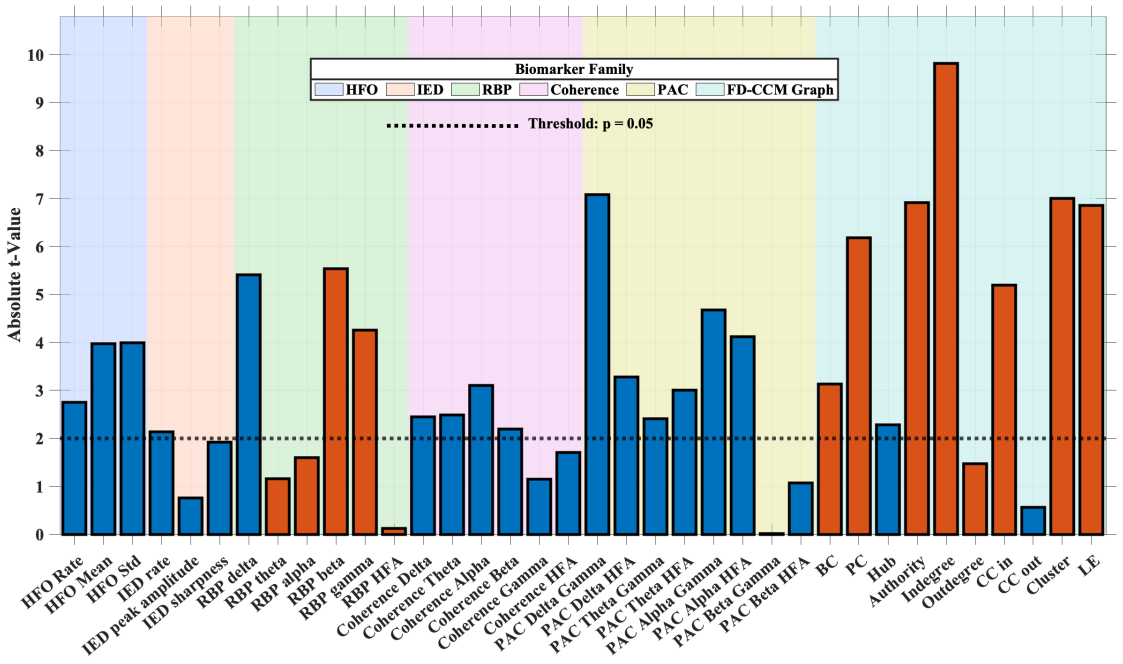

Figure 4.3: Absolute t-statistic plot showing the statistical separability of features between SoZ and non-SoZ regions. Blue bars indicate features with higher mean values in SoZ, while orange bars indicate features elevated in non-SoZ. The dashed line represents the statistical significance threshold ($p = 0.05$).

especially significant in the γ band, highlighting disrupted cross-frequency coupling in epileptogenic regions.

4. *Lower graph centrality measures at SoZ*: Graph-based features, such as indegree, clustering coefficient, and LE, show significantly lower values in SoZ regions, reflecting reduced importance during the interictal phase. These features achieve the highest t-values ($\sim$ 6–9) among all feature sets, demonstrating strong statistical separability and superior discriminative potential.
5. *Coherence and IED features with lower t-values and poor discriminability*: Coherence and IED-related features exhibit t-values below the threshold for significance (t¡2). These features show limited to poor separability and lower discriminative power compared to other feature sets.

6. *Increase in t-value as we progress from univariate features (HFOs, IEDs, RBP) to bivariate PAC and multivariate graph centralities*: The progression from univariate to multivariate features demonstrates increasing t-values and improved discriminability.

4.4.5 Exploring Generalizability and Benchmarking Against Existing Studies

To assess the generalizability of the proposed methods, we validated the results using a second independent dataset. This dataset serves as a benchmark for evaluating how well the features and models trained on one dataset perform when applied to entirely different data, ensuring that the methodology is not overfit to the characteristics of the primary dataset. Table 4.3 summarizes the performance of the 19 features identified earlier in classifying SoZ vs. non-SoZ electrodes. The evaluation includes both internal (within-dataset) and cross-dataset training and testing to assess the generalizability and robustness of the models.

Table 4.3: Mean performance metrics for SoZ vs. non-SoZ classification. The first two cases report mean performance from 10-fold cross-validation within primary and secondary datasets, respectively. For the third and fourth cases, one dataset is used entirely for training and the other for testing to evaluate cross-dataset generalizability.

Train data	**Test data**	**SS**	**SP**	**AU-ROC**	**AU-PRC**	**ACC**
Primary	Primary	0.83	0.90	0.86	0.48	0.89
Secondary	Secondary	0.80	0.88	0.82	0.45	0.86
Secondary	Primary	0.76	0.83	0.78	0.39	0.81
Primary	Secondary	0.79	0.85	0.80	0.41	0.84

Table 4.4: Performance Comparison of the Proposed Method with Prior Works on Machine Learning Algorithms for Identifying SoZ from Interictal iEEG Recordings. Studies utilizing ictal recordings or stimulated seizures are excluded from the comparison.

Method	**Dataset description**	**Algorithm description**	**Performance**
Marsh *et al.* [191] (2010)	30-min iEEG sampled at 200 Hz from 19 children with intractable epilepsy	Spikes in iEEG are identified by two automated detectors through the analysis of amplitude, duration, and waveform features, with thresholds and filters.	Electrodes with the highest mean spike frequency were within the SoZ in 11 of 19 patients (SS = 57.9%)
Geertsema *et al.* [192] (2015)	5-min 65-channel iEEG recorded at 1 kHz from nine individuals with intractable TLE at slow-wave sleep	Autoregressive model residual variation (ARR) features estimated by calculating the coefficient of variation of the r-values obtained through AR modeling.	AUC = 0.82
Varatharajah *et al.* [100] (2018)	2-hr long interictal iEEG recorded at 32 kHz from the study group of 82 subjects	RBF-SVM classifier with features extracted from a combination of SoZ biomarkers - rates of HFOs, IEDs, and PAC	Acc = 74.22%, SS = 70.36%, SP = 75.09%, and AUC = 0.79

Method	Dataset description	Algorithm description	Performance
Jiang *et al.* [193] (2022)	10-min SEEG from 27 subjects with focal epilepsy at 2 kHz or 1 kHz sampling rate	Random forest classifier with the mean within-frequency inward information flow features at each electrode estimated using DTF	Acc = 88%, Precision = 95%, Recall (SS): 76%, and **AUC = 0.94**
Bernabei *et al.* [183] (2022)	60-sec interictal iEEG from 166 DRE patients recorded from multiple centers at 512 Hz or 1024 Hz	Random forest classifier using a combination of power spectral density and coherence features of different frequency bands	AUC = 0.77
Miao *et al.*[181] (2023)	1-hr interictal ECoG data from 7 subjects recorded at 2 kHz	RBF-SVM classifier with PAC features of mean vector length modulation index	AUC = 0.774 ± 0.187
Rekola *et al.* [194] (2024)	12-hr long interictal SEEG from 63 subjects recorded at 512 Hz or 2048 Hz	Logistic regression classifier using features from interictal spike activity score, presence/absence of gamma activity, and focal background slowing activities	AUC = 0.79
Proposed method	5-min iEEG recordings from 27 subjects with F_s ranging from 500 Hz to 1024 Hz	MLP classifier with top-ranked features extracted from different biomarkers - FD-CCM-based graph centrality, HFO, PAC, RBP, IED, and coherence	**Acc = 89%, SS = 83%, SP = 90%**, and AUC = 0.86

To contextualize the proposed method's results, we benchmark our findings against prior studies that have evaluated SoZ localization using interictal iEEG/EEG. These studies provide a comparative baseline to evaluate the relative effectiveness of our approach across various metrics. Table 4.4 summarizes the outcomes of these studies alongside our results, emphasizing key differences in methodologies, feature sets, and datasets. The proposed method surpassed most prior works across nearly all metrics. Although one method achieved a higher AUC [193], it exhibited lower sensitivity in identifying SoZ and matched the accuracy of the proposed method.

Discussion

Key findings

The study emphasizes the importance of network-level features in capturing the connectivity dynamics of the epileptic brain. Although univariate biomarkers such as IED spikes, HFO rate, and RBP are well-established indicators of epileptogenic activity, their standalone performance is limited, with moderate to low t-values. Their limitedness in depicting the complexity of epileptogenic activity emphasizes the exploration of other effective biomarkers through further research. The bivariate characteristics of PAC considerably improved and manifested changes in cross-frequency coupling characteristics of epileptogenic regions. The FD-CCM-based graph centrality measures (e.g., Indegree, LE, and Clustering Coefficient) exhibited the highest t-values, providing robust separation between SoZ and non-SoZ regions, and outperforming features extracted from other biomarkers with a mean classification AUC of 0.8.

This study also highlights the importance of integrating diverse feature sets to improve SoZ identification. The combined approach achieved superior performance by leveraging features from different biomarkers, with a mean AUC of 0.86, mean sensitivity of 83%, specificity of 90%, and accuracy of 89%. This integration outperformed models relying on individual feature sets, including the top-ranked graph-based feature (Indegree) and PAC feature (PAC Alpha-Gamma). The combined feature set surpassed existing methods in sensitivity, outperforming the previous state-of-the-art by 5%. While one prior work achieved a higher AUC [193] than the proposed method, it did so at the cost of reduced

sensitivity despite matching the proposed method's accuracy. Furthermore, the combined feature outperformed a comparable approach that integrated biomarkers [100], increasing accuracy and specificity by about 14% each, sensitivity by over 10%, and AUC by 0.06. These results suggest that combining network-level connectivity measures with local signal dynamics provides complementary insights and significantly enhances the model's capability to identify SoZs accurately. Finally, the proposed method demonstrates generalizability through its strong performance across two independent datasets, as evidenced by both internal and cross-dataset evaluations of accuracy, sensitivity, and specificity.

Graph centralities as a potential biomarker for SoZ and seizure progression

Our analysis of the interictal state revealed that multiple FD-CCM-based graph centralities exhibited significantly lower values in seizure onset zone (SoZ) regions when compared to non-SoZ regions. This finding points to a state of diminished network influence or 'quiescence' from these epileptogenic regions between seizures. Interestingly, the previous chapter has shown that the same graph centrality measures are elevated in SoZ regions during ictal periods [156], exhibiting an increased connectivity and hyper-synchronization during seizures. This dynamic behavior of interictal quiescence followed by ictal activation suggests that these metrics could be a potential biomarker for studying seizure progression in the future. By tracking the temporal evolution of these centrality metrics over more extended periods, it may be possible to monitor how the epileptic network reorganizes and evolves, offering insights into the progression from interictal to ictal phase.

It is important to note that there are different ways to generate brain graphs, including undirected measures of functional connectivity and directed measures of effective connectivity. Even within effective connectivity, various methods exist that operate in either the time or frequency domain. This study uses FD-CCM-based graphs to capture the causal interactions between brain regions in the frequency domain. Thus, the resulting centrality measures should be interpreted within this context as indicators of a region's influence and importance in frequency-specific neural communication.

Limitations and Future Directions

This study employed specific preprocessing steps that may limit the generalizability of the results. Specifically, our choice of referencing for SEEG data, as discussed earlier, presents a limitation. This necessary difference in referencing may have also contributed to the performance disparities observed between the two modalities. Future studies could validate these findings on SEEG datasets where a complete electrode mapping is available, allowing for a direct comparison of referencing schemes.

While care was taken to use detection methods that minimize false positives, such as by selecting HFO algorithms that are robust to artifacts arising from sharp transients like IEDs, each of the baseline biomarker extraction processes could be further refined. For example, the HFO analysis could be enhanced by considering isolated HFOs separately from IED-coupled HFOs, as the former are thought to be a more specific marker of epileptogenic tissue. Similarly, IED detection could be improved with more advanced algorithms, and bivariate measures like PAC and coherence could be fine-tuned by exploring a wider range of frequency pairings. Since the primary focus of this paper was to introduce the FD-CCM graph as a potential biomarker for SoZ identification from interictal recordings and demonstrate the complementary nature of other features, fine-tuning these individual biomarkers presents a clear opportunity for future research to improve the hybrid model's overall performance.

Though the feasibility of transfer learning is established, the smaller subject sizes of the datasets (27 and 17, respectively) and the lower cross-dataset performance compared to internal validation metrics highlight the need for additional datasets and external validation with larger, more diverse patient populations to confirm the approach's broader clinical applicability and statistical significance. The current model was trained exclusively on patients with good surgical outcomes (Engel I). As expected, when this model was applied to patients with seizure recurrence (Engel Class II-IV), its performance showed a lower mean and greater variability. This finding highlights the challenge of real-world clinical application, where patient outcomes are not known a priori. Therefore, the most critical future step is to move from this retrospective analysis to a forward-looking evaluation. A prospective study is the gold standard for proving clinical utility. This would involve applying the model to a new cohort of unevaluated pre-surgical candidates to test its utility

as a diagnostic aid alongside the current standard of care.

Future research can extend beyond clinical validation to enhance the model's fundamental capabilities, such as incorporating space-based connectivity, addressing the computational complexity of the advanced features, and focusing on developing optimized algorithms and exploring hardware acceleration to create lightweight frameworks. This would improve scalability and pave the way for real-time clinical applications.

Chapter 5

Patient-Specific Long-Term Seizure Prediction via Multi-Model Classification

5.1 Chapter Overview

The first part of this dissertation (Chapters 3 and 4) successfully developed and validated a generalizable, network-based framework for localizing the SoZs. Having addressed the critical spatial challenge of "where" seizures originate, this second part of the thesis pivots to the temporal challenge of "when" they will occur. As established in Chapter 1, reliable, long-term seizure prediction remains a crucial unmet need for patients with DRE, with the potential to enable advisory systems and closed-loop therapeutic devices.

As outlined in the literature review in Chapter 2, a primary hurdle that has long plagued the field of seizure prediction is the failure of most existing approaches to account for intra-subject variability. Most prediction models rely on cohort-based data or, if patient-specific, assume a single, monolithic model suffices for all of a patient's seizures. This overlooks the clinical and electrophysiological reality that a single individual can have multiple seizure types, each with its own distinct preictal pattern. This failure to model intra-subject heterogeneity is a key reason for the high false-positive and false-negative rates seen in prior work.

This chapter aims to overcome these limitations by introducing a **subject-specific seizure prediction framework that explicitly models this intra-subject heterogeneity**. The objective is to identify and cluster seizure-specific preictal patterns from long-term (one to two weeks) iEEG recordings. To achieve this, we first extract absolute, relative, and ratio power spectral density (PSD) features from twelve frequency bands. We then apply the Minimum Uncertainty and Sample Elimination (MUSE) algorithm for unsupervised feature selection to form seizure-specific feature sets. Based on the similarity of these features, seizures are grouped into distinct clusters, and a separate classifier is trained for each identified cluster. The final predictions from this ensemble of models are then combined using a grid-optimized k-of-N voting strategy to enhance reliability.

The content of this chapter is adapted from [195]. The following sections will demonstrate the power of this multi-model approach. When evaluated on long-term recordings from ten patients, the clustering framework is shown to significantly improve mean sensitivity from 89.17% to 98.54% while simultaneously reducing the mean false positive rate from 1.15 to 0.62 per day. These findings highlight the critical value of modeling seizure diversity within individuals and support the development of more personalized and interpretable seizure forecasting systems.

5.2 Materials and Methods

5.2.1 Data Description and Preprocessing

This study retrospectively analyzes long-term iEEG recordings collected over one to two week duration from ten patients with medically refractory epilepsy who underwent presurgical monitoring at the University of Minnesota Epilepsy Clinic. Data acquisition was approved by the University of Minnesota Institutional Review Board (Study Number: 1106M01362, PI: Henry, Approval Date: April 2014). The study was conducted in accordance with the Declaration of Helsinki, and all subjects provided informed written consent prior to participation.

For each subject, multiple 24-hour *sessions* were recorded continuously at a 2 kHz sampling rate over one to two weeks. A total of 87 days (2088 hours) of iEEG with 80 seizures was recorded across the ten subjects, where the number of seizures for each

subject varied from 3 to 17 (mean: 8.0, median: 6.5, standard deviation: 3.69). Given the unpredictable and sparse nature of seizure occurrence, some sessions contained multiple seizures, while others contained none. Seizure episodes were manually annotated by clinical experts using visual analysis, with metadata including onset and offset times, seizure types, and electrodes involved in seizure onset and early propagation. These clinically validated annotations served as ground truth for algorithm development and performance evaluation.

Table 5.1 provides a comprehensive summary of clinical and recording characteristics for ten patients with drug-resistant focal epilepsy who underwent invasive EEG at the University of Minnesota Epilepsy Clinic for presurgical evaluation. For each subject, it includes age, gender, days of recording, electrode coverage across SoZ and early propagation (EP) regions, clinical syndrome, etiology, surgical intervention, and Engel outcome classification. The *Engel outcome scale* is a standardized classification used to evaluate seizure outcomes following epilepsy surgery [196].

To focus the analysis on relevant neural activity, features are extracted only from a subset of electrodes corresponding to seizure-onset and early-propagation electrodes, as determined by the neurologists. This expert-driven selection reduced the number of analyzed iEEG channels from approximately 150 to a more targeted set of 9–20 per patient, thereby limiting the potential for confounding events from irrelevant brain regions. While this approach leverages clinical expertise, other automated methods for channel selection exist, such as those based on mutual information or feature ranking. Notably, recent research in the related field of seizure detection has shown that reduced EEG montages, whether expert-driven or algorithmic, can maintain high performance with as few as 3–21 electrodes, supporting the viability of a targeted channel selection strategy [197, 198].

Signal preprocessing included two stages of second-order IIR Butterworth filters: a notch filter centered at 60 Hz with a ± 1 Hz bandwidth to attenuate power line noise and its harmonics (e.g., 120 Hz, 180 Hz, etc.), and a high-pass filter with a cutoff frequency of 0.25 Hz to remove baseline drift and low-frequency artifacts.

Table 5.1: Subject Demographics and Clinical Characteristics

SID	Age	Sex	# Szs	Days rec.	Days w/ Sz	# SoZ + EP Elec.	Seizure Types	Clinical Syndrome	Etiological Factors	Surgical Treatment	Engel Outcome
20	21	F	6	7	2	9	6 CPS	mTLE	Right hippocampal sclerosis	Right anterior temporal lobectomy	I.A at 2 years
21	29	M	7	13	5	11	6 CPS and 1 CPS-GTC	mTLE	Severe TBI	Right superior temporal gyrus resection	I.C at 6 years
22	53	M	6	19	5	20	2 CPS, 2 SPS, and 2 CPS-GTC	mTLE	Subtle bilateral hippocampal alteration	Right anterior temporal lobectomy	I.A at 1 year
24	47	M	6	9	3	10	4 CPS and 2 CPS-GTC	mTLE	Severe TBI	Right anterior temporal lobectomy	II.C at 3 years
25	52	M	17	9	6	16	9 CPS, 6 SPS and 2 CPS-GTC	mTLE (unclear lateralization)	None	Responsive Neurostimulation	III.A at 5 years
26	49	M	10	13	4	19	8 CPS and 2 CPS-GTC	mTLE	Right hippocampal sclerosis	Electrode sites resected (RAT/ RMT/ RPT)	III.A at 6 months
27	27	F	11	8	3	14	11 CPS	mTLE	Auto immune limbic encephalitis	Responsive Neurostimulation	III.A at 5 years
28	35	F	8	8	6	12	5 CPS and 3 CPS-GTC	Neo cortical TLE	None	Left temporal-hippocampal resection	III.A at 1 year
29	29	M	3	7	2	12	3 CPS-GTC	mTLE	Right hippocampus smaller than left	Resection of right amygdalo-hippocampus and temporal cortex	I.C at 4 years
32	34	F	6	8	4	18	6 CPS	mTLE	None	Right anterior temporal lobectomy	II.A at 4 years

5.2.2 Data Split Protocol

Each subject's dataset consisted of multiple 24-hour sessions, though not all sessions contained epileptic activity. For most subjects, the majority of sessions were non-seizure days. To ensure robust model evaluation, the data split is constructed to include representative samples from both seizure (positive) and non-seizure (negative) classes. If the test set lacked any seizure activity, sensitivity could not be computed; hence, it is ensured that each test fold included at least one session with recorded seizures.

A subject-specific k-fold cross-validation protocol is implemented, where k equals the number of sessions containing at least one seizure (denoted as N_{SZ}). Here, $k = N_{\mathrm{SZ}}$. Let N be the total number of sessions for a given subject:

$$N = N_{\mathrm{SZ}} + N_{\mathrm{NSZ}},$$

where N_{NSZ} is the number of sessions without seizures. The remaining N_{NSZ} non-seizure sessions are then partitioned across the k folds as evenly as possible, with some folds potentially receiving one more non-seizure session than others. In each cross-validation fold, one subset consisting of a seizure session and associated non-seizure sessions is held out as the test set, and the remaining $k-1$ folds are used for model creation and training.

Performance metrics are computed per fold, and the overall mean (and standard deviation) across folds is reported. Each fold consists of a held-out 24-hour session with at least one seizure and may also include multiple interictal-only sessions. The evaluation metrics include sensitivity, which is defined as the percentage of seizures in the test fold correctly predicted, and FPR, which is defined as the number of incorrect preictal alarms per 24 hours. Sensitivity varies depending on the number of seizures in the test fold. For instance, in a fold with only one seizure, a missed detection results in 0% sensitivity, whereas in a fold with four seizures, missing one yields a sensitivity of 75%. While this introduces variability in penalizing missed seizures, it reflects real-world usage scenarios where predictions are evaluated over entire recording sessions rather than isolating individual seizures for cross-validation.

5.2.3 Feature Extraction

In this work, power spectral density (PSD)-based features are utilized, as they are commonly employed in epileptic classification tasks [96, 97, 199, 133, 200]. The analysis is restricted to PSD-based features to isolate the effect of clustering on model performance, intentionally avoiding multimodal or temporally dynamic features. This design choice enables a controlled proof-of-concept evaluation, with future work aimed at extending the framework to richer, time-aware, and connectivity-based feature spaces. Twelve frequency bands are considered for feature extraction: θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), γ_1 (30–55 Hz), γ_2 (55–80 Hz), γ_3 (80–105 Hz), γ_4 (105–130 Hz), γ_5 (130–150 Hz), γ_6 (150–170 Hz), γ_7 (170–200 Hz), γ_8 (200–250 Hz), and γ_9 (250–300 Hz).

The gamma band is subdivided into multiple sub-bands to enhance sensitivity to diverse and localized high-frequency oscillations, such as ripples and fast ripples, which are strongly associated with seizure onset [200]. This finer spectral resolution is used to improve discriminability for machine learning models, capture subject-specific biomarkers, and align with neurophysiological evidence suggesting that distinct gamma sub-ranges reflect different pathological processes.

The following PSD features are extracted from each selected electrode (seizure onset zones or early propagation regions) using a sliding 4-second window with 50% overlap:

- *Absolute spectral power* in each of the twelve frequency bands, computed using Welch's overlapped segment averaging estimator (12 features per electrode)
- *Relative spectral power* for each of the twelve frequency bands (12 features per electrode)
- *Ratio of spectral powers* for each unique pair of frequency bands ($\binom{12}{2} = 66$ features per electrode)

Ratio features are included in addition to absolute and relative spectral power features because they have been shown to cancel variance changes in white noise under non-stationary (time-varying statistics), high SNR conditions [201, 202, 203, 204].

5.2.4 Quick Introduction to MUSE feature selection algorithm

Minimum Uncertainty and Sample Elimination (MUSE) is a novel iterative feature selection framework [205] that offers several advantages over the commonly used Minimum Redundancy Maximum Relevance (mRMR) algorithm [178] in large dataset applications.

Prior work that introduced MUSE has shown an average increase of 5.80% in sensitivity and 9.96% in specificity compared to mRMR, highlighting its effectiveness across various classifiers. Additionally, the MUSE algorithm reduces the risk of overfitting by retaining significant feature samples across iterations, thereby enhancing robustness in big data scenarios. Finally, MUSE maintains a constant time complexity per iteration, in contrast to the increasing time cost associated with mRMR, thus ensuring computational efficiency during the feature selection process. A detailed comparison of MUSE with mRMR is discussed in [205].

To explain the MUSE algorithm, it is helpful to consider the data as a matrix, where each *row* is a data sample and each *column* is a feature. The core strategy of MUSE is to iteratively simplify the classification problem. At each step, the most discriminative feature is selected, and the data samples that it classifies with high confidence are eliminated from subsequent iterations. This process forces the algorithm to focus on finding new features that can distinguish the remaining, more challenging data points. The algorithm works in an iterative loop with the following steps:

1. Select the Best Feature: At the start of each iteration, the algorithm searches through all available feature columns to find the single most informative one as the one that best separates the remaining data samples into their classes (i.e., the one with the lowest conditional entropy).

2. Group Samples and Check for Purity: The algorithm takes this selected feature column and groups its values into bins. For each bin, it checks the *purity* by determining if the samples within that bin belong mostly to a single class. A bin is considered *pure* if its impurity score is below a small threshold.

3. Eliminate Data:

    - Remove Rows: Any data sample (*row*) that falls into one of these *pure* bins is considered effectively classified. These rows are permanently removed from the data matrix for all future iterations.
    - Remove the Column: The feature (*column*) that was just used for this step is also removed from the matrix, so it will not be considered again.

4. Repeat: The algorithm then loops back to the first step, operating on the new, smaller data matrix (with fewer rows and fewer columns) that *survived* the elimination. This process is continued, each time finding the next best feature to classify the remaining, more challenging samples.

5.2.5 MUSE-Based Seizure Model Creation

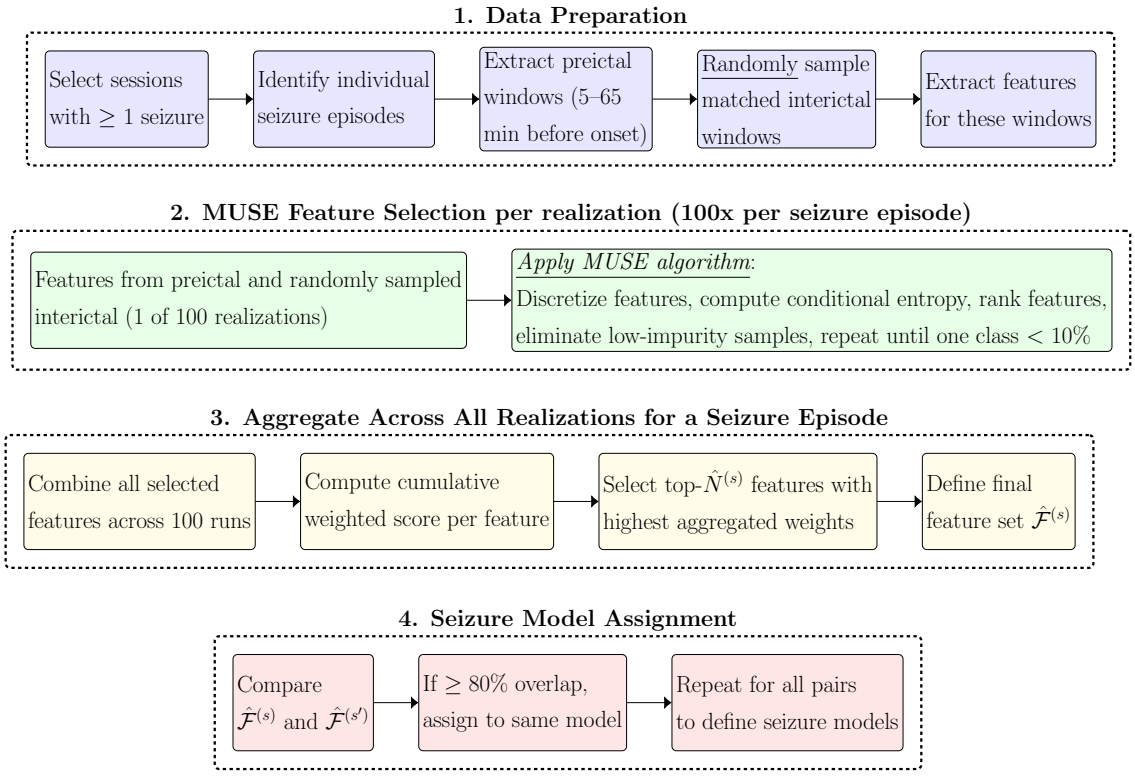

Figure 5.1: *Workflow for seizure model creation from training data using MUSE*: The pipeline consists of four stages—data preparation, MUSE-based feature selection for each realization, aggregating ranked features across iterations, and clustering seizure episodes into models. This approach enables the identification of seizure-specific preictal patterns used for personalized seizure prediction.

In this work, the MUSE algorithm is adopted to capture seizure-specific preictal patterns. The full pipeline is illustrated in Figure 5.1, which outlines each stage from data preparation to model assignment. Detailed pseudocodes for the different algorithms are included in Appendix B.1. The process begins by selecting all training sessions that contain at least one seizure event. For each seizure episode, the corresponding preictal window, defined as the 60-minute interval from 65 minutes to 5 minutes before seizure onset, is identified. A fixed 60-minute preictal window is used to maintain consistency across seizures. An equally sized set of continuous interictal windows is randomly drawn from non-seizure periods within the same session to ensure class balance. Input features are then extracted from these windows.

Step 1: MUSE-Based Feature Ranking Per Realization

Let s index a seizure episode. For each seizure s, feature selection is performed independently across $R = 100$ realizations, where each realization $i \in \{1, 2, \ldots, R\}$ used a random subset of interictal windows combined with fixed preictal data to construct a binary classification problem.

In each realization, the MUSE algorithm is applied to the combined feature matrix of the preictal features and randomly sampled matched interictal features. MUSE performs greedy forward feature selection by iteratively selecting the feature that most improves class separation (e.g., through the reduction in the overlap of class-conditional distributions).

This results in a declining number of samples that survive after each feature is added. The selection process is terminated once the number of surviving samples in either class falls below a predefined threshold—set here at 10% of the original sample count. Only the survival curve of the class that first reaches this 10% threshold is used to quantify the reduction induced by each selected feature. This class is referred to as the *dominant or early-stopping class* for this realization.

Figure 5.2 illustrates a survival curve showing the percentage of samples remaining after each feature is added, separately for each class. The steepness of the drop at each step indicates the discriminative value of the selected feature. For example, in the curve shown, the interictal class (class 0) drops from 100% to 72% after the first feature is added, corresponding to a raw reduction of 28 percentage points.

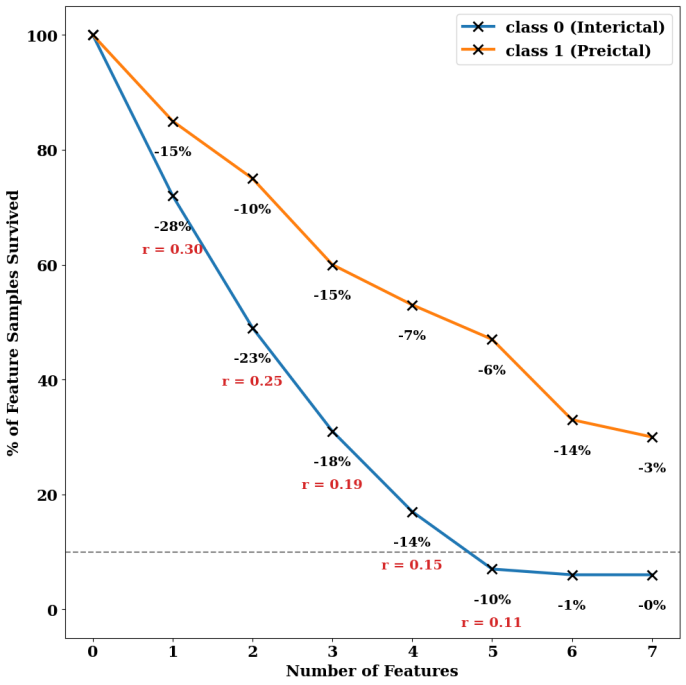

Figure 5.2: Example survival curve illustrating the percentage of training data samples retained after each feature is selected using the MUSE algorithm. Class 0 (interictal) is the dominant or early stopping class in this example, as its survival curve first reached below 10% of samples. Reductions in survival for class 0 (e.g., -28%, -23%, etc.) are used to compute normalized weights for each selected feature.

Let $y_{j,i}^{(s)}$ denote the percentage of surviving samples of the dominant (early-stopping) class after the j-th feature is selected in realization i for seizure s, and $N_{\text{top},i}^{(s)}$ refers to the total number of top-ranked features required to drop the survival rate to below 10%. The reduction attributed to feature $F_{j,i}^{(s)}$ is computed as:

$$\Delta y_{j,i}^{(s)} = y_{j-1,i}^{(s)} - y_{j,i}^{(s)},$$

and the normalized weight is given by:

$$r_{j,i}^{(s)} = \frac{\Delta y_{j,i}^{(s)}}{\sum_{k=1}^{N_{\text{top},i}^{(s)}} \Delta y_{k,i}^{(s)}}.$$

This normalized reduction $r_{j,i}^{(s)} \in [0,1]$ reflects the relative contribution of feature $F_{j,i}^{(s)}$ to class separation and sample pruning.

Let $\mathcal{F}_i^{(s)} = \{F_{1,i}^{(s)}, \ldots, F_{N_{\text{top},i}^{(s)},i}^{(s)}\}$ denote the top-ranked set of features selected in realization i, and $\mathcal{R}_i^{(s)} = \{r_{1,i}^{(s)}, \ldots, r_{N_{\text{top},i}^{(s)},i}^{(s)}\}$ their corresponding normalized weights.

In the example shown in Figure 5.2, the five selected features produce reductions of 28%, 23%, 18%, 14%, and 10%, respectively, which are normalized to yield weights of approximately 0.30, 0.25, 0.19, 0.15, and 0.11. These values reflect the relative sample pruning power of each feature and are used in the aggregation step described next.

Step 2: Aggregating Feature Rankings Across Realizations

Let $\mathcal{U}^{(s)}$ denote the union of all features selected across the R realizations for seizure s. To derive a robust ranking, each unique feature $f \in \mathcal{U}^{(s)}$ is assigned an aggregated weight across all realizations:

$$w^{(s)}(f) = \sum_{i=1}^{R} \sum_{j=1}^{N_{\text{top},i}^{(s)}} \mathbb{I}\left[F_{j,i}^{(s)} = f\right] \cdot r_{j,i}^{(s)},$$

where $\mathbb{I}[\cdot]$ is the indicator function that equals 1 if feature f is selected in realization i at step j, and 0 otherwise. This aggregated score incorporates both the frequency of selection and the strength of contribution, favoring features that consistently induce early and substantial drops in survival rate.

Let $\hat{N}^{(s)} = \max\{N_{\text{top},1}^{(s)}, \ldots, N_{\text{top},R}^{(s)}\}$ denote the maximum number of selected features across realizations. The final representative feature set is then defined as:

$$\hat{\mathcal{F}}^{(s)} = \text{Top-}\hat{N}^{(s)} \text{ features in } \mathcal{U}^{(s)} \text{ ranked by } w^{(s)}(f).$$

Although feature stability is not explicitly enforced, this reduction-weighted aggregation strategy emphasizes robust and highly discriminative features while suppressing those inconsistently selected or redundant across realizations.

Step 3: Clustering Seizures into Models

Seizure episodes are grouped into common models based on the overlap of their representative feature sets. Two seizures s and s' are assigned to the same model if:

$$\frac{|\hat{\mathcal{F}}^{(s)} \cap \hat{\mathcal{F}}^{(s')}|}{\max\left(|\hat{\mathcal{F}}^{(s)}|, |\hat{\mathcal{F}}^{(s')}|\right)} \geq 0.8.$$

An 80% feature overlap threshold is heuristically chosen to ensure high similarity between seizures, such that at least 80% of the features used to distinguish the preictal activity of one seizure are shared with another. The resulting clusters represent distinct algorithm-identified seizure models, potentially reducing false positives when their outputs are aggregated during prediction.

The MUSE-based seizure modeling approach and its application to the patient with Subject ID (SID) 22 are described in detail in B.2.

5.2.6 Training Cluster-Specific Classifiers

For each preictal cluster identified during the model creation stage, a separate binary classifier is trained to distinguish preictal (positive) from interictal (negative) windows. Training is performed using only the subset of features selected as characteristic to the given cluster, i.e., the representative features obtained through the MUSE-based selection and aggregation procedure. This ensures that the classifier is tailored to the specific spectral dynamics reflected in each seizure model.

A random forest (RF) classifier is employed for classification. Each tree is trained on a bootstrap sample of the data using a random subset of the available features, which improves generalization and reduces overfitting. The classifier predicts the label associated with each 4-second EEG window, corresponding to the time segment from which features are extracted. Hyperparameter optimization is performed using grid search over a range of commonly used values, including the number of trees (100, 200, 500), maximum tree depth (None, 5, 10, 20), and the number of features considered at each split (square root or log base 2 of the total number of features). The decision to use an RF classifier is based on a balance of performance, robustness, and the specific goals of this study. Our primary goal is to validate the efficacy of our proposed multi-model, clustering-based pipeline. Compared

to boosting methods like XGBoost that require careful hyperparameter tuning to prevent overfitting, or deep learning models that are highly data-intensive, RF performs reliably with minimal tuning, which is a significant advantage in patient-specific applications with limited data. That said, for future studies involving longer recording durations with a much higher number of seizures across subjects, deep learning architectures may offer advantages in capturing more complex, long-term dependencies.

To address the substantial class imbalance between preictal and interictal segments, the parameter `class_weight` is fixed to `balanced`. This weighting strategy adjusts the importance of each class inversely proportional to its frequency in the training data, ensuring that the minority class (preictal) is not underrepresented during training. Cross-validation within the training set is used to select the optimal hyperparameter configuration, based on validation metrics of sensitivity and FPR.

5.2.7 Aggregating Predictions with Grid-search Optimized *k*-of-*N* Scheme

To improve the reliability of seizure forecasting, the binary outputs from all trained cluster-specific classifiers are aggregated using a k-of-N voting rule, where both k and N are explicitly defined as counts of 4-second windows. Each classifier generated predictions every 2 seconds using a 4-second sliding window with 50% overlap, resulting in 30 predictions per minute.

Outputs from all cluster-specific models are combined via a union rule: if any model labeled a window as preictal, the ensemble label for that window is also preictal. A rolling buffer of the most recent N window-level predictions is maintained. An alarm is triggered if at least k of these N windows are classified as preictal. This voting rule is applied strictly at the window level. To avoid redundant alerts and reduce false positives during prolonged ictal or postictal periods, a *1-hour refractory period* is enforced after each alarm. The common practice for seizure prediction algorithms is to set the refractory period equal to the sum of the seizure prediction horizon (SPH) and seizure occurrence period (SOP) [206, 207]. Given the variable but short SOP duration, the refractory period is chosen to equal the SPH (1 hour).

The k-of-N scheme is optimized through grid search using the training folds. Sensitivity is defined as the proportion of seizure episodes correctly flagged before onset, and the FPR

is computed as the number of false alarms per 24 hours (or per session). The optimal combination of k and N identified in training is then evaluated on the held-out test fold. The grid search is conducted over the following window-level configurations:

$$N \in \{30\ (1\ \text{min}),\ 150\ (5\ \text{min}),\ 300\ (10\ \text{min}),\ 450\ (15\ \text{min}),\ 600\ (20\ \text{min})\}$$

$$k \in \{8\ (0.25\ \text{min}),\ 15\ (0.5\ \text{min}),\ 30\ (1\ \text{min}),\ 60\ (2\ \text{min}),\ 90\ (3\ \text{min}),\ 150\ (5\ \text{min})\},\ k < N$$

This format emphasizes that the voting thresholds are applied over a fixed number of 4-second windows, not over time durations directly. Lower values of k and N typically improved sensitivity but increased false positives, while higher values favored specificity at the cost of missed detections. The best-performing combination is selected based on overall performance across the training folds for all subjects.

5.3 Results

Table 5.2 summarizes seizure characteristics, the number of seizure types per subject, the number of clusters identified by the algorithm, median number of features selected per model, and cross-validated performance metrics (sensitivity and FPR) for both the clustered and non-clustered pipelines using MUSE-based feature selection. The performance using mRMR [178], a commonly used feature selection technique, is also presented for comparison. For the baseline mRMR, the number of top-ranked features is matched to the number of top-ranked features obtained using MUSE feature selection without clustering, i.e., with higher number of features shown in the Table. The final column indicates whether clustering improved sensitivity, reduced FPR, or both. This unified summary supports the observations presented in both subsections below.

5.3.1 Seizure Model Identification

The number of seizure clusters identified per subject ranged from 1 to 4, with higher cluster counts observed in patients with multiple clinically defined seizure types (e.g., SID 25 and SID 28). However, several subjects required more clusters than annotated types, highlighting that additional electrophysiological heterogeneity may exist beyond what is

Table 5.2: Subject-wise summary of cluster modeling complexity and fold-wise prediction performance

SID	# folds	# of szs	# sz types	Engel class	Sensitivity of test folds (%)			FPR of test folds (/day)			# of models (range)	Median # of selected feats.		Cluster effect
					w/o cluster mRMR ranking	w/o cluster MUSE ranking	with MUSE clustering	w/o cluster mRMR ranking	w/o cluster MUSE ranking	with MUSE clustering		w/o cluster	with cluster	
20	2	6	1	I	**100.00 ± 0.00**	**100.0 ± 0.0**	**100.0 ± 0.0**	0.88 ± 0.18	0.88 ± 0.18	**0.58 ± 0.12**	1	**7**	**7**	FPR ↓
21	5	7	2	I	**100.00 ± 0.00**	**100.0 ± 0.0**	**100.0 ± 0.0**	0.90 ± 0.30	0.90 ± 0.32	**0.57 ± 0.44**	1–3	15	**12**	FPR ↓
22	5	6	3	I	**100.00 ± 0.00**	60.0 ± 54.77	**100.0 ± 0.0**	0.61 ± 0.25	0.58 ± 0.21	**0.41 ± 0.19**	3–4	28	**16**	SS ↑ and FPR ↓
24	3	6	2	II	91.67 ± 14.43	91.67 ± 14.43	**100.0 ± 0.0**	1.22 ± 0.51	1.11 ± 0.38	**0.56 ± 0.20**	2–3	22	**15**	SS ↑ and FPR ↓
25	6	17	3	III	87.50 ± 17.08	94.44 ± 13.61	**97.22 ± 6.81**	1.92 ± 0.83	1.92 ± 0.80	**1.25 ± 0.42**	3–4	34	**20**	SS ↑ and FPR ↓
26	4	10	2	III	85.42 ± 14.43	77.08 ± 31.46	**100.0 ± 0.0**	1.27 ± 0.38	1.02 ± 0.24	**0.54 ± 0.16**	2–3	19	**12**	SS ↑ and FPR ↓
27	3	11	1	III	**94.44 ± 9.62**	88.89 ± 19.24	86.11 ± 12.73	1.00 ± 0.00	1.00 ± 0.00	**0.67 ± 0.34**	1–2	7	6	SS ↓ and FPR ↓
28	6	8	2	III	94.45 ± 13.61	**100.0 ± 0.0**	**100.0 ± 0.0**	1.67 ± 0.75	1.50 ± 0.84	**0.42 ± 0.49**	2–3	21	**17**	FPR ↓
29	2	3	1	I	75.00 ± 35.36	75.0 ± 35.36	**100.0 ± 0.0**	0.17 ± 0.24	**0.16 ± 0.23**	**0.16 ± 0.23**	1	8	**6**	SS ↑
32	4	6	1	II	75.00 ± 28.87	**100.0 ± 0.0**	**100.0 ± 0.0**	1.38 ± 0.25	1.38 ± 0.25	**0.62 ± 0.25**	1	**5**	**5**	FPR ↓

Note: The number of folds corresponds to the number of 24-hour sessions containing at least one seizure. Sensitivity and FPR are reported as mean ± standard deviation across folds. *# sz types* refers to the number of clinically identified seizure types. *Median # Features* refers to the median number of features selected across all cross-validation folds for each subject. *Clustering Effect* indicates whether clustering improved sensitivity (SS), reduced false positive rate (FPR), or both.
Bolded entries denote the best-performing case in terms of highest sensitivity, lowest FPR, and fewest selected features.

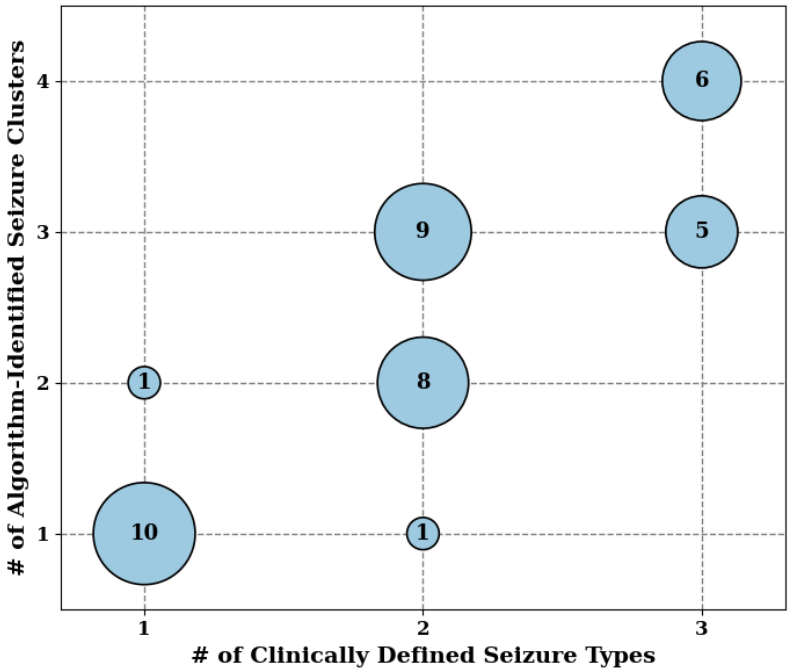

Figure 5.3: Relationship between clinically identified seizure types and algorithm-identified seizure clusters. Circle size reflects the number of subject-fold instances sharing each seizure type–cluster pairing. The plot summarizes model assignments aggregated across all cross-validation folds.

Table 5.3: Overall average performance across all subjects, weighted by number of folds

Test fold performance	**Without cluster mRMR ranking**	**Without cluster MUSE ranking**	**With MUSE preictal clustering**
Sensitivity (%)	91.04 ± 8.39	89.17 ± 13.62	98.54 ± 3.67
FPR(/day)	1.21 ± 0.48	1.15 ± 0.47	0.62 ± 0.29

discernible from seizure type labels alone. This relationship is illustrated in Figure 5.3. In contrast, subjects with a single type of seizure profile (e.g., SID 21, 29, 32) are mostly best represented using a single model.

Table 5.4: Aggregated test performance across all cross-validation folds for each subject, comparing three approaches based on the total number of missed seizures and the overall percentage of seizures predicted: mRMR without clustering, MUSE without clustering, and MUSE with clustering

SID	# sz	# of missed seizures			Overall % of all seizures predicted across all folds		
		w/o cluster mRMR	w/o cluster MUSE	w/ cluster MUSE	w/o cluster mRMR	w/o cluster MUSE	w/ cluster MUSE
20	6	0	0	0	100	100	100
21	7	0	0	0	100	100	100
22	6	0	2	0	100	66.67	100
24	6	1	1	0	83.33	83.33	100
25	17	3	2	1	82.35	88.24	94.12
26	10	2	3	0	80	70	100
27	11	1	2	2	90.91	81.82	81.82
28	8	1	0	0	87.5	100	100
29	3	1	1	0	66.67	66.67	100
32	6	2	0	0	66.67	100	100
Overall	80	11	11	3	86.25	86.25	96.25

To further illustrate how seizure clusters within a subject differ in spectral representation, Figure 5.4 shows a radar plot of the top-ranked features across three clusters for a representative subject-fold (SID 24). Each cluster exhibits a distinct feature profile, with differing reliance on relative band power versus cross-band ratios. Notably, Cluster 1 shows strong emphasis on high gamma and theta power, while Cluster 2 favors ratio-based features such as θ/γ and α/γ. These patterns support the hypothesis that different seizure types or morphologies within a subject manifest distinct preictal dynamics, reinforcing the need for individualized cluster-specific classifiers. While this figure presents a generalized view based on the kinds of features selected, the actual electrodes contributing to each model may differ, reflecting spatial variability in seizure onset and propagation patterns.

It is to be noted that while seizure complexity increases with the variety of clinical seizure types, the models identified by the algorithm do not necessarily correspond to these

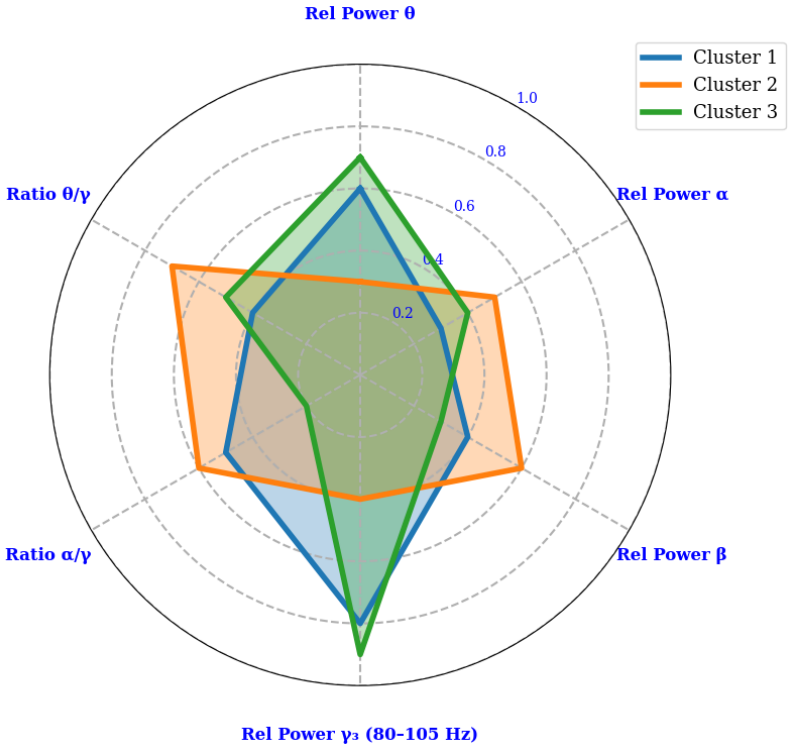

Figure 5.4: Radar plot of major overlapping top-ranked spectral and ratio features across three seizure clusters for a representative subject (SID 24). Each cluster is represented by a separate line, showing how the relative importance of shared features varies across models (contributing electrodes may vary by model).

clinical classifications. This is evident in certain folds where the number of algorithm-derived models exceeds the number of clinically labeled seizure types. Moreover, some models group seizures of different clinical types that originate from similar brain regions (e.g., for SID 24, a CPS and a GTC episode are grouped into the same model). This suggests that the unsupervised model formation is primarily influenced by electrode location and spectral features, leading to the grouping of seizures with shared spatial origins or preictal dynamics. The impact of this clustering-based modeling strategy on prediction accuracy and false alarm reduction is detailed in the following performance analysis.

5.3.2 Prediction Performance Summary

Table 5.3 summarizes the overall performance metrics aggregated across all subjects in Table 5.2, with values weighted by the number of folds per subject to account for variability in test instances. The optimal k-of-N rule selected through grid search is $k = 30$ of $N = 150$ overlapping windows, corresponding to a decision threshold based on 30 of the past 150 consecutive 4-second windows. The two unclustered methods, mRMR-based and MUSE-based feature selection, yield comparable performance in terms of both sensitivity and FPR. In contrast, the MUSE-based clustering approach consistently outperforms both baselines. The remainder of the results section focuses on comparing the MUSE-based clustering approach with the unclustered baseline that also employs MUSE for feature selection. A separate comparison between MUSE and mRMR feature selection methods is presented in Section 5.4.2.

A consistent improvement in FPR is observed with clustering, along with higher or comparable sensitivity, indicating the benefit of cluster-specific modeling. The clustering-based modeling pipeline improved performance in most subjects, particularly those with multiple seizure types or high seizure burden. With clustering, 9 out of 10 subjects achieved improved sensitivity, with eight of them achieving 100%. SID 27, the only exception, achieved 86.11% average sensitivity due to two missed seizures, one of which is identified by the unclustered models. The clustered pipeline also consistently reduced the FPR. For example, SID 25 showed a drop in FPR from 1.92 to 1.25/day, while SID 22 dropped from 0.58 to 0.41/day. Across subjects, clustering also led to a reduction in the number of features per model. For example, the median number of features dropped from 34 to 20 in SID 25 and from 28 to 16 in SID 22. This supports the hypothesis that clustering yields simpler, seizure-specific models that reduce redundancy in spectral-spatial representations. Figure 5.5 illustrates the subject-wise changes in mean sensitivity, mean FPR, and the median number of features required for classification.

To provide a more holistic view of performance not influenced by seizure distribution within the cross-validation splits, an aggregated analysis is performed by combining predictions from all test folds. The results, summarized in Table **??**, reveal a significant performance gap between the methods. The two approaches without clustering (mRMR and MUSE) failed to predict an overall 11 seizures, with the specific number of misses for

each method varying by subject. In stark contrast, the MUSE-based clustering approach proved far more robust. It successfully predicted 96.25% of all seizures within the SPH, missing only 3 of the 80 total seizures analyzed, and achieved perfect 100% sensitivity for 8 of the 10 subjects.

5.4 Discussion

5.4.1 Modeling Seizure Heterogeneity

This study highlights the benefit of modeling intra-subject seizure variability in long-term iEEG recordings. The proposed clustering framework based on our prior MUSE feature ranking algorithm addresses a key limitation of prior subject-specific methods that assume a single preictal representation per individual. By identifying seizure-specific clusters and training specialized models, the framework captures latent electrophysiological diversity that may not be reflected in clinical seizure type labels. In 72.5% of subject-fold instances, the number of seizure clusters equaled or exceeded the number of clinically annotated seizure types, with a clear linear trend observed between the two. This suggests that seizure morphology and preictal dynamics are more diverse than conventional classification captures.

5.4.2 Significance of feature ranking mechanism: mRMR vs. MUSE

While the overall performance of mRMR and MUSE-based feature selection methods is similar, yielding average sensitivities of 91.04 ± 8.39 and 89.17 ± 13.62, and FPRs of 1.21 ± 0.48 and 1.15 ± 0.47, respectively, subject-level differences indicate varying strengths. MUSE achieved higher sensitivity than mRMR in subjects 25, 28, and 32. Subjects 25 and 28 had both a higher number of seizures (17 and 8) and multiple seizure types (3 and 2), along with more seizure-containing days (6 each), suggesting larger within-subject data availability. Subject 32, though having fewer seizures (6) and only one type, also showed improved sensitivity with MUSE (100.0% vs. 75.0%). Conversely, mRMR performed better in subjects 22, 26, and 27, who had fewer seizures (6–11) and a mix of seizure type diversity. In addition to sensitivity differences, MUSE also showed a tendency toward lower FPRs, with a mean FPR of 1.15 ± 0.47 compared to 1.21 ± 0.48 for mRMR. Notably,

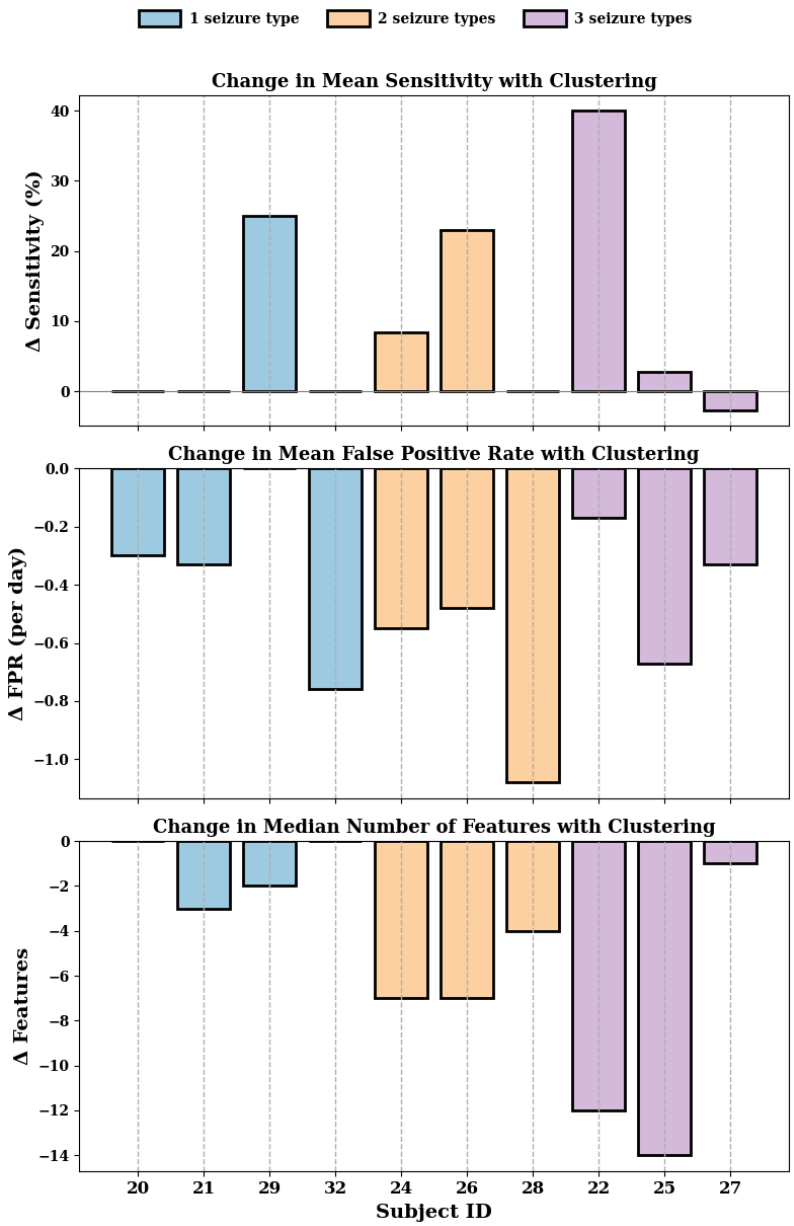

Figure 5.5: Subject-wise changes in model performance and complexity with clustering: The figure illustrates the improvement in mean sensitivity, reduction in mean FPR, and reduction in the median number of selected features for each subject when clustering is applied. Subjects are grouped by the number of distinct seizure types identified by the epileptologist, highlighting greater benefit in patients with clinically observed seizure heterogeneity.

MUSE yielded equal or lower FPR in all subjects. These patterns suggest that while either method may perform better depending on the subject, MUSE shows a tendency to perform well in higher-burden cases when more data are available per subject with greater seizure variability.

5.4.3 Impact of clustering on Prediction Performance

The clustering-based approach preserved or improved sensitivity across most subjects, achieving a mean of 98.5% (SD = 5.3) compared to 89.2% (SD = 25.7) without clustering. The average FPR decreased from 1.15 to 0.62 alarms/day. Notably, subjects with multiple seizure types or greater seizure burden experienced the most significant improvements. For example, SID 25 showed a reduction in FPR from 1.92 to 1.25/day, and SID 22 saw a sensitivity increase from 60% to 100% across folds.

In several folds, clustering enabled the detection of seizures that are missed in the baseline approach (e.g., SID 22, folds 1 and 4). These gains are attributed to the ability of the clustering pipeline to isolate seizures with unique preictal characteristics and train specialized models with smaller, more relevant feature sets. Both pipelines performed comparably in subjects with only one seizure pattern (e.g., SID 20, 21, 29), although clustering offered marginal reductions in false positives.

5.4.4 Model Complexity and Interpretability

Clustering contributed significantly to improving model simplicity and interpretability. The median number of features required per model decreased from 22 to 14, representing a 36.4% reduction in feature dimensionality without compromising classification performance. This reduction supports the notion that clustering enables the creation of seizure-specific models that capture essential biomarkers while discarding redundant information.

Analysis of feature selection frequency across all clusters revealed a consistent dominance of relative spectral power and power ratio features (Fig. 5.6). In particular, the most frequently selected are gamma power (30–300 Hz when grouped) and cross-band ratios involving alpha, theta, and gamma bands. These features likely reflect key neurophysiological changes in excitability and network synchronization that precede seizure onset. High-frequency gamma activity has been linked to local cortical activation, while

power ratios capture shifts in spectral balance across frequency bands; both may vary between seizure subtypes.

The gamma range is decomposed into nine sub-bands spanning 30 to 300 Hz to further investigate the contribution of gamma band features. A comparison between the grouped gamma band and its constituent sub-bands revealed that γ_1 (30–55 Hz), γ_3 (80–105 Hz), and γ_4 (105–130 Hz) are most frequently selected. This suggests that discriminative information is concentrated in specific portions of the gamma spectrum, underscoring the utility of sub-band resolution in improving feature efficiency.

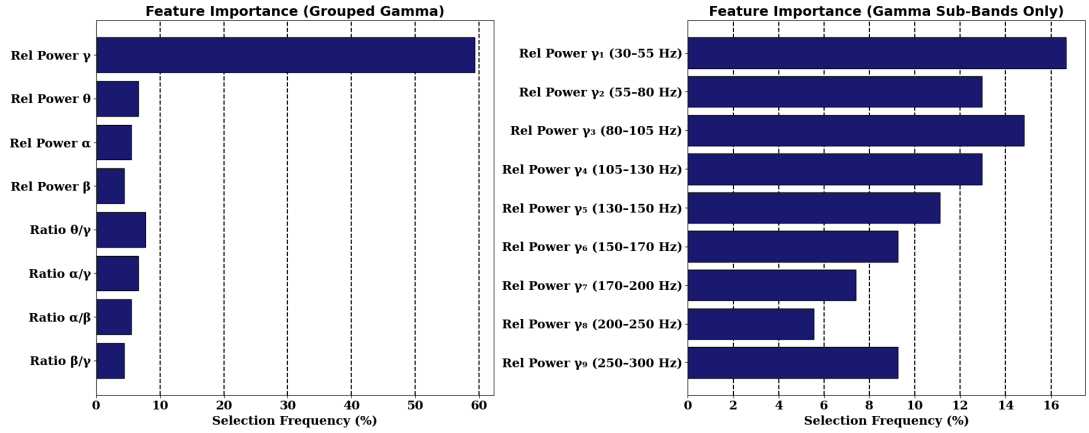

Figure 5.6: Feature selection frequency across seizure models: Left panel shows normalized frequencies with gamma grouped as a single band; right panel shows the distribution across individual gamma sub-bands (γ_1–γ_9), highlighting peaks in selection within low and mid-gamma ranges.

The distribution of top features also varied across clusters within the same subject, suggesting that seizures with distinct morphological or spatial characteristics emphasize different spectral patterns. This finding reinforces the importance of cluster-specific modeling: a single feature set may not generalize across all seizures for a given patient, especially in the presence of electrophysiological heterogeneity.

5.4.5 The Role of the k-of-N Threshold in Reducing False Alarms

An essential component of the prediction pipeline is the k-of-N voting scheme, which ensures the stability and reliability of alarms. Due to the inherent variability of iEEG signals, relying on a single window's prediction can lead to spurious false positives, as transient fluctuations might mimic preictal patterns by chance. A clinically viable system must distinguish these momentary events from a sustained progression into a preictal state.

Table 5.5: Comparison of FPR (mean ± SD) with and without k-of-N for MUSE-based clustered approach

SID	**FPR with k-of-N**	**FPR w/o k-of-N**
20	0.58 ± 0.12	5.42 ± 1.12
21	0.57 ± 0.44	5.91 ± 4.56
22	0.41 ± 0.19	5.32 ± 2.46
24	0.56 ± 0.20	5.47 ± 1.95
25	1.25 ± 0.42	10.81 ± 3.95
26	0.54 ± 0.16	8.02 ± 2.37
27	0.67 ± 0.34	8.56 ± 4.35
28	0.42 ± 0.49	6.77 ± 7.89
29	0.16 ± 0.23	4.03 ± 5.79
32	0.62 ± 0.25	5.38 ± 2.17

The k-of-N rule addresses this challenge by requiring a persistence of preictal classifications over a more extended period before triggering an alarm. The dramatic impact of this mechanism is quantified in Table 5.5. As shown, the window-level predictions without this thresholding result in a very high FPRs, ranging from approximately 4 to 11 alarms per day across subjects. With the k-of-N rule applied, the FPR drops by nearly an order of magnitude to a clinically acceptable level, typically below 1.0 per day. While the MUSE-based clustering helps further reduce the FPR from an unclustered baseline by creating more specific models, the k-of-N rule is the key to avoiding false alarm triggers.

5.4.6 Failure Cases and Missed Seizures

The proposed clustering-based method missed three seizures across all test folds. In SID 25, the missed seizure occurred in a fold where the ictal onset is localized to electrodes LST 1,2 (left superior temporal), a region not represented in the training data, which included seizures originating from LAT (left anterior temporal), RAT (right anterior temporal), and RPT (right posterior temporal) regions. Similarly, in SID 27, one missed seizure had a right temporal onset, while the remaining ten seizures used for training are left temporal, making this event structurally distinct. Another missed seizure in SID 27 exhibited a diffuse onset pattern detected by the non-clustered baseline but missed by the clustered model. This case may reflect a limitation of the clustering approach, that it promotes simpler, seizure-specific models by selecting a minimal subset of features, but may underrepresent more atypical seizure dynamics outside of learned clusters.

Notably, all missed seizures occurred in subjects with Engel Class III outcomes, reflecting poor surgical response despite treatment based on presurgical SoZ localization. This observation raises the possibility that the clinical ground truth used to guide treatment may have been incomplete—i.e., certain seizure onset zones may not have been captured during monitoring. These findings emphasize the importance of modeling approaches that can accommodate seizure variability and motivate future extensions that integrate uncertainty or novelty detection to flag outlier events.

5.4.7 Comparison with Prior Work on Seizure Prediction Using Multi-Day Recordings

Table 6.3 summarizes representative studies on seizure prediction using multi-day EEG recordings. These works span various modalities—including invasive iEEG, subdural, scalp, and sub-scalp recordings—and differ in recording duration, number of seizures, algorithmic complexity, and evaluation criteria.

Early approaches relied on nonlinear dynamical features extracted from limited datasets and achieved moderate performance [108, 109]. Subsequent studies expanded to larger cohorts and longer monitoring durations, such as using multi-month ambulatory iEEG, and subdural iEEG [135, 147]. While these works demonstrated feasibility in real-world conditions, subject-wise variability remained a key limitation. Recent studies have introduced

Table 5.6: Representative seizure prediction studies in humans utilizing multi-day iEEG recordings

Year	**Method**	**Dataset Description**	**Feature and Model Used**	**Performance**
2003	Iasemidis *et al.* [108]	iEEG from 5 patients (28–32 sites, 200 Hz), 55 seizures over 18 days (**short-term**)	Adaptive Seizure Prediction Algorithm (ASPA) using Short-Term Maximum Lyapunov Exponents (STLmax)	SS = 82%, FPR = 0.16/hr (3.84/day)
2005	Iasemidis *et al.* [109]	iEEG from 2 subjects, 25 seizures over 107.5 hrs (**short-term**)	ASPA + T-index tracking of seizure precursors	SS = 81.8%, FPR = 0.15/hr (3.6/day)
2013	Cook *et al.*[135]	15 patients, NeuroVista iEEG (16 channels, 400 Hz) over 4 months (**long-term**)	Beta-band (16–32 Hz) power + k-NN classifier	SS = 65–100%
2016	Brinkmann *et al.*[146]	Crowd-sourced challenge: 230 hrs of iEEG (3 × 8 subdural grid) from 2 patients, sampled at 5 kHz with a total of 46 seizures (canine data used for the contest not described here) (**fragmented long-term**)	Top ranked model: Spectral, entropy, correlation, and statistical features + Lasso-regularized GLM classifier	SS = 71% @ 75% specificity (includes canine data)
2018	Kuhlmann *et al.*[147]	Crowd-sourced seizure prediction challenge: 3 patients, 1-year 16-channel subdural iEEG, 1100+ seizures (f_s=400 Hz) (**long-term**)	Top ranked model: Spectral, AR, fractal, coherence features; SVM, k-NN, GLM, XGBoost	Top ranked model: SS = 58% @ 75% specificity
2025	**Proposed model**	10 subjects, 2088 hrs iEEG (f_s=2 kHz), 80 seizures (**long-term**)	MUSE-based clustering of PSD features + random forest classifier	**Mean SS = 98.54%, FPR = 0.62/day**

deep learning-based models (e.g., LSTM, CNN) to exploit temporal and spatial information in scalp and sub-scalp EEG [113]. While these approaches report promising sensitivity and reduced false alarms, they often require extensive training data, complex architectures, or proprietary acquisition systems.

In contrast, the proposed method applies interpretable PSD features, specifically absolute, relative, and ratio power across twelve frequency bands, and uses unsupervised clustering to capture intra-subject seizure variability. Each cluster is assigned a dedicated classifier to improve model specificity, and predictions are aggregated using an optimized k-of-N voting scheme. Applied to high-resolution (2 kHz) iEEG data from 10 subjects spanning 87 days (2088 hours) and 80 seizures, the method achieves a mean fold sensitivity of 98.54% and a FPR of 0.62/day, outperforming prior studies in both sensitivity and FPR without relying on deep learning. While the proposed approach outperforms all prior approaches, readers are encouraged to consider contextual differences, such as recording duration, number of seizure days, and total seizure count, that contribute to the modeling complexity in each study.

The proposed system is also practical for real-time implementation. The total processing time per window is roughly 370 ms, leaving a significant lead time within the 2-second interval. The test was conducted on a MacBook Air (M1, 2020) with an 8-core CPU, a 7-core GPU, and 16 GB of RAM. The bulk of the computation is spent on feature extraction to get the spectral powers, which may take about 350 ms using an efficient FFT hardware. After that, the actual classification is very fast; the parallel random forest predictions only take 15 ms, and the final aggregation step adds less than 1 ms. This leaves plenty of headroom of over 1.6 seconds, confirming the system is not computationally constrained.

5.4.8 Limitations and Future Work

This study is limited by its retrospective design and a relatively small cohort of ten subjects. Furthermore, this study was conducted using high-resolution iEEG data to establish a proof of concept. Generalizing this framework to scalp EEG represents an important future direction. Still, it would require significant adaptation of the feature set to address the challenges inherent to non-invasive recordings, such as a lower signal-to-noise ratio, muscle artifacts, and reduced spatial resolution. This would likely involve replacing or

modifying the feature with alternatives more robust to these conditions. The core multi-model approach, however, remains a promising strategy for non-invasive applications.

While cross-validation was conducted using seizure-containing sessions, further validation is needed to assess generalization across more extended recording periods, additional patients, and rare seizure types. Our multi-model framework is architecturally designed to address this challenge via a lifelong learning approach. In a prospective setting, events that cannot be confidently classified could be flagged as novelties. After clinical verification, these events could form a new cluster used to train a new *expert* model, which is then added to the classification ensemble, allowing the system to adapt over time without complete retraining.

Although the evaluation is retrospective, each test fold includes at least two 24-hour sessions, and predictions are generated using a rolling k-of-N voting scheme with a 1-hour refractory period, approximating key aspects of real-time deployment. Nonetheless, full streaming simulations across continuous recordings can be explored in future work to more closely reflect prospective performance.

The clustering approach is based on feature overlap to capture seizure-specific preictal dynamics. While orthogonality between cluster-specific classifiers is not assumed, redundancy is expected to be reduced through the use of distinct, entropy-weighted feature sets. Future work will assess inter-model agreement more formally to quantify potential redundancy across clusters. Additionally, incorporating temporal evolution and spatial context from electrode montages may enhance model specificity.

Finally, while binary predictions are used for simplicity, incorporating probabilistic outputs or risk scores from individual models could enable more nuanced, dynamically weighted ensemble decisions. Future work will explore semi-supervised clustering, prospective validation, and integration with adaptive or lifelong learning frameworks.

Chapter 6

Patient-Specific Prediction based on the Novel AMIFD Biomarker

6.1 Chapter Overview

In the previous chapter, we successfully demonstrated that a patient-specific, multi-model framework is a powerful and necessary strategy to address intra-subject seizure heterogeneity. That approach, which used standard PSD features, confirmed that clustering seizure types is critical for improving prediction accuracy, as evidenced by the significant improvement in sensitivity to 98.54% and a 50% reduction in the FPR, confirming the importance of the multi-model framework itself.

This chapter introduces a new, patient-specific prediction pipeline based on a novel biomarker: the **Absolute Mean Instantaneous Frequency Difference (AMIFD)**. As detailed in the methodology chapter, this bivariate feature is designed not to measure average power, but to capture the instantaneous, phase-based frequency dynamics between brain regions. We hypothesize that this measure of phase synchrony is a potent and effective biomarker for characterizing preictal states.

The primary objective of this research is to present and validate this AMIFD-based pipeline. To ensure a fair and robust evaluation that properly accounts for seizure variability (as established in Chapter 5), we will employ the same multi-model classification strategy. The approach uses the MUSE feature ranking technique to automatically cluster

different seizure types based on their top-ranked AMIFD preictal feature signatures. This data-driven grouping enables the creation of an ensemble of expert models—in this case, specialized random forest classifiers—which are then combined using a k-of-N aggregation logic to minimize false alarms.

The effectiveness of the foundational biomarker was confirmed, with AMIFD showing a significantly greater effect size than several standard electrophysiological features ($p < 0.03$). The subsequent sections will show that this AMIFD-driven, multi-model framework achieved a mean sensitivity of 92.08% and a false positive rate of 1.32/day when tested on 10 epileptic subjects. This performance, a substantial improvement over non-clustered baselines, establishes AMIFD as a powerful and viable biomarker for seizure prediction, further solidifying that the patient-specific, multi-model strategy is a reliable pathway toward clinically accurate prediction systems.

6.2 Materials and Methods

6.2.1 Dataset description and Data Split Scheme

The patient cohort and long-term iEEG dataset used to validate this feature pipeline are identical to those introduced in the previous chapter. A complete description of the patient demographics, recording specifications, and preprocessing steps is available in Section 5.2.1. A critical component of this validation is adhering to the same data splitting and evaluation protocol used in Chapter 5 (5.2.2). This ensures a direct and fair comparison between the PSD-based approach and the AMIFD-based approach. We followed the identical cross-validation scheme to segment the data for training and testing, as illustrated in Fig. 6.1.

6.2.2 A Novel Bivariate Measure: Absolute Mean Instantaneous Frequency Difference (AMIFD)

This paper proposes a novel bivariate feature, the *Absolute Mean Instantaneous Frequency Difference* (AMIFD), as a biomarker for seizure prediction. The feature is designed to quantify the level of synchronization between two distinct EEG/iEEG signals by comparing their *mean instantaneous frequencies*. Its core principle is based on the concept of

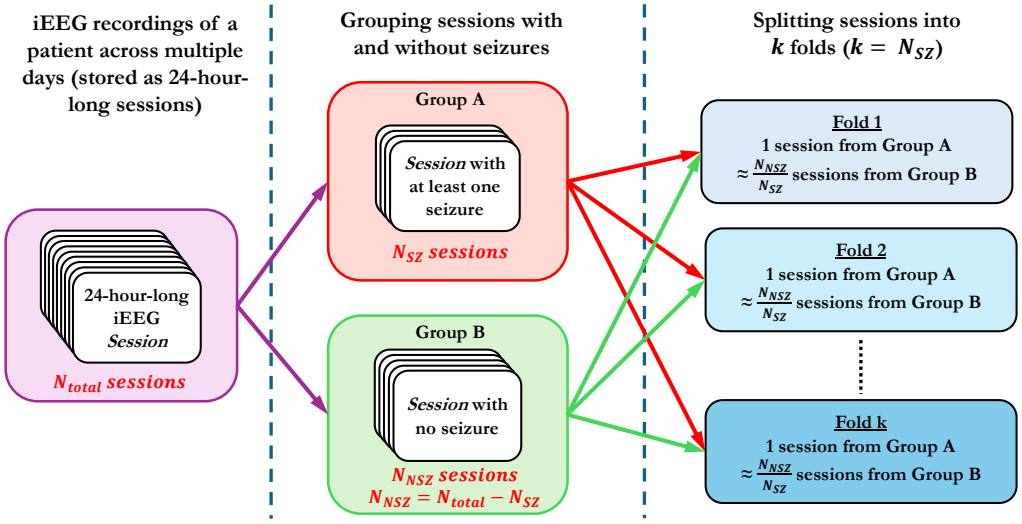

Figure 6.1: Illustration of the subject-specific k-fold data-split scheme. The total number of folds (k) is determined by the number of sessions containing seizures (N_{SZ}). Non-seizure sessions are distributed among the folds. In each iteration, one fold is held out for testing, and the remaining folds are used for training.

phase-locking [208]; if two signals are highly synchronized, their mean instantaneous frequencies will be nearly identical, resulting in a small AMIFD value. Conversely, a larger frequency difference suggests that the signals are less synchronized. The AMIFD is therefore calculated between pairs of iEEG signals from different electrodes to measure the coupling between different brain regions across various frequency bands. By tracking this measure, the algorithm aims to identify the subtle changes in brain state that precede a seizure.

Suppose $s_i(t)$ is a band-limited signal represented by the real-valued iEEG time series from channel i. The analytic form of $s_i(t)$, denoted by $a_i(t)$ is represented by a complex exponential with time-varying instantaneous amplitude ($A_i(t)$) and instantaneous frequency ($\phi_i(t)$):

$$a_i(t) = A_i(t)\ e^{j\phi_i(t)} \tag{6.1}$$

The instantaneous frequency is thus the argument of its complex analytic representation

$a_i(t)$.

$$\phi_i(t) = Arg\{a_i(t)\} \tag{6.2}$$

Within a small analysis window, the instantaneous phase can also be modeled as a linear function of time with an additive stochastic process (φ_i):

$$\phi_i(t) = 2\pi f_i t + \varphi_i \tag{6.3}$$

where f_i is the mean instantaneous frequency of the signal.

Since the slope of the instantaneous phase represents the mean instantaneous frequency, least-squares linear regression is used to estimate f_i for a given signal $s_i(t)$. The estimated value of f_i, denoted by $\hat{f}_i$ from a window size of M is given by:

$$\hat{f}_i = \frac{1}{2\pi} \frac{\sum_{k=1}^{M}(t[k] - \bar{t})(\phi_i[k] - \bar{\phi}_i)}{\sum_{k=1}^{M}(t[k] - \bar{t})^2} \tag{6.4}$$

Here, t is the time vector $[\frac{1}{f_s}, \frac{2}{f_s}, \cdots, \frac{M}{f_s}]^T$ with mean $\bar{t}$ and $t[k] = \frac{k}{f_s}$. ϕ_i is the instantaneous phase vector with mean $\bar{\phi}_i$.

The AMIFD between signals from channels i and j is thus estimated as:

$$AMIFD_{i,j}(t) = |\hat{f}_i - \hat{f}_j| \tag{6.5}$$

6.2.3 Feature extraction and Assigning Class Labels

Features are extracted from the preprocessed iEEG data using 4-second sliding windows with a 50% overlap, which results in a new feature calculation every 2 seconds. The AMIFD features are extracted between every pairs of channels analyzed for the following twelve frequency bands: θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), γ_1 (30–55 Hz), γ_2 (55–80 Hz), γ_3 (80–105 Hz), γ_4 (105–130 Hz), γ_5 (130–150 Hz), γ_6 (150–170 Hz), γ_7 (170–200 Hz), γ_8 (200–250 Hz), and γ_9 (250–300 Hz). Segmenting the gamma band into multiple sub-bands enables more sensitive detection of features from various high-frequency oscillations, such as ripples and fast ripples, which are considered strong indicators of seizure onset [5, 42].

To avoid the high computational cost of applying 12 separate band-pass filters for each of the many channels analyzed, we employ a more efficient method. This technique leverages the Fast Fourier Transform (FFT) to generate the analytic signal for each frequency

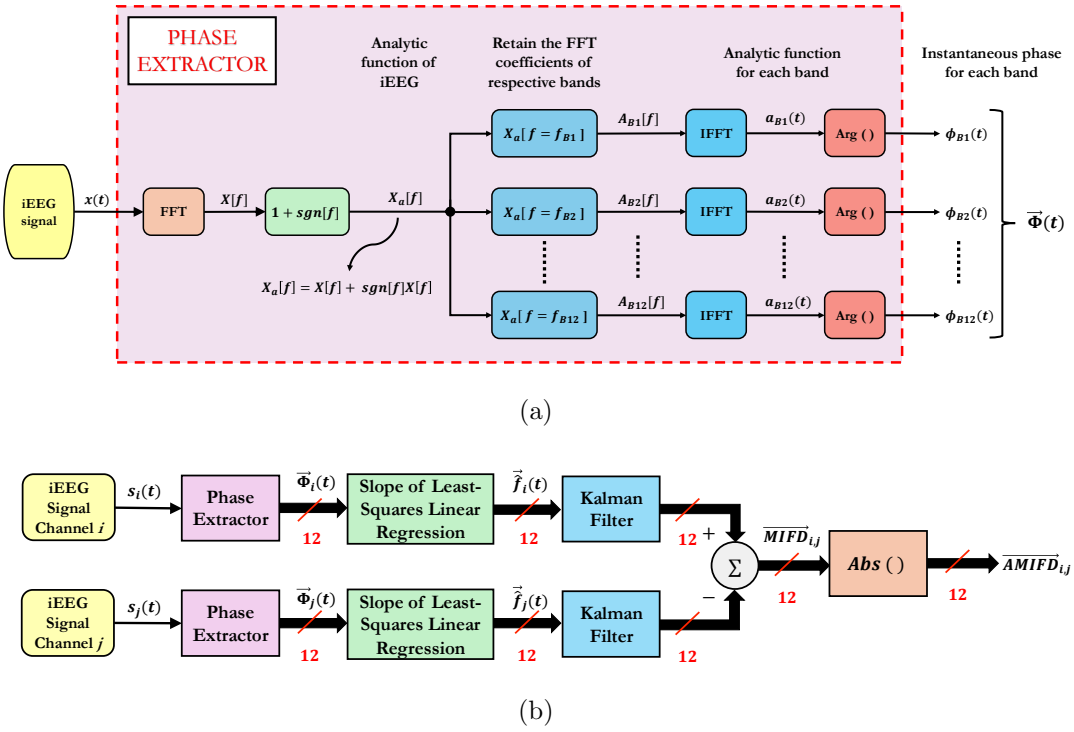

Figure 6.2: Flow charts for (a) the *Phase Extractor* to compute the instantaneous phases for a single channel signal in 12 frequency bands and (b) computing the mean instantaneous frequency (MIF), MIF difference (MIFD), and absolute MIFD (AMIFD) for the 12 frequency bands for signals from two channels.

band. As depicted in Fig. 6.2a, the analytic signal $a(t)$ for a signal band-limited to B is obtained as follows:

$$a_B(t) = \mathcal{F}^{-1}[S[f = B] + sgn(f)S[f = B]] \tag{6.6}$$

where $s(t) \overset{\mathcal{F}}{\longleftrightarrow} S[f]$, $\mathcal{F}^{-1}$ is the inverse Fourier transform, and $sgn(f)$ is the signum function.

The practical implementation of this involves calculating the FFT of the iEEG signal once, retaining the coefficients corresponding to each band of interest (Bi), and then applying the inverse transform. In this case, there are twelve distinct bands, which can be

represented as $B1, B2, \cdots, B12$. This single-FFT strategy is substantially less computationally intensive than an approach based on multiple individual filters. The instantaneous phase for each band, $\phi_{Bi}(t)$, is obtained from the complex argument of the analytic representation $a_{Bi}(t)$, as depicted in Fig. 6.2. The AMIFD estimation from the instantaneous phase follows the same logic described in Section 6.2.2, with a slight modification. A post-processing step, involving the application of a second-order Kalman filter [209], is used to smooth irregular effects in the estimated instantaneous frequencies before computing their difference, thereby reducing noise that can degrade prediction capabilities. The state-space model previously detailed in [210] and in the supplementary materials of [96] is adopted for the Kalman filter.

The preictal period, a 60-minute window from 65 to 5 minutes before a seizure, serves as the Seizure Prediction Horizon (SPH). The subsequent ictal window (5 minutes before to 1 hour after the seizure) is masked off during training. All other data beyond these periods are labeled interictal. A preictal alarm triggered within the SPH constitutes a successful prediction.

6.2.4 Seizure Model Creation and Classification

The core pipeline for seizure model creation and classification remains identical to that of Chapter 5. This process, which is illustrated in Figure 6.3, consists of three main stages:

1. **Unsupervised Feature Selection and Clustering**: The MUSE algorithm is applied on a per-seizure basis to identify the top-ranked AMIFD preictal feature signatures for each seizure. Seizures are then grouped into clusters based on the similarity of these feature sets (refer to Section 5.2.5).
2. **Ensemble Classifier Training**: A separate, specialized classifier (e.g., Random Forest) is independently trained for each identified seizure cluster, creating an ensemble of expert models.
3. **Aggregated Prediction**: The final prediction alarm is triggered using the same k-of-N aggregation and voting logic described in Chapter 5 to enhance reliability and minimize false alarms.

Note that the initial input feature space is high-dimensional. For the proposed method, N electrodes generate $12 \times \binom{N}{2}$ features, resulting in a feature count ranging from 432 to

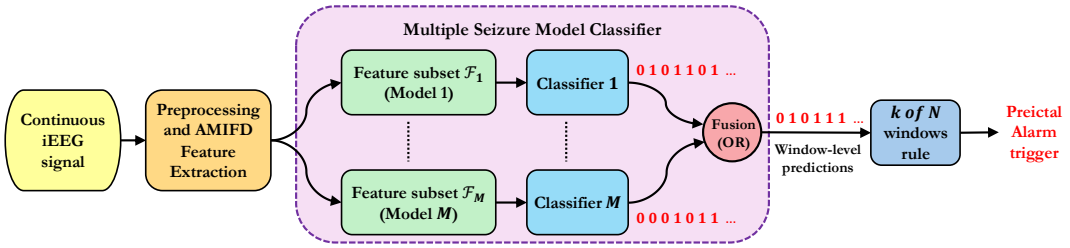

Figure 6.3: The proposed multi-model classification and prediction aggregation pipeline. Continuous iEEG signals are processed to extract features, which are then fed into M cluster-specific classifiers. The window-level predictions are fused and passed through a k-of-N voting rule to trigger a final preictal alarm.

2280 depending on the subject. Since many of these features are not informative, they are discarded during feature selection with the MUSE algorithm. The final number of unique features used will be substantially smaller, as seen in the results.

6.3 Results

6.3.1 Comparative Effect Size of the AMIFD Biomarker

To quantitatively evaluate the ability of the AMIFD biomarker to distinguish between *interictal* and *preictal* states, a comparative analysis was performed against five standard EEG feature types. All feature classes use the same twelve frequency bands. For a robust comparison of the most effective examples of each method, this analysis focused on the top 10 performing features from each category, with performance data aggregated across all subjects. The effect size for these selected features was calculated using Cohen's d [211]. Cohen's d is a standardized measure of effect size that quantifies the magnitude of the difference between two groups in pooled standard deviation units. Larger absolute values of Cohen's d signify a greater degree of separation between the groups, which translates to a higher potential for accurate and reliable classification. The distribution of these values for the top-performing features is presented in Figure 6.4.

The AMIFD biomarker consistently yielded a large effect size, indicating strong class

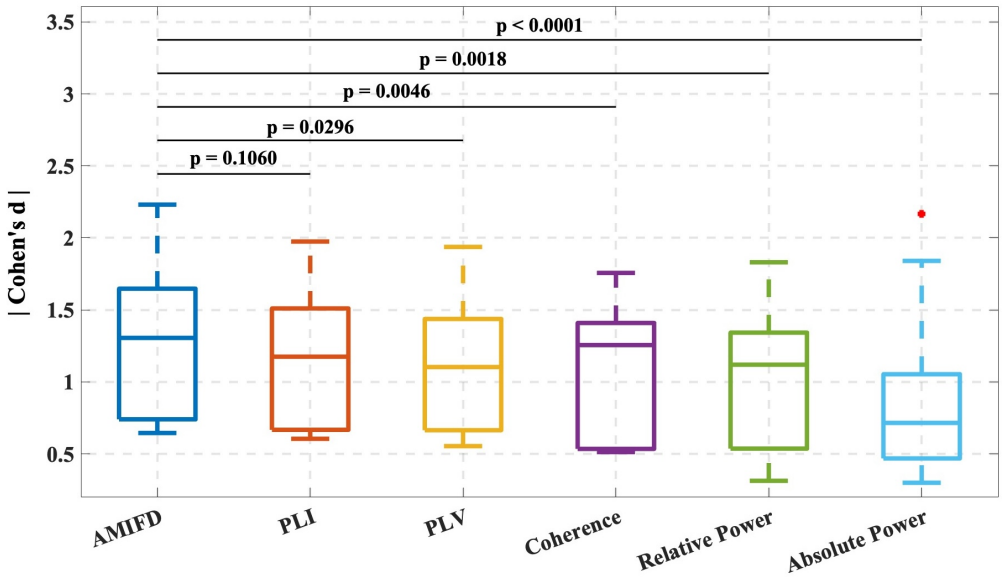

Figure 6.4: Comparative effect size of AMIFD against standard biomarkers. Box plots show the distribution of Cohen's d values for the top 10 performing features from each category, discriminating between interictal and preictal states. Horizontal lines indicate post-hoc statistical comparisons of each feature group against AMIFD, with corresponding p-values.

separability. Post-hoc statistical tests confirmed that the effect size of AMIFD was significantly greater than that of four of the five features tested: Absolute Power ($p < 0.0001$), Relative Power ($p = 0.0018$), Coherence ($p = 0.0046$), and Phase Locking Value (PLV) ($p = 0.0296$). While not reaching statistical significance, the AMIFD biomarker exhibited a trend toward a larger median effect size and a higher maximum observed value compared to the Phase Lag Index (PLI) ($p = 0.1060$).

6.3.2 Classifier Performance using AMIFD features

The seizure prediction performance is quantitatively evaluated by first comparing a pipeline using MUSE-based feature selection with clustering to a pipeline without it. To establish a performance benchmark, a baseline pipeline using the common mRMR feature selection technique [178] without clustering was also included for comparison. Therefore, the analysis was conducted on three distinct classification schemes:

1. **mRMR-NC**: A pipeline using the common minimum Redundancy Maximum Relevance (mRMR) feature selection method with **N**o **C**lustering.
2. **MUSE-NC**: A pipeline using **MUSE**-based feature selection with **N**o **C**lustering.
3. **MUSE-C**: A pipeline using **MUSE**-based feature selection with data **C**lustering prior to model training.

The performance metrics are summarized in Table 6.1 for subject-wise test folds performance, and in Table 6.2 for the overall mean performance across all subjects. The optimal k-of-N threshold through grid search is chosen as $k = 30$ of $N = 150$ overlapping windows, translating to a decision threshold based on 30 of the past 150 consecutive windows. To ensure a fair comparison against the baseline, the number of top-ranked features used in the mRMR-NC approach was matched to the quantity selected by the MUSE-NC pipeline, which corresponds to the larger feature set presented in Table 6.1.

A quantitative analysis revealed distinct differences in prediction efficacy and model complexity across these methods. The MUSE-C method, which incorporated clustering, demonstrated the highest mean sensitivity in seven of the ten subjects. Notably, it achieved a perfect mean sensitivity of 100.00% in three subjects (SID 20, 21, 29). The two non-clustered methods, mRMR-NC and MUSE-NC, produced nearly identical sensitivity results

Table 6.1: Subject-Specific Performance and Complexity for Clustered vs. Non-Clustered Prediction Pipelines using AMIFD Features

SID	# folds	# szs	# sz types	Engel class	Sensitivity of test folds (%)			FPR of test folds (/day)			# models (range)	Median # of selected feats.	
					mRMR-NC	MUSE-NC	MUSE-C	mRMR-NC	MUSE-NC	MUSE-C		NC	C
20	2	6	1	I	87.50 ± 17.68	87.50 ± 17.68	**100.00 ± 0.00**	1.29 ± 0.06	1.17 ± 0.24	**1.00 ± 0.00**	2	11	**8**
21	5	7	2	I	**100.00 ± 0.00**	93.33 ± 14.91	**100.00 ± 0.00**	0.90 ± 0.43	0.90 ± 0.37	**0.87 ± 0.18**	2–3	24	**15**
22	5	6	3	I	**90.00 ± 22.36**	**90.00 ± 22.36**	**90.00 ± 22.36**	0.92 ± 0.46	0.88 ± 0.24	**0.83 ± 0.16**	4	35	**22**
24	3	6	2	II	**91.67 ± 14.43**	**91.67 ± 14.43**	**91.67 ± 14.43**	1.56 ± 0.19	1.44 ± 0.19	**1.33 ± 0.00**	3-4	31	**20**
25	6	17	3	III	76.39 ± 38.88	80.56 ± 22.15	**87.50 ± 14.67**	2.92 ± 1.63	2.67 ± 0.41	**2.25 ± 0.42**	3	39	**31**
26	4	10	2	III	85.42 ± 17.18	85.42 ± 17.18	**93.75 ± 12.50**	1.52 ± 0.35	1.46 ± 0.32	**1.29 ± 0.21**	3-4	26	**18**
27	3	11	1	III	**86.11 ± 12.73**	**86.11 ± 12.73**	80.56 ± 17.35	1.17 ± 0.44	1.11 ± 0.19	**0.89 ± 0.19**	3	15	**7**
28	6	8	2	III	83.33 ± 40.82	83.33 ± 40.82	**94.44 ± 13.61**	1.75 ± 0.88	1.33 ± 0.52	**1.17 ± 0.41**	3	30	18
29	2	3	1	I	75.00 ± 35.36	75.00 ± 35.36	**100.00 ± 0.00**	1.04 ± 0.41	1.00 ± 0.00	**0**.71 ± 0.06	1-2	11	**7**
32	4	6	1	II	75.00 ± 28.87	75.00 ± 28.87	**87.50±25.00**	2.50 ± 0.41	2.38 ± 0.25	**2.13 ± 0.25**	2	8	**5**

Note: Sensitivity and FPR are reported as mean ± standard deviation across folds. NC and C denote pipelines with No Clustering and with Clustering, respectively; mRMR and MUSE refer to the feature selection algorithm used. *# sz types* refers to the number of clinically identified seizure types. *Median # Features* refers to the median number of features selected across all cross-validation folds for each subject. **Bolded entries** denote the best-performing case in terms of highest sensitivity, lowest FPR, and fewest selected features.

Table 6.2: Overall performance across all folds of all subjects, summarized from Table 6.1

Performance metric	**mRMR-NC**	**MUSE-NC**	**MUSE-C**
Sensitivity (%)	85.21 ± 25.84	85.00 ± 23.05	**92.08 ± 14.85**
FPR (/day)	1.65 ± 1.01	1.51 ± 0.71	**1.32 ± 0.60**

for most subjects. With respect to the FPR, the MUSE-C method yielded the lowest mean FPR for all ten subjects evaluated. In contrast, the mRMR-NC method consistently produced the highest FPR.

From a model complexity perspective, the MUSE-C pipeline required a smaller median number of unique features for every subject compared to the non-clustered MUSE-NC method. For example, for subject 22, the feature count was reduced from 35 to 22 following clustering. The number of models generated by the MUSE-C pipeline was low, ranging from one to four across all subjects.

6.4 Discussion

6.4.1 Effect Size and Robustness of the AMIFD Biomarker

The AMIFD biomarker demonstrates exceptional discriminative power in this study. A key finding from the comparative analysis is that its median Cohen's d value was higher than that of all other tested biomarkers, including Absolute Power, Relative Power, Coherence, PLV, and PLI. The observed superiority is reinforced by high statistical significance in four of these five comparisons: against Absolute Power, Relative Power, Coherence, and PLV.

The strong empirical performance is complemented by a fundamental theoretical advantage inherent to AMIFD's design. The biomarker is based on instantaneous frequency, which is the mathematical derivative of the instantaneous phase. Since raw iEEG signals are non-stationary, the cumulative instantaneous phase often exhibits strong, non-stationary trends. The derivative operation inherently removes these trends, yielding a more stationary signal. This makes the AMIFD biomarker more robust and reliable for long-term seizure prediction, as it is relatively less susceptible to the random drifts and baseline shifts that can corrupt non-stationary signals.

6.4.2 Performance Improvement using Multiple Seizure Models

The adoption of multiple seizure models (MUSE-C) is the key for the enhanced prediction performance observed in this study. A single, generalized model, as used in the mRMR-NC and MUSE-NC pipelines, is forced to learn an aggregate representation of preictal states. This approach can obscure the unique electrophysiological signatures of distinct seizure types, especially in patients who experience multiple types. The model must find a *one-size-fits-all* solution, which can limit its ability to predict every seizure event accurately.

The MUSE-C method overcomes this constraint by first clustering seizures based on their preictal feature representations. It allows for the training of multiple specialized models, each tailored to the specific preictal patterns of a particular seizure cluster. This is empirically supported by the data in Table 6.1, where the MUSE-C method not only improved sensitivity but also consistently lowered the FPR across all subjects. The superior performance is achieved with a smaller median number of features per model. This supports the hypothesis that a patient's seizures have distinct subtypes, each best predicted by an efficient, specialized model using a compact set of highly relevant features, rather than accommodating a larger, more varied set of features needed for a generalized model.

6.4.3 Comparison with Existing Long-term Seizure Prediction Methods

To contextualize the performance of the proposed algorithm, Table 6.3 summarizes representative studies on seizure prediction from multi-day EEG/iEEG recordings. These works vary significantly in their recording modalities, monitoring durations, dataset sizes, and the criteria used for performance evaluation. Note that a comparison with the PSD-based framework, discussed in the previous chapter, is not provided in the table as the two methods were developed around the same time. A comparison between the two will be discussed in the conclusion chapter.

Table 6.3: Representative seizure prediction studies in humans utilizing multi-day EEG recordings

Year	**Method**	**Dataset Description**	**Feature and Model Used**	**Performance**
2003	Iasemidis *et al.* [108]	iEEG from 5 patients (28–32 sites, 200 Hz), 55 seizures over 18 days (**short-term**)	Adaptive Seizure Prediction Algorithm (ASPA) using Short-Term Maximum Lyapunov Exponents (STLmax)	SS = 82%, FPR = 0.16/hr (3.84/day)
2005	Iasemidis *et al.* [109]	iEEG from 2 subjects, 25 seizures over 107.5 hrs (**short-term**)	ASPA + T-index tracking of seizure precursors	SS = 81.8%, FPR = 0.15/hr (3.6/day)
2013	Cook *et al.*[135]	15 patients, NeuroVista iEEG (16 channels, 400 Hz) over 4 months (**long-term**)	Beta-band (16–32 Hz) power + k-NN classifier	SS = 65–100%
2016	Brinkmann *et al.*[146]	Crowd-sourced challenge: 230 hrs of iEEG (3 × 8 subdural grid) from 2 patients, sampled at 5 kHz with a total of 46 seizures (canine data used for the contest not described here) (**fragmented long-term**)	Top ranked model: Spectral, entropy, correlation, and statistical features + Lasso-regularized GLM classifier	SS = 71% @ 75% specificity (includes canine data)
2018	Kuhlmann *et al.*[147]	Crowd-sourced seizure prediction challenge: 3 patients, 1-year 16-channel subdural iEEG, 1100+ seizures (f_s=400 Hz) (**long-term**)	Top ranked model: Spectral, AR, fractal, coherence features; SVM, k-NN, GLM, XGBoost	Top ranked model: SS = 58% @ 75% specificity

Year	Method	Dataset Description	Feature and Model Used	Performance
2019	Wei *et al.* [110]	15 patients, 22-lead EEG (f_s=500 Hz), 540 hrs, 164 seizures (**short-term**)	CNN for features + LSTM for preictal sequence learning	SS = 91.88%, SP = 86.13%, ACC = 93.4%, FPR = 0.04/hr (0.96/day)
2021	Stirling *et al.* [112]	Sub-scalp EEG from 5 subjects (f_s=250 Hz) with refractory focal epilepsy, totaling 128 seizures across 55–106 days per subject (**long-term**).	Ensemble: Random Forest regressor + Logistic Regression	SS = 66%, ACC = 83%, AUC = 0.88
2022	Viana *et al.* [113]	6 subjects, 11,000+ hrs sqEEG (f_s=207 Hz), 12–36 seizures/patient (**long-term**)	Subject-specific LSTM trained on preictal/interictal segments	SS = 67–80%, AUC = 0.59–0.79, Time-in-warning = 10.9–44.4%
2025	**Proposed model**	10 subjects, 2088 hrs iEEG (f_s=2 kHz), 80 seizures (**long-term**)	MUSE-based clustering of AMIFD features + random forest classifier	**Mean SS = 92.08%, FPR = 1.32/day**

Early work in this area often employed nonlinear dynamical features on smaller datasets, achieving moderate success [108, 109]. Subsequent large-scale studies demonstrated the feasibility of long-term prediction using ambulatory iEEG, though subject-wise variability remained a significant hurdle [135, 147]. More recently, deep learning architectures like LSTMs and CNNs have been applied to scalp and sub-scalp EEG to leverage spatio-temporal patterns, reporting promising results but often at the cost of high computational complexity and the need for extensive training data [113]. In contrast, the proposed framework, built upon a novel and robust AMIFD biomarker and using a multi-model approach,

achieves a mean sensitivity of **92.08%** and a FPR of **1.32/day**. Our method's sensitivity surpasses all prior work discussed in Table 6.3. Its FPR is also lower, except for that of [110], which analyzed short-term data, unlike the long-term recordings used in this study. While direct comparisons are challenging, readers are encouraged to consider the contextual differences between studies, including recording durations and seizure counts, when evaluating performance.

6.4.4 Avenues for Improvement

It is important to acknowledge the limiting aspects of this study that provide avenues for future research. The study is designed retrospectively on a small cohort of ten subjects. The challenge of generalizing from a limited dataset is particularly highlighted by specific missed alarms where seizures originated from onset zones not represented in the training data. For instance, a seizure in SID 27 with a right temporal onset was missed because the model was trained exclusively on seizures from the left, making the event structurally distinct. Although the evaluation approximated real-time conditions using a k-of-N voting scheme, true clinical viability must be confirmed through future validation on larger, more diverse patient populations in a prospective setting.

The current framework did not incorporate spatial information from the electrode montage, and did not formally quantify the potential redundancy between the cluster-specific models. Future work may improve model specificity by including this spatial context and moving beyond binary predictions to incorporate probabilistic risk scores. Integrating the system with advanced techniques such as adaptive or lifelong learning frameworks would also be a valuable next step toward creating a more nuanced model that evolves with the patient over time.

Chapter 7

Conclusion and Future Direction

7.1 Summary of Thesis Contributions

This doctoral research was driven by the critical need for more advanced and reliable computational methods to manage drug-resistant epilepsy. While significant progress has been made in computational neuroengineering, patient outcomes remain hindered by two fundamental challenges: the precise localization of the epileptogenic zone (EZ) for surgical planning and the reliable, long-term prediction of seizures. The work in this dissertation addresses both challenges by developing and validating novel, patient-specific frameworks to analyze complex brain network dynamics.

The primary contributions of this thesis are twofold:

1. The research moves beyond traditional, static biomarkers to establish a dynamic network signature of the Seizure Onset Zone (SoZ). By applying effective connectivity and graph-based analysis to both ictal and interictal iEEG, this work demonstrates that the SoZ is not a static target. Instead, it is a dynamic region that transitions from a state of network "quiescence" between seizures to one of "hyper-activation," where it becomes the central hub driving ictogenesis.
2. This dissertation presents a complete, dual-component framework for long-term seizure prediction that addresses the key limitations of current systems. This framework successfully addresses the problem of intra-patient seizure variability by introducing a multi-model classification scheme that trains "expert" models for distinct seizure

types. It simultaneously solves the problem of signal non-stationarity by introducing the Absolute Mean Instantaneous Frequency Difference (AMIFD), a novel biomarker inherently robust to long-term signal drift.

The methods developed in this thesis collectively provide a more nuanced understanding of epileptogenic networks and offer robust, validated pathways toward more personalized and clinically viable diagnostic and advisory systems.

7.2 Integration of Key Research Findings

This section consolidates the primary results of the four publications that constitute this doctoral work. The findings are organized thematically into two main research thrusts: the identification of the Seizure Onset Zone (SoZ) and the forecasting of impending seizures.

7.2.1 A Dynamic Network Signature for SoZ Identification

A foundational challenge in epilepsy surgery is the accurate localization of the SoZ. Traditional localization methods often rely on biomarkers such as high-frequency oscillations (HFOs) and spectral power changes. These markers, while helpful, primarily describe localized activity and often fail to capture the complex, network-level interactions that define an epileptogenic zone. The research in this thesis proposed that the SoZ is better understood as a dynamic network hub, whose functional role changes dramatically between resting and seizure states.

To test this hypothesis, this work introduced a novel graph-theoretic framework to analyze the effective connectivity (EC) of the epileptogenic network. After comparing three different EC measures, *Frequency-Domain Convergent Cross-Mapping (FD-CCM)* was identified as the most powerful method for revealing directional connectivity patterns.

The first significant finding, derived from ictal (during-seizure) recordings, demonstrated that SoZ electrodes function as the central drivers of seizure activity. In this state, SoZ electrodes showed significantly elevated centrality measures, indicating they are hyperactive hubs in the epileptic network. This graph-based approach proved highly discriminative, with a supervised machine learning model achieving a mean AU-ROC of 0.93 in separating SoZ from non-SoZ electrodes.

While effective, this ictal-based analysis relies on capturing seizure events, which can be difficult and time-consuming in a clinical setting. The second component of this research, therefore, adapted this robust FD-CCM framework to the more readily available interictal (resting-state) data. This investigation revealed a crucial and complementary pattern: between seizures, the SoZ electrodes exist in a state of network *quiescence* characterized by significantly lower centrality values. This finding suggests a state of functional isolation from the broader brain network during the resting phase.

By combining these two findings, this thesis establishes a novel dynamic signature of the SoZ. This signature characterizes the epileptogenic zone as a region that transitions from a state of network isolation to one of network hyper-activation. This dynamic profile is a much richer biomarker than static measures. To validate its clinical utility, a hybrid classifier was developed that integrated this new graph-based biomarker with traditional ones (HFOs, Phase-Amplitude Coupling, and relative band power). This integrated model, using only interictal data, achieved a robust mean accuracy of 89% and a mean sensitivity of 82% for SoZ identification, demonstrating its practical value for surgical planning.

7.2.2 A Novel Framework for Patient-Specific Seizure Prediction

The second major thrust of this dissertation was to develop a seizure prediction framework robust enough for long-term clinical use. The primary obstacle to this goal is intra-patient heterogeneity: the fact that a single patient can exhibit multiple, distinct seizure types, each with its own unique preictal signature. A *one-size-fits-all* patient-specific model will inevitably fail by either averaging or entirely missing these diverse signatures.

The core contribution of this research in seizure forecasting is the development of a patient-specific, multi-model classification framework designed to solve this exact problem. This framework uses the Minimum Uncertainty and Sample Elimination (MUSE) algorithm to analyze the preictal features of each individual seizure, automatically clustering seizures with similar signatures. A separate, specialized classifier is then trained for each cluster. The power of this adaptive framework was demonstrated in two independent validation studies using different feature classes.

Validation with Power Spectral Density Features

In the first validation, the multi-model framework was applied using a comprehensive set of Power Spectral Density (PSD) features, including absolute, relative, and ratio-based measures. These features are well-established but were used here to test the hypothesis that modeling heterogeneity itself was the key to improving performance.

The results were unequivocal. The multi-model (clustered) approach was directly compared with a single-model (non-clustered) approach using the same PSD features. The multi-model framework dramatically improved performance, increasing the mean sensitivity from 89.2% to 98.5% while simultaneously halving the false positive rate from 1.15 to 0.62 alarms per day. This finding provided definitive proof of two concepts: first, that significant latent electrophysiological heterogeneity exists beyond standard clinical labels, and second, that the proposed multi-model framework can successfully capture it.

Validation with a Novel Synchronization Biomarker (AMIFD)

In a second, independent study, the framework's robustness was tested with a completely different type of feature: a novel, purpose-built biomarker introduced in this thesis as the *Absolute Mean Instantaneous Frequency Difference (AMIFD)*.

This feature was designed to be a robust, bivariate measure of synchronization that is less susceptible to the signal drift and baseline shifts common in long-term recordings. The AMIFD biomarker was first validated on its own, demonstrating a statistically significant larger effect size (Cohen's d) in separating preictal from interictal states compared to other synchronization features like coherence ($p < 0.005$) and Phase Locking Value ($p < 0.03$).

This new AMIFD feature set was then integrated into the same multi-model clustering framework. Once again, the framework proved its value by significantly improving performance over a single-model baseline. The final AMIFD-based pipeline achieved a strong mean sensitivity of 92.1% and a low false-positive rate of 1.32 alarms per day.

7.3 Limitations of the Present Work

While the results are promising, it is essential to acknowledge the limitations of this research, which also define the pathway for future work.

1. **Retrospective Analysis**: All methods were developed and validated retrospectively using pre-recorded iEEG data. While the prediction pipeline was designed to simulate real-time use (e.g., using a refractory period and session-based cross-validation), its true performance must be confirmed in a prospective, real-time clinical trial.

2. **Invasive Data**: The high-fidelity results in this thesis were achieved using invasive iEEG. This data has a high signal-to-noise ratio but is only available from a small population of pre-surgical candidates. The generalizability of these specific features (especially FD-CCM and AMIFD) to non-invasive scalp EEG is not yet known.

3. **Limited Cohort Size**: Consistent with the challenges of iEEG research, the studies in this thesis were conducted on limited patient cohorts (N=10 for the core prediction studies and N=27 for the primary SoZ study). Validation on larger, more diverse patient populations is required to confirm the broad applicability of these findings.

4. **Challenge with novel seizure types**: The multi-model prediction framework is adaptive but still relies on data seen during training. As noted in the analysis, some missed seizures were *novel* events—seizure types that were not represented in the training data. This highlights a fundamental challenge for all patient-specific models.

7.4 Future Research Directions

The successful validation of these novel frameworks for SoZ identification and seizure prediction lays the foundation for three primary avenues of future work: prospective clinical validation, methodological integration, and translational expansion into broader clinical applications.

7.4.1 Prospective Clinical Validation

The most critical next step is to move from retrospective analysis to prospective validation.

- **Real-time, Prospective Trial:** The gold standard will be to implement the multi-model seizure prediction framework in a real-time, forward-looking clinical trial. This

is the only way to truly assess its performance and clinical viability, including its long-term stability and its impact on patient quality of life.

- **Larger, Diverse Cohorts:** Both the SoZ identification and seizure prediction frameworks must be validated on larger, more diverse patient populations. This will be essential to confirm their statistical power and ensure they generalize beyond the initial cohorts, which were primarily composed of patients with temporal lobe epilepsy.

7.4.2 Methodological Integration and Enhancement

This research produced several distinct, high-performance methodologies. The next logical step is to integrate them and enhance their capabilities.

- **Cross-Pollination of Methods:** This thesis presents an opportunity to "cross-pollinate" its own findings. For instance, the **AMIFD biomarker**, which proved so robust for prediction, should be tested as a novel feature for **interictal SoZ identification**. Conversely, the **dynamic centrality signature** used for SoZ localization could be explored as a powerful, network-based feature for **seizure prediction**.

- **Adaptive *Lifelong Learning* Framework:** The multi-model framework successfully addresses known seizure types but is still vulnerable to "novel" events not seen in training. A significant future goal is to evolve this into an adaptive **"lifelong learning" system**. Such a system would be designed to detect a novel seizure type, flag it for clinical verification, and then automatically train and add a *new* expert classifier to its ensemble on the fly, allowing the system to learn and adapt to the patient over time.

7.4.3 Translational and Clinical Expansion

Finally, a major long-term goal is to move these advanced methods beyond the specialized iEEG environment and make them accessible to a broader patient population.

- **Adaptation to Non-Invasive Scalp EEG:** A key challenge will be adapting these frameworks for use with non-invasive scalp EEG. While the high signal-to-noise ratio

of iEEG was ideal for development, the true translational impact of these methods will come from their successful application to scalp EEG. The robustness of the AMIFD feature, in particular, may make it a promising candidate for this difficult transition.

- **Optimized Hardware Implementation:** As demonstrated in this work, the computational load of these algorithms is feasible for real-time applications. Future work should focus on further optimizing these algorithms for low-power, implantable, or wearable hardware. This would pave the way for a fully contained advisory system, moving this research from a computational proof-of-concept to a tangible clinical device.

References

[1] Hans O Lüders, Imad Najm, Dileep Nair, Peter Widdess-Walsh, and William Bingman. The epileptogenic zone: general principles. *Epileptic disorders*, 8(2):1–9, 2006.

[2] Epilepsy foundation. `https://www.epilepsy.com`.

[3] Roland D Thijs, Rainer Surges, Terence J O'Brien, and Josemir W Sander. Epilepsy in adults. *The Lancet*, 393(10172):689–701, 2019. (doi: 10.1016/S0140-6736(18)32596-0).

[4] Carl E Stafstrom and Lionel Carmant. Seizures and epilepsy: an overview for neuroscientists. *Cold Spring Harbor perspectives in medicine*, 5(6):a022426, 2015. (doi: 10.1101/cshperspect.a022426).

[5] Sai Sanjay Balaji and Keshab K. Parhi. Seizure onset zone identification from ieeg: A review. *IEEE Access*, 10:62535–62547, 2022. DOI: 10.1109/ACCESS.2022.3182716.

[6] Hans Berger. Über das elektroenkephalogramm des menschen. *Archiv für psychiatrie und nervenkrankheiten*, 87(1):527–570, 1929.

[7] Gyorgy Buzsaki. *Rhythms of the Brain*. Oxford university press, 2006.

[8] Neep Hazarika, Jean Zhu Chen, Ah Chung Tsoi, and Alex Sergejew. Classification of eeg signals using the wavelet transform. *Signal processing*, 59(1):61–72, 1997.

[9] Richard J Staba, Matt Stead, and Gregory A Worrell. Electrophysiological biomarkers of epilepsy. *Neurotherapeutics*, 11(2):334–346, 2014.

[10] Anatol Bragin, Jerome Engel Jr, Charles L Wilson, Itzhak Fried, and Gary W Mathern. Hippocampal and entorhinal cortex high-frequency oscillations (100–500 Hz) in human epileptic brain and in kainic acid-treated rats with chronic seizures. *Epilepsia*, 40(2):127–137, 1999.

[11] Alard Roebroeck, Anil K Seth, and Pedro Valdes-Sosa. Causal Time Series Analysis of Functional Magnetic Resonance Imaging Data. In *NIPS mini-symposium on causality in time series*, pages 65–94. PMLR, 2011.

[12] Sandeep Avvaru, Noam Peled, Nicole R Provenza, Alik S Widge, and Keshab K Parhi. Region-level functional and effective network analysis of human brain during cognitive task engagement. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:1651–1660, 2021.

[13] Ed Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature reviews neuroscience*, 10(3):186–198, 2009.

[14] Karl J Friston. Functional and effective connectivity: a review. *Brain connectivity*, 1(1):13–36, 2011.

[15] Tingting Xu, Kathryn R Cullen, Bryon Mueller, Mindy W Schreiner, Kelvin O Lim, S Charles Schulz, and Keshab K Parhi. Network analysis of functional brain connectivity in borderline personality disorder using resting-state fmri. *NeuroImage: Clinical*, 11:302–315, 2016.

[16] André M Bastos and Jan-Mathijs Schoffelen. A tutorial review of functional connectivity analysis methods and their interpretational pitfalls. *Frontiers in systems neuroscience*, 9:175, 2016.

[17] Patrick A Stokes and Patrick L Purdon. A study of problems encountered in granger causality analysis from a neuroscience perspective. *Proceedings of the national academy of sciences*, 114(34):E7063–E7072, 2017.

[18] David Marc Anton Mehler and Konrad Paul Kording. The lure of misleading causal statements in functional connectivity research. *arXiv preprint arXiv:1812.03363*, 2018.

[19] Types of seizures. `https://www.hopkinsmedicine.org/health/conditions-and-diseases/epilepsy/types-of-seizures`.

[20] Proposal for revised clinical and electroencephalographic classification of epileptic seizures. *Epilepsia*, 22(4):489–501, 1981, https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1528-1157.1981.tb06159.x.

[21] Temporal Lobe Epilepsy (TLE). `https://www.epilepsy.com/learn/types-epilepsy-syndromes/temporal-lobe-epilepsy-aka-tle`.

[22] Robert S. Fisher, J. Helen Cross, Jacqueline A. French, Norimichi Higurashi, Edouard Hirsch, Floor E. Jansen, Lieven Lagae, Solomon L. Moshé, Jukka Peltola, Eliane Roulet Perez, Ingrid E. Scheffer, and Sameer M. Zuberi. Operational classification of seizure types by the international league against epilepsy: Position paper of the ilae commission for classification and terminology. *Epilepsia*, 58(4):522–530, 2017, https://onlinelibrary.wiley.com/doi/pdf/10.1111/epi.13670.

[23] Ingrid E Scheffer, Samuel Berkovic, Giuseppe Capovilla, Mary B Connolly, Jacqueline French, Laura Guilhoto, Edouard Hirsch, Satish Jain, Gary W Mathern, Solomon L Moshé, et al. ILAE classification of the epilepsies: position paper of the ILAE Commission for Classification and Terminology. *Epilepsia*, 58(4):512–521, 2017.

[24] Wilder Penfield. Epileptogenic lesions. *Acta Neurol Belg*, 56:75–88, 1956.

[25] Herbert H Jasper, Genevieve Arfel-Capdeville, and Theodore Rasmussen. Evaluation of EEG and cortical electrographic studies for prognosis of seizures following surgical excision of epileptogenic lesions. *Epilepsia*, 2(2):130–137, 1961.

[26] Felix Rosenow and Hans Lüders. Presurgical evaluation of epilepsy. *Brain*, 124(9):1683–1700, 2001.

[27] Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005.

[28] A James Rowan, Ruud J Veldhuisen, and Nico JD Nagelkerke. Comparative evaluation of sleep deprivation and sedated sleep eegs as diagnostic aids in epilepsy. *Electroencephalography and Clinical Neurophysiology*, 54(4):357–364, 1982.

[29] Julia Jacobs, Pierre LeVan, Rahul Chander, Jeffery Hall, François Dubeau, and Jean Gotman. Interictal high-frequency oscillations (80–500 Hz) are an indicator of seizure onset areas independent of spikes in the human epileptic brain. *Epilepsia*, 49(11):1893–1907, 2008.

[30] JD Jirsch, E Urrestarazu, P LeVan, A Olivier, F Dubeau, and J Gotman. High-frequency oscillations during human focal seizures. *Brain*, 129(6):1593–1608, 2006.

[31] Matthias Dümpelmann, Julia Jacobs, and Andreas Schulze-Bonhage. Temporal and spatial characteristics of high frequency oscillations as a new biomarker in epilepsy. *Epilepsia*, 56(2):197–206, 2015.

[32] Su Liu, Zhiyi Sha, Altay Sencer, Aydin Aydoseli, Nerse Bebek, Aviva Abosch, Thomas Henry, Candan Gurses, and Nuri Firat Ince. Exploring the time–frequency content of high frequency oscillations for automated identification of seizure onset zone in epilepsy. *Journal of neural engineering*, 13(2):026026, 2016.

[33] Urszula Malinowska, Gregory K Bergey, Jaroslaw Harezlak, and Christophe C Jouny. Identification of seizure onset zone and preictal state based on characteristics of high frequency oscillations. *Clinical Neurophysiology*, 126(8):1505–1513, 2015.

[34] Andrew Matsumoto, Benjamin H Brinkmann, S Matthew Stead, Joseph Matsumoto, Michal T Kucewicz, W Richard Marsh, Frederic Meyer, and Gregory Worrell. Pathological and physiological high-frequency oscillations in focal human epilepsy. *Journal of neurophysiology*, 110(8):1958–1964, 2013.

[35] Shuang Wang, Irene Z Wang, Juan C Bulacio, John C Mosher, Jorge Gonzalez-Martinez, Andreas V Alexopoulos, Imad M Najm, and Norman K So. Ripple classification helps to localize the seizure-onset zone in neocortical epilepsy. *Epilepsia*, 54(2):370–376, 2013.

[36] Rafeed Alkawadri, Nicolas Gaspard, Irina I Goncharova, Dennis D Spencer, Jason L Gerrard, Hitten Zaveri, Robert B Duckrow, Hal Blumenfeld, and Lawrence J Hirsch. The spatial and signal characteristics of physiologic high frequency oscillations. *Epilepsia*, 55(12):1986–1995, 2014.

[37] Hisako Fujiwara, Hansel M Greiner, Ki Hyeong Lee, Katherine D Holland-Bouley, Joo Hee Seo, Todd Arthur, Francesco T Mangano, James L Leach, and Douglas F Rose. Resection of ictal high-frequency oscillations leads to favorable surgical outcome in pediatric epilepsy. *Epilepsia*, 53(9):1607–1617, 2012.

[38] Claire Haegelen, Piero Perucca, Claude-Edouard Châtillon, Luciana Andrade-Valença, Rina Zelmann, Julia Jacobs, D Louis Collins, François Dubeau, André Olivier, and Jean Gotman. High-frequency oscillations, extent of surgical resection, and surgical outcome in drug-resistant focal epilepsy. *Epilepsia*, 54(5):848–857, 2013.

[39] Jounhong Ryan Cho, Dae Lim Koo, Eun Yeon Joo, Dae Won Seo, Seung-Chyul Hong, Premysl Jiruska, and Seung Bong Hong. Resection of individually identified high-rate high-frequency oscillations region is associated with favorable outcome in neocortical epilepsy. *Epilepsia*, 55(11):1872–1883, 2014.

[40] Robert S Fisher, WR Webber, Ronald P Lesser, Santiago Arroyo, and Sumio Uematsu. High-frequency EEG activity at the start of seizures. *Journal of clinical neurophysiology: official publication of the American Electroencephalographic Society*, 9(3):441–448, 1992.

[41] Roger D Traub, Miles A Whittington, Eberhard H Buhl, Fiona EN LeBeau, Andrea Bibbig, Stewart Boyd, Helen Cross, and Torsten Baldeweg. A possible role for gap junctions in generation of very fast EEG oscillations preceding the onset of, and perhaps initiating, seizures. *Epilepsia*, 42:153–170, 2001.

[42] Greg A Worrell, Landi Parish, Stephen D Cranstoun, Rachel Jonas, Gordon Baltuch, and Brian Litt. High-frequency oscillations and seizure generation in neocortical epilepsy. *Brain*, 127(7):1496–1506, 2004.

[43] Houman Khosravani, Nikhil Mehrotra, Michael Rigby, Walter J Hader, C Robert Pinnegar, Neelan Pillay, Samuel Wiebe, and Paolo Federico. Spatial localization

and time-dependant changes of electrographic high frequency oscillations in human temporal lobe epilepsy. *Epilepsia*, 50(4):605–616, 2009.

[44] Richard J Staba, Charles L Wilson, Anatol Bragin, Itzhak Fried, and Jerome Engel Jr. Quantitative analysis of high-frequency oscillations (80–500 Hz) recorded in human epileptic hippocampus and entorhinal cortex. *Journal of neurophysiology*, 88(4):1743–1752, 2002.

[45] Justin A Blanco, Matt Stead, Abba Krieger, Jonathan Viventi, W Richard Marsh, Kendall H Lee, Gregory A Worrell, and Brian Litt. Unsupervised classification of high-frequency oscillations in human neocortical epilepsy and control patients. *Journal of neurophysiology*, 104(5):2900–2912, 2010.

[46] Krit Charupanit, Indranil Sen-Gupta, Jack J Lin, and Beth A Lopour. Amplitude of high frequency oscillations as a biomarker of the seizure onset zone. *Clinical Neurophysiology*, 131(11):2542–2550, 2020.

[47] Kevin J Staley and F Edward Dudek. Interictal spikes and epileptogenesis. *Epilepsy currents*, 6(6):199–202, 2006.

[48] ARTHUR A WARD Jr. The epileptic spike. *Epilepsia*, 1(1-5):600–606, 1959.

[49] Daniel Lachner-Piza, Julia Jacobs, Jonas C Bruder, Andreas Schulze-Bonhage, Thomas Stieglitz, and Matthias Dümpelmann. Automatic detection of high-frequency-oscillations and their sub-groups co-occurring with interictal-epileptic-spikes. *Journal of neural engineering*, 17(1):016030, 2020.

[50] Christian G Bénar, Laeticia Chauvière, Fabrice Bartolomei, and Fabrice Wendling. Pitfalls of high-pass filtering for detecting epileptic oscillations: a technical note on "false" ripples. *Clinical Neurophysiology*, 121(3):301–310, 2010.

[51] Nicolas Roehri, Francesca Pizzo, Stanislas Lagarde, Isabelle Lambert, Anca Nica, Aileen McGonigal, Bernard Giusiano, Fabrice Bartolomei, and Christian-George Bénar. High-frequency oscillations are not better biomarkers of epileptogenic tissues than spikes. *Annals of neurology*, 83(1):84–97, 2018.

[52] Julia Jacobs, Joyce Y Wu, Piero Perucca, Rina Zelmann, Malenka Mader, Francois Dubeau, Gary W Mathern, Andreas Schulze-Bonhage, and Jean Gotman. Removing high-frequency oscillations: a prospective multicenter study on seizure outcome. *Neurology*, 91(11):e1040–e1052, 2018.

[53] Clive WJ Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438, 1969.

[54] Marcin Jan Kaminski and Katarzyna J Blinowska. A new method of the description of the information flow in the brain structures. *Biological cybernetics*, 65(3):203–210, 1991.

[55] Luiz A Baccalá and Koichi Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological cybernetics*, 84(6):463–474, 2001.

[56] Bhim M Adhikari, Charles M Epstein, and Mukesh Dhamala. Localizing epileptic seizure onsets with Granger causality. *Physical Review E*, 88(3):030701, 2013.

[57] Eun-Hyoung Park and Joseph R Madsen. Granger causality analysis of interictal iEEG predicts seizure focus and ultimate resection. *Neurosurgery*, 82(1):99–109, 2018.

[58] Anil K Seth. A MATLAB toolbox for Granger causal connectivity analysis. *Journal of neuroscience methods*, 186(2):262–273, 2010.

[59] Mark P Richardson. Large scale brain models of epilepsy: dynamics meets connectomics. *Journal of Neurology, Neurosurgery & Psychiatry*, 83(12):1238–1248, 2012.

[60] Aristidis Likas, Nikos Vlassis, and Jakob J Verbeek. The global k-means clustering algorithm. *Pattern recognition*, 36(2):451–461, 2003.

[61] Ana Coito, Gijs Plomp, Mélanie Genetti, Eugenio Abela, Roland Wiest, Margitta Seeck, Christoph M Michel, and Serge Vulliemoz. Dynamic directed interictal connectivity in left and right temporal lobe epilepsy. *Epilepsia*, 56(2):207–217, 2015.

[62] Olivier David, Isabelle Guillemain, Sandrine Saillet, Sebastien Reyt, Colin Deransart, Christoph Segebarth, and Antoine Depaulis. Identifying neural drivers with functional MRI: an electrophysiological validation. *PLoS biology*, 6(12):e315, 2008.

[63] Boris Petrovich Bezruchko, Vladimir Ivanovich Ponomarenko, Mikhail Dmitrievich Prokhorov, Dmitrii Alekseevich Smirnov, and Peter Alexander Tass. Modeling nonlinear oscillatory systems and diagnostics of coupling between them using chaotic time series analysis: applications in neurophysiology. *Physics-Uspekhi*, 51(3):304–310, 2008.

[64] Daniele Marinazzo, Mario Pellicoro, and Sebastiano Stramaglia. Kernel method for nonlinear Granger causality. *Physical review letters*, 100(14):144103, 2008.

[65] Meng Hu and Hualou Liang. A copula approach to assessing Granger causality. *NeuroImage*, 100:125–134, 2014.

[66] Daniele Marinazzo, Wei Liao, Huafu Chen, and Sebastiano Stramaglia. Nonlinear connectivity by Granger causality. *Neuroimage*, 58(2):330–338, 2011.

[67] Mingzhou Ding, Steven L Bressler, Weiming Yang, and Hualou Liang. Short-window spectral analysis of cortical event-related potentials by adaptive multivariate autoregressive modeling: data preprocessing, model validation, and variability assessment. *Biological cybernetics*, 83(1):35–45, 2000.

[68] Maciej Kamiński, Mingzhou Ding, Wilson A Truccolo, and Steven L Bressler. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biological cybernetics*, 85(2):145–157, 2001.

[69] Mattia F Pagnotta and Gijs Plomp. Time-varying mvar algorithms for directed connectivity analysis: Critical comparison in simulations and benchmark eeg data. *PloS one*, 13(6):e0198846, 2018.

[70] David Pascucci, Maria Rubega, and Gijs Plomp. Modeling time-varying brain networks with a self-tuning optimized kalman filter. *PLoS computational biology*, 16(8):e1007566, 2020.

[71] Pieter Van Mierlo, Evelien Carrette, Hans Hallez, Kristl Vonck, Dirk Van Roost, Paul Boon, and Steven Staelens. Accurate epileptogenic focus localization through time-variant functional connectivity analysis of intracranial electroencephalographic signals. *Neuroimage*, 56(3):1122–1133, 2011.

[72] Pieter Van Mierlo, Evelien Carrette, Hans Hallez, Robrecht Raedt, Alfred Meurs, Stefaan Vandenberghe, Dirk Van Roost, Paul Boon, Steven Staelens, and Kristl Vonck. Ictal-onset localization through connectivity analysis of intracranial EEG signals in patients with refractory epilepsy. *Epilepsia*, 54(8):1409–1418, 2013.

[73] James Massey. Causality, feedback and directed information. In *Proc. Int. Symp. Inf. Theory Applic.(ISITA-90)*, pages 303–305. Citeseer, 1990.

[74] Rakesh Malladi, Giridhar Kalamangalam, Nitin Tandon, and Behnaam Aazhang. Identifying seizure onset zone from the causal connectivity inferred using directed information. *IEEE Journal of Selected Topics in Signal Processing*, 10(7):1267–1283, 2016.

[75] Ying Liu and Selin Aviyente. Directed information measure for quantifying the information flow in the brain. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 2188–2191. IEEE, 2009.

[76] Yonathan Murin, Jeremy Kim, Josef Parvizi, and Andrea Goldsmith. SozRank: A new approach for localizing the epileptic seizure onset zone. *PLoS computational biology*, 14(1):e1005953, 2018.

[77] Joost B Wagenaar, Benjamin H Brinkmann, Zachary Ives, Gregory A Worrell, and Brian Litt. A multimodal platform for cloud-based collaborative research. In *2013 6th international IEEE/EMBS conference on neural engineering (NER)*, pages 1386–1389. IEEE, 2013.

[78] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford InfoLab, 1999.

[79] Astrid Von Stein and Johannes Sarnthein. Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta synchronization. *International journal of psychophysiology*, 38(3):301–313, 2000.

[80] Ole Jensen and Laura L Colgin. Cross-frequency coupling between neuronal oscillations. *Trends in cognitive sciences*, 11(7):267–269, 2007.

[81] Ryan T Canolty, Erik Edwards, Sarang S Dalal, Maryam Soltani, Srikantan S Nagarajan, Heidi E Kirsch, Mitchel S Berger, Nicholas M Barbaro, and Robert T Knight. High gamma power is phase-locked to theta oscillations in human neocortex. *science*, 313(5793):1626–1628, 2006.

[82] Ryan T Canolty and Robert T Knight. The functional role of cross-frequency coupling. *Trends in cognitive sciences*, 14(11):506–515, 2010.

[83] Shennan A Weiss, Garrett P Banks, Guy M McKhann Jr, Robert R Goodman, Ronald G Emerson, Andrew J Trevelyan, and Catherine A Schevon. Ictal high frequency oscillations distinguish two types of seizure territories in humans. *Brain*, 136(12):3796–3808, 2013.

[84] Shennan A Weiss, Athena Lemesiou, Robert Connors, Garrett P Banks, Guy M McKhann, Robert R Goodman, Binsheng Zhao, Christopher G Filippi, Mark Nowell, Roman Rodionov, et al. Seizure localization using ictal phase-locked high gamma: A retrospective surgical outcome study. *Neurology*, 84(23):2320–2328, 2015.

[85] M Le Van Quyen, C_ Adam, J Martinerie, M Baulac, S Clemenceau, and F Varela. Spatio-temporal characterizations of non-linear changes in intracranial activities prior to human temporal lobe seizures. *European Journal of Neuroscience*, 12(6):2124–2134, 2000.

[86] Mina Amiri, Birgit Frauscher, and Jean Gotman. Phase-amplitude coupling is elevated in deep sleep and in the onset zone of focal epileptic seizures. *Frontiers in human neuroscience*, 10:387, 2016.

[87] Nuria E Cámpora, Camilo J Mininni, Silvia Kochen, and Sergio E Lew. Seizure localization using pre ictal phase-amplitude coupling in intracranial electroencephalography. *Scientific reports*, 9(1):1–8, 2019.

[88] Shennan A Weiss, Iren Orosz, Noriko Salamon, Stephanie Moy, Linqing Wei, Maryse A Van't Klooster, Robert T Knight, Ronald M Harper, Anatol Bragin, Itzhak Fried, et al. Ripples on spikes show increased phase-amplitude coupling in mesial temporal lobe epilepsy seizure-onset zones. *Epilepsia*, 57(11):1916–1930, 2016.

[89] Michael X Cohen. Assessing transient cross-frequency coupling in EEG data. *Journal of neuroscience methods*, 168(2):494–499, 2008.

[90] Angela CE Onslow, Rafal Bogacz, and Matthew W Jones. Quantifying phase–amplitude coupling in neuronal network oscillations. *Progress in biophysics and molecular biology*, 105(1-2):49–57, 2011.

[91] Adriano BL Tort, Robert Komorowski, Howard Eichenbaum, and Nancy Kopell. Measuring phase-amplitude coupling between neuronal oscillations of different frequencies. *Journal of neurophysiology*, 104(2):1195–1210, 2010.

[92] Adriano BL Tort, Mark A Kramer, Catherine Thorn, Daniel J Gibson, Yasuo Kubota, Ann M Graybiel, and Nancy J Kopell. Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a t-maze task. *Proceedings of the National Academy of Sciences*, 105(51):20517–20522, 2008.

[93] Jose M Hurtado, Leonid L Rubchinsky, and Karen A Sigvardt. Statistical method for detection of phase-locking episodes in neural oscillations. *Journal of neurophysiology*, 91(4):1883–1898, 2004.

[94] Bahareh Elahian, Mohammed Yeasin, Basanagoud Mudigoudar, James W Wheless, and Abbas Babajani-Feremi. Identifying seizure onset zone from electrocorticographic recordings: a machine learning approach based on phase locking value. *Seizure*, 51:35–42, 2017.

[95] M Dümpelmann. Early seizure detection for closed loop direct neurostimulation devices in epilepsy. *Journal of neural engineering*, 16(4):041001, 2019.

[96] Yun Park, Lan Luo, Keshab K Parhi, and Theoden Netoff. Seizure prediction with spectral power of eeg using cost-sensitive support vector machines. *Epilepsia*, 52(10):1761–1770, 2011.

[97] Zisheng Zhang and Keshab K Parhi. Low-complexity seizure prediction from ieeg/seeg using spectral power and ratios of spectral power. *IEEE transactions on biomedical circuits and systems*, 10(3):693–706, 2015.

[98] Mohammad Khubeb Siddiqui, Ruben Morales-Menendez, Xiaodi Huang, and Nasir Hussain. A review of epileptic seizure detection using machine learning classifiers. *Brain informatics*, 7(1):1–18, 2020.

[99] U Rajendra Acharya, Yuki Hagiwara, and Hojjat Adeli. Automated seizure prediction. *Epilepsy & Behavior*, 88:251–261, 2018.

[100] Yogatheesan Varatharajah, Brent Berry, Jan Cimbalnik, Vaclav Kremen, Jamie Van Gompel, Matt Stead, Benjamin Brinkmann, Ravishankar Iyer, and Gregory Worrell. Integrating artificial intelligence with real-time intracranial EEG monitoring to automate interictal identification of seizure onset zones in focal epilepsy. *Journal of neural engineering*, 15(4):046035, 2018.

[101] Dakun Lai, Xinyue Zhang, Kefei Ma, Zichu Chen, Wenjing Chen, Heng Zhang, Han Yuan, and Lei Ding. Automated detection of high frequency oscillations in intracranial eeg using the combination of short-time energy and convolutional neural networks. *IEEE Access*, 7:82501–82511, 2019.

[102] AV Medvedev, GI Agoureeva, and AM Murro. A long short-term memory neural network for the detection of epileptiform spikes and high frequency oscillations. *Scientific reports*, 9(1):1–10, 2019.

[103] Jeff Craley, Emily Johnson, Christophe C Jouny, David Hsu, Raheel Ahmed, and Archana Venkataraman. SZLoc: A multi-resolution architecture for automated

epileptic seizure localization from scalp EEG. In *Medical Imaging with Deep Learning*, 2022.

[104] Juan C Bulacio, Lara Jehi, Chong Wong, Jorge Gonzalez-Martinez, Prakash Kotagal, Dileep Nair, Imad Najm, and William Bingaman. Long-term seizure outcome after resective surgery in patients evaluated with intracranial electrodes. *Epilepsia*, 53(10):1722–1730, 2012.

[105] L.D. Iasemidis. Epileptic seizure prediction and control. *IEEE Transactions on Biomedical Engineering*, 50(5):549–558, 2003. (doi: 10.1109/TBME.2003.810705).

[106] Florian Mormann, Ralph G. Andrzejak, Christian E. Elger, and Klaus Lehnertz. Seizure prediction: the long and winding road. *Brain*, 130(2):314–333, 09 2006. (doi: 10.1093/brain/awl241).

[107] Levin Kuhlmann, Klaus Lehnertz, Mark P Richardson, Björn Schelter, and Hitten P Zaveri. Seizure prediction—ready for a new era. *Nature Reviews Neurology*, 14(10):618–630, 2018. (doi: 10.1038/s41582-018-0055-2).

[108] L.D. Iasemidis, Deng-Shan Shiau, W. Chaovalitwongse, J.C. Sackellares, P.M. Pardalos, J.C. Principe, P.R. Carney, A. Prasad, B. Veeramani, and K. Tsakalis. Adaptive epileptic seizure prediction system. *IEEE Transactions on Biomedical Engineering*, 50(5):616–627, 2003. (doi: 10.1109/TBME.2003.810689).

[109] Leon D Iasemidis, D-S Shiau, Panos M Pardalos, Wanpracha Chaovalitwongse, K Narayanan, Awadhesh Prasad, Konstantinos Tsakalis, Paul R Carney, and J Chris Sackellares. Long-term prospective on-line real-time seizure prediction. *Clinical neurophysiology*, 116(3):532–544, 2005. (doi: 10.1016/j.clinph.2004.10.013).

[110] Xiaoyan Wei, Lin Zhou, Zhen Zhang, Ziyi Chen, and Yi Zhou. Early prediction of epileptic seizures using a long-term recurrent convolutional network. *Journal of neuroscience methods*, 327:108395, 2019. (doi: 10.1016/j.jneumeth.2019.108395).

[111] Benjamin H Brinkmann, Nicholas M Gregg, and Gregory A Worrell. Seizure forecasting in epilepsy: From computation to clinical practice. *Epilepsy*, pages 451–490, 2021. (doi: 10.1002/9781119431893.ch26).

[112] Rachel E. Stirling, Matias I. Maturana, Philippa J. Karoly, Ewan S. Nurse, Kate McCutcheon, David B. Grayden, Steven G. Ringo, John M. Heasman, Rohan J. Hoare, Alan Lai, Wendyl D'Souza, Udaya Seneviratne, Linda Seiderer, Karen J. McLean, Kristian J. Bulluss, Michael Murphy, Benjamin H. Brinkmann, Mark P. Richardson, Dean R. Freestone, and Mark J. Cook. Seizure forecasting using a novel sub-scalp ultra-long term eeg monitoring system. *Frontiers in Neurology*, Volume 12 - 2021, 2021. (doi: 10.3389/fneur.2021.713794).

[113] Pedro F. Viana, Tal Pal Attia, Mona Nasseri, Jonas Duun-Henriksen, Andrea Biondi, Joel S. Winston, Isabel Pavão Martins, Ewan S. Nurse, Matthias Dümpelmann, Andreas Schulze-Bonhage, Dean R. Freestone, Troels W. Kjaer, Mark P. Richardson, and Benjamin H. Brinkmann. Seizure forecasting using minimally invasive, ultra-long-term subcutaneous electroencephalography: Individualized intrapatient models. *Epilepsia*, 64(S4):S124–S133, 2023. (doi: 10.1111/epi.17252).

[114] Alessio Burrello, Lukas Cavigelli, Kaspar Schindler, Luca Benini, and Abbas Rahimi. Laelaps: An Energy-Efficient Seizure Detection Algorithm from Long-term Human iEEG Recordings without False Alarms. In *2019 Design, Automation Test in Europe Conference Exhibition (DATE)*, pages 752–757, 2019. (doi: 10.23919/DATE.2019.8715186).

[115] Keshab K. Parhi and Zisheng Zhang. Discriminative ratio of spectral power and relative power features derived via frequency-domain model ratio with application to seizure prediction. *IEEE Transactions on Biomedical Circuits and Systems*, 13(4):645–657, 2019.

[116] Anouk van Westrhenen, Tessa Souhoka, Maaike E Ballieux, and Roland D Thijs. Seizure detection devices: Exploring caregivers' needs and wishes. *Epilepsy & Behavior*, 116:107723, 2021.

[117] Samyak Shah, Erie Gonzalez Gutierrez, Jennifer L Hopp, James Wheless, Antonio Gil-Nagel, Gregory L Krauss, and Nathan E Crone. Prospective multicenter study of continuous tonic-clonic seizure monitoring on apple watch in epilepsy monitoring units and ambulatory environments. *Epilepsy & Behavior*, 158:109908, 2024.

[118] Sebastian Thrun and Tom M Mitchell. Lifelong robot learning. *Robotics and autonomous systems*, 15(1-2):25–46, 1995.

[119] Mojtaba Bandarabadi, César A Teixeira, Jalil Rasekhi, and António Dourado. Epileptic seizure prediction using relative spectral power features. *Clinical Neurophysiology*, 126(2):237–248, 2015. (doi: 10.1016/j.clinph.2014.05.022).

[120] Ryan Chen and Keshab K. Parhi. Seizure prediction using convolutional neural networks and sequence transformer networks. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, pages 6483–6486, 2021. (doi: 10.1109/EMBC46164.2021.9629732).

[121] Florian Mormann, Thomas Kreuz, Christoph Rieke, Ralph G Andrzejak, Alexander Kraskov, Peter David, Christian E Elger, and Klaus Lehnertz. On the predictability of epileptic seizures. *Clinical neurophysiology*, 116(3):569–587, 2005.

[122] Maryann D'Alessandro, Rosana Esteller, George Vachtsevanos, Arthur Hinson, Javier Echauz, and Brian Litt. Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. *IEEE transactions on biomedical engineering*, 50(5):603–615, 2003.

[123] Brian Litt and Javier Echauz. Prediction of epileptic seizures. *The Lancet Neurology*, 1(1):22–30, 2002.

[124] Piotr Mirowski, Deepak Madhavan, Yann LeCun, and Ruben Kuzniecky. Classification of patterns of eeg synchronization for seizure prediction. *Clinical neurophysiology*, 120(11):1927–1940, 2009.

[125] Nilufer Ozdemir and Esen Yildirim. Patient specific seizure prediction system using hilbert spectrum and bayesian networks classifiers. *Computational and mathematical methods in medicine*, 2014, 2014.

[126] Ardalan Aarabi and Bin He. Seizure prediction in hippocampal and neocortical epilepsy using a model-based approach. *Clinical Neurophysiology*, 125(5):930–940, 2014.

[127] Nishant Sinha, Justin Dauwels, Marcus Kaiser, Sydney S Cash, M Brandon Westover, Yujiang Wang, and Peter N Taylor. Predicting neurosurgical outcomes in focal epilepsy patients using computational modelling. *Brain*, 140(2):319–332, 2017.

[128] Mona Nasseri, Tal Pal Attia, Boney Joseph, Nicholas M Gregg, Ewan S Nurse, Pedro F Viana, Gregory Worrell, Matthias Dümpelmann, Mark P Richardson, Dean R Freestone, et al. Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning. *Scientific reports*, 11(1):21935, 2021. (doi: 10.1038/s41598-021-01449-2).

[129] Rachel E. Stirling, David B. Grayden, Wendyl D'Souza, Mark J. Cook, Ewan Nurse, Dean R. Freestone, Daniel E. Payne, Benjamin H. Brinkmann, Tal Pal Attia, Pedro F. Viana, Mark P. Richardson, and Philippa J. Karoly. Forecasting seizure likelihood with wearable technology. *Frontiers in Neurology*, Volume 12 - 2021, 2021. (doi: 10.3389/fneur.2021.704060).

[130] Isabell Kiral-Kornek, Subhrajit Roy, Ewan Nurse, Benjamin Mashford, Philippa Karoly, Thomas Carroll, Daniel Payne, Susmita Saha, Steven Baldassano, Terence O'Brien, et al. Epileptic seizure prediction using big data and deep learning: toward a mobile system. *EBioMedicine*, 27:103–111, 2018. (doi: 10.1016/j.ebiom.2017.11.032).

[131] Sha Lu, Lin Liu, Jiuyong Li, Jordan Chambers, Mark J. Cook, and David B. Grayden. Leveraging Channel Coherence in Long-Term iEEG Data for Seizure Prediction. *IEEE Journal of Biomedical and Health Informatics*, 29(8):5541–5548, 2025.

[132] Chang Li, Zhiwei Deng, Rencheng Song, Xiang Liu, Ruobing Qian, and Xun Chen. Eeg-based seizure prediction via model uncertainty learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:180–191, 2022.

[133] Lulu Ge and Keshab K Parhi. Applicability of hyperdimensional computing for seizure prediction using LBP and PSD features from iEEG. In *2023 IEEE 66th International Midwest Symposium on Circuits and Systems (MWSCAS)*, pages 1065–1069. IEEE, 2023. (doi: 10.1109/MWSCAS57524.2023.10405889).

[134] Xuanjie Qiu, Fang Yan, and Haihong Liu. A difference attention resnet-lstm network for epileptic seizure detection using eeg signal. *Biomedical Signal Processing and Control*, 83:104652, 2023.

[135] Mark J Cook, Terence J O'Brien, Samuel F Berkovic, Michael Murphy, Andrew Morokoff, Gavin Fabinyi, Wendyl D'Souza, Raju Yerra, John Archer, Lucas Litewka, et al. Prediction of seizure likelihood with a long-term, implanted seizure advisory system in patients with drug-resistant epilepsy: a first-in-man study. *The Lancet Neurology*, 12(6):563–571, 2013. (doi: 10.1016/s1474-4422(13)70075-9).

[136] Philippa J Karoly, Vikram R Rao, Nicholas M Gregg, Gregory A Worrell, Christophe Bernard, Mark J Cook, and Maxime O Baud. Cycles in epilepsy. *Nature Reviews Neurology*, 17(5):267–284, 2021. (doi: 10.1038/s41582-021-00464-1).

[137] Maxime O Baud, Jonathan K Kleen, Emily A Mirro, Jason C Andrechak, David King-Stephens, Edward F Chang, and Vikram R Rao. Multi-day rhythms modulate seizure risk in epilepsy. *Nature communications*, 9(1):88, 2018. (doi: 10.1038/s41467-017-02577-y).

[138] Maxime O. Baud, Timothée Proix, Nicholas M. Gregg, Benjamin H. Brinkmann, Ewan S. Nurse, Mark J. Cook, and Philippa J. Karoly. Seizure forecasting: Bifurcations in the long and winding road. *Epilepsia*, 64(S4):S78–S98, 2023. (doi: 10.1111/epi.17311).

[139] Petros Koutsouvelis, Bartlomiej Chybowski, Alfredo Gonzalez-Sulser, Shima Abdullateef, and Javier Escudero. Preictal period optimization for deep learning-based epileptic seizure prediction. *Journal of neural engineering*, 21(6):066040, 2024. (doi: 10.1088/1741-2552/ad9ad0).

[140] Gonçalo Costa, César Teixeira, and Mauro F Pinto. Comparison between epileptic seizure prediction and forecasting based on machine learning. *Scientific Reports*, 14(1):5653, 2024. (doi: 10.1038/s41598-024-56019-z).

[141] Eric B. Geller, Tara L. Skarpaas, Robert E. Gross, Robert R. Goodman, Gregory L. Barkley, Carl W. Bazil, Michael J. Berg, Gregory K. Bergey, Sydney S. Cash, Andrew J. Cole, Robert B. Duckrow, Jonathan C. Edwards, Stephan Eisenschenk,

James Fessler, Nathan B. Fountain, Alicia M. Goldman, Ryder P. Gwinn, Christianne Heck, Aamar Herekar, Lawrence J. Hirsch, Barbara C. Jobst, David King-Stephens, Douglas R. Labar, James W. Leiphart, W. Richard Marsh, Kimford J. Meador, Eli M. Mizrahi, Anthony M. Murro, Dileep R. Nair, Katherine H. Noe, Yong D. Park, Paul A. Rutecki, Vicenta Salanova, Raj D. Sheth, Donald C. Shields, Christopher Skidmore, Michael C. Smith, David C. Spencer, Shraddha Srinivasan, William Tatum, Paul C. Van Ness, David G. Vossler, Robert E. Wharen Jr, Gregory A. Worrell, Daniel Yoshor, Richard S. Zimmerman, Kathy Cicora, Felice T. Sun, and Martha J. Morrell. Brain-responsive neurostimulation in patients with medically intractable mesial temporal lobe epilepsy. *Epilepsia*, 58(6):994–1004, 2017. (doi: 10.1111/epi.13740).

[142] Eric B Geller, Tara L Skarpaas, Robert E Gross, Robert R Goodman, Gregory L Barkley, Carl W Bazil, Michael J Berg, Gregory K Bergey, Sydney S Cash, Andrew J Cole, et al. Brain-responsive neurostimulation in patients with medically intractable mesial temporal lobe epilepsy. *Epilepsia*, 58(6):994–1004, 2017.

[143] Gregory K Bergey, Martha J Morrell, Eli M Mizrahi, Alica Goldman, David King-Stephens, Dileep Nair, Shraddha Srinivasan, Barbara Jobst, Robert E Gross, Donald C Shields, et al. Long-term treatment with responsive brain stimulation in adults with refractory partial seizures. *Neurology*, 84(8):810–817, 2015.

[144] Babak Razavi, Vikram R. Rao, Christine Lin, Krzysztof A. Bujarski, Sanjay E. Patra, David E. Burdette, Eric B. Geller, Mesha-Gay M. Brown, Emily A. Johnson, Cornelia Drees, Edward F. Chang, Janet E. Greenwood, Christianne N. Heck, Barbara C. Jobst, Ryder P. Gwinn, Nicole M. Warner, and Casey H. Halpern. Real-world experience with direct brain-responsive neurostimulation for focal onset seizures. *Epilepsia*, 61(8):1749–1757, 2020. (doi: 10.1111/epi.16593).

[145] Karla Ivankovic, Alessandro Principe, Justo Montoya-Gálvez, Linus Manubens-Gil, Riccardo Zucca, Pablo Villoslada, Mara Dierssen, and Rodrigo Rocamora. A novel way to use cross-validation to measure connectivity by machine learning allows epilepsy surgery outcome prediction. *NeuroImage*, 306:120990, 2025. (doi: 10.1016/j.neuroimage.2024.120990).

[146] Benjamin H. Brinkmann, Joost Wagenaar, Drew Abbot, Phillip Adkins, Simone C. Bosshard, Min Chen, Quang M. Tieng, Jialune He, F. J. Muñoz-Almaraz, Paloma Botella-Rocamora, Juan Pardo, Francisco Zamora-Martinez, Michael Hills, Wei Wu, Iryna Korshunova, Will Cukierski, Charles Vite, Edward E. Patterson, Brian Litt, and Gregory A. Worrell. Crowdsourcing reproducible seizure forecasting in human and canine epilepsy. *Brain*, 139(6):1713–1722, 03 2016. (doi: 10.1093/brain/aww045).

[147] Levin Kuhlmann, Philippa Karoly, Dean R Freestone, Benjamin H Brinkmann, Andriy Temko, Alexandre Barachant, Feng Li, Jr. Titericz, Gilberto, Brian W Lang, Daniel Lavery, Kelly Roman, Derek Broadhead, Scott Dobson, Gareth Jones, Qingnan Tang, Irina Ivanenko, Oleg Panichev, Timothée Proix, Michal Náhlík, Daniel B Grunberg, Chip Reuben, Gregory Worrell, Brian Litt, David T J Liley, David B Grayden, and Mark J Cook. Epilepsyecosystem.org: crowd-sourcing reproducible seizure prediction with long-term human intracranial eeg. *Brain*, 141(9):2619–2630, 08 2018. (doi: 10.1093/brain/awy210).

[148] Mohamed L Seghier and Cathy J Price. Interpreting and utilising intersubject variability in brain function. *Trends in cognitive sciences*, 22(6):517–530, 2018. (doi: 10.1016/j.tics.2018.03.003).

[149] Joseph West, Zahra Dasht Bozorgi, Jeffrey Herron, Howard J Chizeck, Jordan D Chambers, and Lyra Li. Machine learning seizure prediction: one problematic but accepted practice. *Journal of Neural Engineering*, 20(1):016008, 2023.

[150] Alex Van Esbroeck, Landon Smith, Zeeshan Syed, Satinder Singh, and Zahi Karam. Multi-task seizure detection: addressing intra-patient variation in seizure morphologies. *Machine Learning*, 102(3):309–321, 2016. (doi: 10.1007/s10994-015-5519-7).

[151] Sebastian Böttcher, Elisa Bruno, Nino Epitashvili, Matthias Dümpelmann, Nicolas Zabler, Martin Glasstetter, Valentina Ticcinelli, Sarah Thorpe, Simon Lees, Kristof Van Laerhoven, et al. Intra-and inter-subject perspectives on the detection of focal onset motor seizures in epilepsy patients. *Sensors*, 22(9):3318, 2022. (doi: 10.3390/s22093318).

[152] Yun Zhao, David B Grayden, Mario Boley, Yueyang Liu, Philippa J Karoly, Mark J Cook, and Levin Kuhlmann. Cortical stability and chaos during focal seizures: insights from inference-based modeling. *Journal of Neural Engineering*, 22(3):036021, 2025. (doi: 10.1088/1741-2552/add83f).

[153] Maria Luisa Saggio, Dakota Crisp, Jared M Scott, Philippa Karoly, Levin Kuhlmann, Mitsuyoshi Nakatani, Tomohiko Murai, Matthias Dümpelmann, Andreas Schulze-Bonhage, Akio Ikeda, Mark Cook, Stephen V Gliske, Jack Lin, Christophe Bernard, Viktor Jirsa, and William C Stacey. A taxonomy of seizure dynamotypes. *eLife*, 9:e55632, jul 2020. (doi: 10.7554/eLife.55632).

[154] Lulu Ge and Keshab K. Parhi. Applicability of hyperdimensional computing to seizure detection. *IEEE Open Journal of Circuits and Systems*, 3:59–71, 2022. (doi: 10.1109/OJCAS.2022.3163075).

[155] Dionisije Sopic, Tomas Teijeiro, David Atienza, Amir Aminifar, and Philippe Ryvlin. Personalized seizure signature: An interpretable approach to false alarm reduction for long-term epileptic seizure detection. *Epilepsia*, 64(S4):S23–S33, 2023. (doi: 10.1111/epi.17176).

[156] Sai Sanjay Balaji and Keshab K Parhi. Seizure onset zone (SOZ) identification using effective brain connectivity of epileptogenic networks. *Journal of Neural Engineering*, 21(3):036053, jun 2024. doi: 10.1088/1741-2552/ad5938.

[157] John M. Bernabei, Adam Li, Andrew Y. Revell, Rachel J. Smith, Kristin M. Gunnarsdottir, Ian Z. Ong, Kathryn A. Davis, Nishant Sinha, Sridevi Sarma, and Brian Litt. "HUP iEEG Epilepsy Dataset", 2023.

[158] Alan C Evans, Micheline Kamber, DL Collins, and David MacDonald. An MRI-based probabilistic atlas of neuroanatomy. *Magnetic resonance scanning and epilepsy*, pages 263–274, 1994.

[159] Jerome Engel Jr. Update on surgical treatment of the epilepsies: summary of the second international palm desert conference on the surgical treatment of the epilepsies (1992). *Neurology*, 43(8):1612–1612, 1993.

[160] Thomas M Cover. *Elements of information theory*. John Wiley & Sons, 1999.

[161] Andrea Brovelli, Daniel Chicharro, Jean-Michel Badier, Huifang Wang, and Viktor Jirsa. Characterization of cortical networks and corticocortical functional connectivity mediating arbitrary visuomotor mapping. *Journal of Neuroscience*, 35(37):12643–12658, 2015.

[162] Sandeep Avvaru and Keshab K Parhi. Effective Brain Connectivity Extraction by Frequency-Domain Convergent Cross-Mapping (FDCCM) and its Application in Parkinson's Disease Classification. *IEEE Transactions on Biomedical Engineering*, 2023.

[163] George Sugihara, Robert May, Hao Ye, Chih-hao Hsieh, Ethan Deyle, Michael Fogarty, and Stephan Munch. Detecting causality in complex ecosystems. *science*, 338(6106):496–500, 2012.

[164] David A Dickey and Wayne A Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: journal of the Econometric Society*, pages 1057–1072, 1981.

[165] Sebastian Wallot and Dan Mønster. Calculation of average mutual information (AMI) and false-nearest neighbors (FNN) for the estimation of embedding parameters of multidimensional time series in MATLAB. *Frontiers in psychology*, 9:1679, 2018.

[166] Etienne Combrisson et al. Group-level inference of information-based measures for the analyses of cognitive brain networks from neurophysiological data. *NeuroImage*, 258:119347, September 2022.

[167] Etienne Combrisson et al. Frites: A Python package for functional connectivity analysis and group-level statistics of neurophysiological data. *Journal of Open Source Software*, 7(79):3842, 2022.

[168] Alex Bavelas. Communication patterns in task-oriented groups. *The journal of the acoustical society of America*, 22(6):725–730, 1950.

[169] Giorgio Fagiolo. Clustering in complex directed networks. *Physical Review E*, 76(2):026107, 2007.

[170] Vito Latora and Massimo Marchiori. Economic small-world behavior in weighted networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 32:249–263, 2003.

[171] Douglas R White and Stephen P Borgatti. Betweenness centrality measures for directed graphs. *Social networks*, 16(4):335–346, 1994.

[172] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107–117, 1998.

[173] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

[174] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[175] Preya Shah, John M Bernabei, Lohith G Kini, Arian Ashourvan, Jacqueline Boccanfuso, Ryan Archer, Kelly Oechsel, Sandhitsu R Das, Joel M Stein, Timothy H Lucas, et al. High interictal connectivity within the resection zone is associated with favorable post-surgical outcomes in focal epilepsy patients. *NeuroImage: Clinical*, 23:101908, 2019.

[176] F. Pedregosa et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[177] T Head et al. scikit-optimize/scikit-optimize, September 2020.

[178] Hanchuan Peng, Fuhui Long, and Chris Ding. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on pattern analysis and machine intelligence*, 27(8):1226–1238, 2005.

[179] Tingting Xu, Massoud Stephane, and Keshab K Parhi. Abnormal neural oscillations in schizophrenia assessed by spectral power ratio of MEG during word processing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(11):1148–1158, 2016.

[180] Lucia R Quitadamo, Elaine Foley, Roberto Mai, Luca De Palma, Nicola Specchio, and Stefano Seri. EPINETLAB: a software for seizure-onset zone identification from intracranial EEG signal in epilepsy. *Frontiers in neuroinformatics*, 12:45, 2018.

[181] Yao Miao, Yasushi Iimura, Hidenori Sugano, Kosuke Fukumori, and Toshihisa Tanaka. Seizure onset zone identification using phase-amplitude coupling and multiple machine learning approaches for interictal electrocorticogram. *Cognitive Neurodynamics*, 17(6):1591–1607, 2023.

[182] Sai Sanjay Balaji and Keshab K. Parhi. Seizure onset zone (soz) identification from interictal intracranial electroencephalogram. *Scientific Reports*, 15(1), 11 2025.

[183] John M Bernabei, Nishant Sinha, T Campbell Arnold, Erin Conrad, Ian Ong, Akash R Pattnaik, Joel M Stein, Russell T Shinohara, Timothy H Lucas, Dani S Bassett, et al. Normative intracranial eeg maps epileptogenic tissues in focal epilepsy. *Brain*, 145(6):1949–1961, 2022.

[184] Kristin Gunnarsdottir, Adam Li, Rachel Smith, Joon Kang, Anna Korzeniewska, Nathan Crone, Adam Rouse, Jennifer Cheng, Michael Kinsman, Patrick Landazuri, Utku Uysal, Carol Ulloa, Nathaniel Cameron, Iahn Cajigas, Jonathan Jagid, Andres Kanner, Turki Elarjani, Manuel Bicchi, Sara Inati, Kareem Zaghloul, Varina Boerwinkle, Sarah Wyckoff, Niravkumar Barot, Jorge Gonzalez-Martinez, and Sridevi Sarma. Epilepsy-iEEG-Interictal-Multicenter-Dataset, 2023.

[185] Kristin M Gunnarsdottir, Adam Li, Rachel J Smith, Joon-Yi Kang, Anna Korzeniewska, Nathan E Crone, Adam G Rouse, Jennifer J Cheng, Michael J Kinsman, Patrick Landazuri, et al. Source-sink connectivity: A novel interictal eeg marker for seizure localization. *Brain*, 145(11):3901–3915, 2022.

[186] Adam Li, Chester Huynh, Zachary Fitzgerald, Iahn Cajigas, Damian Brusko, Jonathan Jagid, Angel O Claudio, Andres M Kanner, Jennifer Hopp, Stephanie Chen, et al. Neural fragility as an eeg marker of the seizure onset zone. *Nature neuroscience*, 24(10):1465–1474, 2021.

[187] Merritt W Brown III, Brenda E Porter, Dennis J Dlugos, Jeff Keating, Andrew B Gardner, Phillip B Storm Jr, and Eric D Marsh. Comparison of novel computer

detectors and human performance for spike detection in intracranial eeg. *Clinical neurophysiology*, 118(8):1744–1752, 2007.

[188] R. Zelmann, F. Mari, J. Jacobs, M. Zijlmans, R. Chander, and J. Gotman. Automatic detector of high frequency oscillations for human recordings with macroelectrodes. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*, pages 2329–2333, 2010.

[189] Yipeng Zhang, Lawrence Liu, Yuanyi Ding, Xin Chen, Tonmoy Monsoor, Atsuro Daida, Shingo Oana, Shaun Hussain, Raman Sankar, Aria Fallah, et al. PyHFO: lightweight deep learning-powered end-to-end high-frequency oscillations analysis application. *Journal of Neural Engineering*, 21(3):036023, 2024.

[190] Adriano B. L. Tort, Mark A. Kramer, Catherine Thorn, Daniel J. Gibson, Yasuo Kubota, Ann M. Graybiel, and Nancy J. Kopell. Dynamic cross-frequency couplings of local field potential oscillations in rat striatum and hippocampus during performance of a t-maze task. *Proceedings of the National Academy of Sciences*, 105(51):20517–20522, 2008.

[191] Eric D Marsh, Bradley Peltzer, Merritt W Brown III, Courtney Wusthoff, Phillip B Storm Jr, Brian Litt, and Brenda E Porter. Interictal eeg spikes identify the region of electrographic seizure onset in some, but not all, pediatric epilepsy patients. *Epilepsia*, 51(4):592–601, 2010.

[192] Evelien E Geertsema, Gerhard H Visser, Demetrios N Velis, Steven P Claus, Maeike Zijlmans, and Stiliyan N Kalitzin. Automated seizure onset zone approximation based on nonharmonic high-frequency oscillations in human interictal intracranial eegs. *International journal of neural systems*, 25(05):1550015, 2015.

[193] Haiteng Jiang, Vasileios Kokkinos, Shuai Ye, Alexandra Urban, Anto Bagić, Mark Richardson, and Bin He. Interictal seeg resting-state connectivity localizes the seizure onset zone and predicts seizure outcome. *Advanced Science*, 9(18):2200887, 2022.

[194] Lauri Rekola, Maria Peltola, Jukka Vanhanen, Juha Wilenius, Eeva-Liisa Metsähonkala, Leena Kämppi, Leena Lauronen, and Päivi Nevalainen. Combined

value of interictal markers and stimulated seizures to estimate the seizure onset zone in stereoelectroencephalography. *Epilepsia*, 65(10):2946–2958, 2024.

[195] Sai Sanjay Balaji, Zisheng Zhang, Zhiyi Sha, Thomas R Henry, and Keshab K Parhi. Patient-specific long-term seizure prediction via multi-model classification. *Journal of Neural Engineering*, 2025. DOI: 10.1088/1741-2552/ae1875.

[196] J Engel. *Surgical treatment of the epilepsies*. Lippincott Williams and Wilkins, Philadelphia, PA, 2 edition, May 1993. (https://lccn.loc.gov/85043520).

[197] Eugene I Shih, Ali H Shoeb, and John V Guttag. Sensor selection for energy-efficient ambulatory medical monitoring. In *Proceedings of the 7th international conference on mobile systems, applications, and services*, pages 347–358, 2009. (doi: 10.1145/1555816.1555851).

[198] Jonas Duun-Henriksen, Troels Wesenberg Kjaer, Rasmus Elsborg Madsen, Line Sofie Remvig, Carsten Eckhart Thomsen, and Helge Bjarup Dissing Sorensen. Channel selection for automatic seizure detection. *Clinical Neurophysiology*, 123(1):84–92, 2012. (doi: 10.1016/j.clinph.2011.06.001).

[199] Shan Liu, Jiang Wang, Shanshan Li, and Lihui Cai. Epileptic seizure detection and prediction in eegs using power spectra density parameterization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:3884–3894, 2023. (doi: 10.1109/TNSRE.2023.3317093).

[200] Rabia Tutuk and Reyhan Zengin. Epileptic seizure detection combining power spectral density and high-frequency oscillations. *International Journal of Applied Mathematics Electronics and Computers*, 11(2):117–127, 2023. (doi: 10.18100/ijamec.1229907).

[201] Tingting Xu, Massoud Stephane, and Keshab K. Parhi. Abnormal neural oscillations in schizophrenia assessed by spectral power ratio of meg during word processing. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 24(11):1148–1158, 2016. (doi: 10.1109/TNSRE.2016.2551700).

[202] Zisheng Zhang and Keshab K. Parhi. Seizure prediction using polynomial svm classification. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 5748–5751, 2015. (doi: 10.1109/EMBC.2015.7319698).

[203] Keshab K Parhi and Zisheng Zhang. Discriminative ratio of spectral power and relative power features derived via frequency-domain model ratio with application to seizure prediction. *IEEE transactions on biomedical circuits and systems*, 13(4):645–657, 2019. (doi: 10.1109/TBCAS.2019.2917184).

[204] Keshab K Parhi and Zisheng Zhang. Method and apparatus for prediction and detection of seizure activity, October 1 2019. US Patent 10,426,365.

[205] Zisheng Zhang and Keshab K Parhi. Muse: Minimum uncertainty and sample elimination based binary feature selection. *IEEE Transactions on Knowledge and Data Engineering*, 31(9):1750–1764, 2018. (doi: 10.1109/TKDE.2018.2865778).

[206] M. Winterhalder, T. Maiwald, H.U. Voss, R. Aschenbrenner-Scheibe, J. Timmer, and A. Schulze-Bonhage. The seizure prediction characteristic: a general framework to assess and compare seizure prediction methods. *Epilepsy Behavior*, 4(3):318–325, 2003. (doi: 10.1016/S1525-5050(03)00105-7).

[207] Joana Batista, Mauro F Pinto, Mariana Tavares, Fábio Lopes, Ana Oliveira, and César Teixeira. Eeg epilepsy seizure prediction: the post-processing stage as a chronology. *Scientific Reports*, 14(1):407, 2024. (doi: 10.1038/s41598-023-50609-z).

[208] Florian Mormann, Klaus Lehnertz, Peter David, and Christian E Elger. Mean phase coherence as a measure for phase synchronization and its application to the eeg of epilepsy patients. *Physica D: Nonlinear Phenomena*, 144(3-4):358–369, 2000.

[209] Simon S Haykin. *Adaptive filter theory*. Pearson Education India, 2002.

[210] Luigi Chisci, Antonio Mavino, Guido Perferi, Marco Sciandrone, Carmelo Anile, Gabriella Colicchio, and Filomena Fuggetta. Real-time epileptic seizure prediction using AR models and support vector machines. *IEEE Transactions on Biomedical Engineering*, 57(5):1124–1132, 2010.

[211] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Routledge, May 2013.

Appendix A

Supplementary Information for SoZ Identification

A.1 HUP dataset: Patient Information

The two chapters dedicated to SOZ identification are based primarily on analysis of a dataset collected at the Hospital of the University of Pennsylvania (HUP) [157]. Table A.1 presents a synopsis of patient details, including epilepsy type, iEEG type, and SoZ location, adapted from [183]. Patients with an Engel I outcome, used for model training, are indicated with an asterisk (*).

Table A.1: Participant Demographics and Clinical Information

Subject ID	Age	Sex	Outcome	Engel	Therapy	Implant	Target	Lesion	Onset Age
HUP060	42	F	F	3A	A	SEEG	FRONTAL	n/a	12
HUP064*	21	M	S	1D	R	ECOG	FRONTAL	L	3
HUP065*	36	M	S	1B	R	ECOG	TEMPORAL	L	2
HUP070*	33	M	S	1B	R	ECOG	FP	NL	12
HUP074*	25	F	S	1C	R	ECOG	TEMPORAL	L	5
HUP075	57	F	F	4A	R	ECOG	TEMPORAL	NL	52
HUP080	n/a	F	F	2C	R	ECOG	TEMPORAL	NL	n/a
HUP082*	56	F	S	1A	R	ECOG	TEMPORAL	L	34

Continued on next page

Table A.1 – continued from previous page

Subject ID	Age	Sex	Outcome	Engel	Therapy	Implant	Target	Lesion	Onset Age
HUP086	25	F	S	2A	R	ECOG	TEMPORAL	NL	17
HUP087*	24	M	S	1D	R	ECOG	FRONTAL	L	19
HUP088*	35	F	S	1D	R	ECOG	TEMPORAL	L	1
HUP089	n/a	n/a	S	1B	R	ECOG	n/a	n/a	n/a
HUP094*	48	F	S	1B	R	ECOG	TEMPORAL	NL	20
HUP097*	n/a	n/a	S	1D	R	ECOG	n/a	n/a	n/a
HUP105*	39	M	S	1A	R	ECOG	TEMPORAL	L	27
HUP106*	45	F	S	1B	R	ECOG	TEMPORAL	NL	24
HUP107*	36	M	S	1A	R	ECOG	TEMPORAL	NL	5
HUP111*	40	F	S	1B	R	ECOG	TEMPORAL	NL	28
HUP112	21	F	F	3A	A	SEEG	FRONTAL	L	0
HUP114	n/a	n/a	F	3A	A	ECOG	n/a	n/a	n/a
HUP116	59	F	S	1A	A	SEEG	MTL	L	42
HUP117	39	M	S	1A	R	SEEG	TEMPORAL	L	12
HUP123	n/a	n/a	S	1A	R	ECOG	n/a	n/a	n/a
HUP126	n/a	n/a	S	1A	A	ECOG	n/a	n/a	n/a
HUP130*	46	F	S	1B	A	SEEG	MFL	NL	20
HUP132	n/a	n/a	F	3A	A	SEEG	n/a	n/a	n/a
HUP133	52	F	F	3A	A	SEEG	MTL	NL	47
HUP134*	n/a	n/a	S	1B	R	SEEG	n/a	n/a	n/a
HUP135	37	M	F	2A	A	SEEG	MTL	NL	34
HUP138	38	M	F	4A	A	SEEG	MTL	L	29
HUP139*	20	M	S	1A	A	SEEG	PARIETAL	L	0
HUP140	47	F	S	1B	A	SEEG	MTL	NL	26
HUP141*	30	M	F	1C	A	SEEG	MTL	NL	14
HUP142*	n/a	n/a	F	1D	A	SEEG	n/a	n/a	n/a
HUP144*	n/a	n/a	S	1D	R	SEEG	n/a	n/a	n/a
HUP146*	n/a	n/a	S	1A	R	SEEG	n/a	n/a	n/a
HUP148*	n/a	n/a	S	1A	A	SEEG	n/a	n/a	n/a
HUP150	17	M	S	1B	A	SEEG	INSULAR	L	4
HUP151	n/a	M	S	2A	A	SEEG	MFL	NL	n/a
HUP157*	25	M	S	1B	A	SEEG	MTL	NL	16
HUP158	32	M	F	3A	A	SEEG	INSULAR	NL	7
HUP160*	n/a	n/a	S	1A	R	SEEG	n/a	n/a	n/a
HUP162	n/a	n/a	F	3A	A	SEEG	n/a	n/a	n/a

Continued on next page

Table A.1 – continued from previous page

Subject ID	Age	Sex	Outcome	Engel	Therapy	Implant	Target	Lesion	Onset Age
HUP163*	42	F	S	1D	A	SEEG	MTL	NL	32
HUP164*	34	F	S	1D	A	SEEG	MTL	L	14
HUP165	21	F	S	3A	A	SEEG	MTL	NL	12
HUP166	n/a	n/a	S	3A	R	SEEG	n/a	n/a	n/a
HUP171	50	M	F	2A	A	SEEG	FRONTAL	NL	6
HUP172	28	F	F	2A	A	SEEG	FRONTAL	NL	2
HUP173	24	F	S	1A	R	SEEG	TEMPORAL	L	12
HUP177*	42	F	S	1A	R	SEEG	TEMPORAL	NL	5
HUP179	20	F	S	3A	R	SEEG	FRONTAL	L	13
HUP180	28	F	S	1A	A	SEEG	FRONTAL	L	2
HUP181	31	F	S	3A	A	SEEG	TEMPORAL	L	16
HUP185*	38	M	S	1A	A	SEEG	MTL	L	9
HUP187	n/a	n/a	S	2A	A	SEEG	n/a	n/a	n/a
HUP188	24	F	F	3A	R	SEEG	FRONTAL	L	9
HUP190	25	M	F	3A	R	SEEG	MTL	NL	12

Abbreviations: F (Female), M (Male); S (Successful/Seizure-Free), F (Failed/Not Seizure-Free); R (Resection), A (Ablation); L (Lesional), NL (Non-Lesional); FP (Frontoparietal), MTL (Mesial Temporal Lobe), MFL (Mesial Frontal Lobe). * Subject included in the training set.

A.2 Selection of time window for feature extraction

In our endeavor to select the most appropriate time window for feature extraction in Chapter 3, we considered a variety of preictal, ictal, and interictal time intervals, encompassing solely the ictal 30-second and 45-second windows, 5-minute interictal window, preictal windows of 30 seconds, 1-minute, and 2 minutes, as well as combined preictal and ictal windows lasting 10, 20, and 30 seconds and centered around the onset. The results derived from the FD-CCCM-based EC graph model are detailed in figures A.1 and A.2. Although the specific values changed slightly for the other two measures, in all three cases, the 30-second window, positioned around the seizure onset and comprising 15 seconds of preictal signal followed by 15 seconds of ictal signal, showed a higher average t-value across the ten features for the two classes. All ten features also exhibited a significant distinction between the two classes for FD-CCM and nine out of ten features for the other two measures.

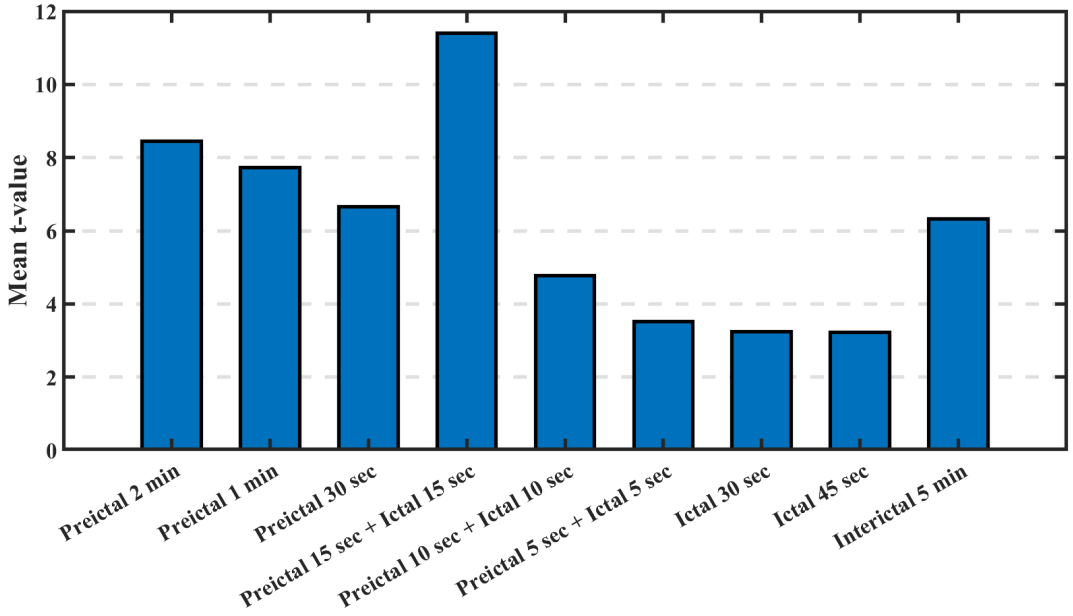

Figure A.1: A bar chart depicting the average t-values of graph centrality features that exhibit statistical significance ($p < 0.05$) in Welch's t-test. This test compares SOZ versus non-SOZ classes using an FD-CCM-based EC model across various time windows. A higher t-value suggests enhanced discriminability.

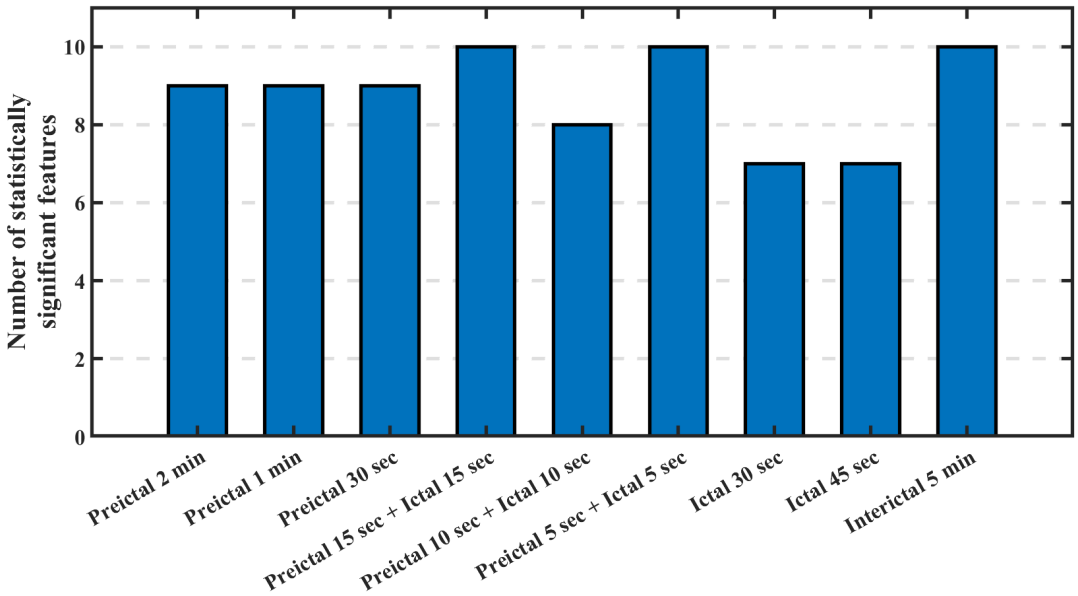

Figure A.2: A bar chart illustrating the count of graph centrality features with statistical significance ($p < 0.05$) in Welch's t-test. This test compares SOZ versus non-SOZ classes using an FD-CCM-based EC model across various time windows.

Appendix B

Supplementary Information for Seizure Prediction

B.1 Algorithms for Seizure Model Creation

This section provides the formal pseudocode for the three-stage process used to create the patient-specific seizure models. The first stage, Feature Ranking and Weighting per Realization, is detailed in Algorithm 1. This step employs the MUSE algorithm over multiple data realizations to identify and weight discriminative preictal features for a single seizure event. The second stage, Aggregation of Feature Rankings Across Realizations, is described in Algorithm 2. This step synthesizes the results from the multiple realizations to create a single, stable, representative feature set for that seizure. Finally, the third stage, Seizure Clustering into Models, is outlined in Algorithm 3. This algorithm groups individual seizures into distinct models based on the similarity of their representative feature sets, forming the basis for the multi-model classification framework.

Algorithm 1 Feature Ranking and Weighting per Realization

Input: A seizure episode s, preictal feature data $\mathcal{X}_{\text{pre}}^{(s)}$, interictal feature pool $\mathcal{X}_{\text{int\_pool}}$, number of realizations $R = 100$, survival threshold $\theta = 0.10$.

Output: For each realization i, a set of top-ranked features $\mathcal{F}_i^{(s)}$ and their normalized weights $\mathcal{R}_i^{(s)}$.

for $i = 1$ **to** R **do**

  Randomly sample interictal data $\mathcal{X}_{\text{int},i}^{(s)}$ from $\mathcal{X}_{\text{int\_pool}}$.

  Construct combined feature matrix: $\mathcal{X}_i^{(s)} \leftarrow \mathcal{X}_{\text{pre}}^{(s)} \cup \mathcal{X}_{\text{int},i}^{(s)}$.

  Initialize surviving sample sets S_{pre} and S_{int}.

  Initialize $\mathcal{F}_i^{(s)} \leftarrow \emptyset$, and step counter $j \leftarrow 1$.

  while $\frac{|S_{\text{pre}}|}{|\mathcal{X}_{\text{pre}}^{(s)}|} \geq \theta$ **or** $\frac{|S_{\text{int}}|}{|\mathcal{X}_{\text{int},i}^{(s)}|} \geq \theta$ **do**

    Select next best feature $F_{j,i}^{(s)}$ using one step of the MUSE algorithm on $\mathcal{X}_i^{(s)}$.

    Update S_{pre} and S_{int} by pruning samples based on $F_{j,i}^{(s)}$.

    Append $F_{j,i}^{(s)}$ to $\mathcal{F}_i^{(s)}$.

    $j \leftarrow j + 1$.

  end while

  Identify the dominant (early-stopping) class.

  Let y_j be the survival percentage of the dominant class after step j.

  for each feature $F_{j,i}^{(s)} \in \mathcal{F}_i^{(s)}$ **do**

    Compute raw reduction: $\Delta y_{j,i}^{(s)} \leftarrow y_{j-1,i}^{(s)} - y_{j,i}^{(s)}$.

  end for

  Let $N_{\text{top},i}^{(s)} \leftarrow |\mathcal{F}_i^{(s)}|$.

  for $j = 1$ **to** $N_{\text{top},i}^{(s)}$ **do**

    Compute normalized weight: $r_{j,i}^{(s)} \leftarrow \frac{\Delta y_{j,i}^{(s)}}{\sum_{k=1}^{N_{\text{top},i}^{(s)}} \Delta y_{k,i}^{(s)}}$.

    Store $r_{j,i}^{(s)}$ in $\mathcal{R}_i^{(s)}$.

  end for

end for

return $\{\mathcal{F}_i^{(s)}, \mathcal{R}_i^{(s)}\}_{i=1}^{R}$

Algorithm 2 Aggregation of Feature Rankings Across Realizations

Input: Ranked features and weights for all R realizations: $\{\mathcal{F}_i^{(s)}, \mathcal{R}_i^{(s)}\}_{i=1}^{R}$.
Output: A single representative feature set for the seizure, $\hat{\mathcal{F}}^{(s)}$.
Construct the set of unique features: $\mathcal{U}^{(s)} \leftarrow \bigcup_{i=1}^{R} \mathcal{F}_i^{(s)}$.
Initialize aggregated weights $w^{(s)}(f) \leftarrow 0$ for all $f \in \mathcal{U}^{(s)}$.
for each feature $f \in \mathcal{U}^{(s)}$ **do**
    Calculate aggregated weight: $w^{(s)}(f) \leftarrow \sum_{i=1}^{R} \sum_{j=1}^{|\mathcal{F}_i^{(s)}|} \mathbb{I}\left[F_{j,i}^{(s)} = f\right] \cdot r_{j,i}^{(s)}$.
end for
Determine max number of features: $\hat{N}^{(s)} \leftarrow \max_i |\mathcal{F}_i^{(s)}|$.
Rank features in $\mathcal{U}^{(s)}$ in descending order based on $w^{(s)}(f)$.
Select the top $\hat{N}^{(s)}$ features from the ranked list to form $\hat{\mathcal{F}}^{(s)}$.
return $\hat{\mathcal{F}}^{(s)}$

Algorithm 3 Seizure Clustering into Models

Input: Representative feature sets for all seizures: $\{\hat{\mathcal{F}}^{(s)}\}_{s=1}^{S_{total}}$, overlap threshold $\tau = 0.8$.
Output: A set of seizure clusters (models), $\mathcal{C} = \{C_1, \ldots, C_M\}$.
Initialize an empty list of clusters, $\mathcal{C} \leftarrow \emptyset$.
for each seizure $s = 1$ **to** S_{total} **do**
    Set `assigned_to_cluster` $\leftarrow$ `false`.
    for each existing cluster $C_k \in \mathcal{C}$ **do**
        Pick a representative seizure $s' \in C_k$.
        Calculate feature overlap: Overlap $\leftarrow \frac{|\hat{\mathcal{F}}^{(s)} \cap \hat{\mathcal{F}}^{(s')}|}{\max\left(|\hat{\mathcal{F}}^{(s)}|, |\hat{\mathcal{F}}^{(s')}|\right)}$.
        if Overlap $\geq \tau$ **then**
            Add seizure s to cluster C_k.
            `assigned_to_cluster` $\leftarrow$ `true`.
            break
        end if
    end for
    if `assigned_to_cluster` is `false` **then**
        Create a new cluster $\{s\}$ and add it to $\mathcal{C}$.
    end if
end for
return $\mathcal{C}$

B.2 Example: Seizure Modeling for Subject (SID 22) Using MUSE and Feature Clustering

A detailed walk-through of the seizure-specific modeling pipeline applied to Subject SID 22 is presented in this section. The dataset spans 19 days of iEEG recordings, including 5 days on which seizures occurred. Using the MUSE algorithm and a clustering-based modeling approach, we identify subject-specific seizure models and validate them in a five-fold cross-validation setting. The data for each fold consists of one seizure day and one or multiple 24-hour recordings from seizure-free days. Table B.1 describes the metadata for all seizures recorded from subject SID22, including seizure type, date, and the corresponding onset and early propagation electrode groups.

Table B.1: Seizure metadata for SID 22, including onset and early propagation electrodes

Sz ID	**Date**	**Type**	**Onset Electrodes**	**Early Propagation Electrodes**
Sz1	Day 3	CPS	RMT 1	RPT1–4, RAT1–2
Sz2	Day 5	CPS	LAT1–4, LMT1–4	LPT1–4
Sz3	Day 6	SPS	LAT1–3, LMT1	LPT1–4, LAT4, LMT2–4
Sz4	Day 7	SPS	LAT1–3, LMT1	LPT1–4, LAT4, LMT2–4
Sz5	Day 20	CPS–GTC	RPT1–4	RAT1–4, RMT1–4, RAP1–4, RPP1–4
Sz6	Day 20	CPS–GTC	RPT1–4	RAT1–4, RMT1–4, RAP1–4, RPP1–4

B.2.1 Cross-Validation Setup (5 Folds)

Subject SID 22 had seizures on five distinct days. In the proposed five-fold cross-validation scheme, each fold uses one seizure day as the test set and the other four seizure days for training. The remaining 14 seizure-free days are distributed across folds to simulate prospective evaluation. One possible data split is described in Table B.2. For the next

subsections, assume that the algorithm is tested on *Fold 1* after being trained on the other folds.

Table B.2: Five-fold cross-validation structure for SID 22

Fold	**Test Day**	**Test Seizure(s)**	**Assigned Non-Seizure Days**	**Total Test Duration**
1	Day 3	Sz1	Days 1, 2, 4	1 seizure + 3 non-seizure days
2	Day 5	Sz2	Days 6, 7, 8	1 seizure + 3 non-seizure days
3	Day 6	Sz3	Days 9, 10, 11	1 seizure + 3 non-seizure days
4	Day 7	Sz4	Days 12, 13, 14	1 seizure + 3 non-seizure days
5	Day 20	Sz5, Sz6	Days 15, 16	1 seizure + 2 non-seizure days

B.2.2 Feature Selection Using MUSE: Example using Sz2

For seizure Sz2, MUSE is applied 100 times, each time using a fixed preictal window paired with randomly sampled, duration-matched interictal segments. In each realization, features are ranked based on their ability to reduce conditional entropy, and the top $N_{\text{top},i}^{(s)}$ features are selected. The final number of features retained, $\hat{N}^{(s)}$, is set to 7, corresponding to the maximum number of features selected in any realization. A representative excerpt showing selected features and their associated entropy values across realizations is shown in B.3. Each feature's overall importance is computed by summing its conditional entropy contributions across all realizations in which it appeared. Table B.4 lists the top 7 features and their entropy-weighted scores.

Table B.3: Excerpt of selected features and entropies across MUSE realizations for Sz2

Realization	**Selected Features**	**Corresponding weights**
1	LAT1_Abs_Theta, LAT2_Rel_Alpha, LAT3_Ratio_(Beta/Theta), LMT2_Ratio_(Alpha/Gamma1), LAT4_Rel_Theta	0.23, 0.22, 0.21, 0.20, 0.19
2	LAT1_Abs_Theta, LMT1_Rel_Gamma1, LMT2_Ratio_(Alpha/Gamma1), LAT4_Rel_Theta	0.27, 0.26, 0.23, 0.21
⋮	⋮	⋮
100	LAT2_Rel_Alpha, LAT3_Ratio_(Beta/Theta), LMT2_Ratio_(Alpha/Gamma1), LMT3_Abs_Alpha	0.28, 0.25, 0.24, 0.22

Table B.4: Top features selected for Sz2 based on entropy-weighted importance across 100 realizations

Feature	**# Realizations Selected**	**Cumulative Weight** $w(f)$
LMT2_Ratio_(Alpha/Gamma1)	90	34.2
LAT2_Rel_Alpha	85	31.5
LAT3_Ratio_(Beta/Theta)	78	28.1
LAT1_Abs_Theta	72	25.2
LMT1_Rel_Gamma1	65	22.1
LAT4_Rel_Theta	60	19.8
LMT3_Abs_Alpha	58	18.6

These features are selected as the representative feature set $\hat{\mathcal{F}}^{(s)}$ for Sz2. They consistently appeared across multiple realizations and contributed significantly to class separation between preictal and interictal windows. A similar entropy-weighted aggregation strategy is applied to determine representative feature sets for all other seizures for the subject. Table B.5 lists the top-ranked representative features for all seizures of SID22.

Table B.5: Top-ranked feature sets per seizure after MUSE aggregation

Sz ID	**Representative Features $\hat{\mathcal{F}}^{(s)}$**
Sz2	LAT1_Abs_Theta, LAT2_Rel_Alpha, LAT3_Ratio_(Beta/Theta), LMT1_Abs_Delta, LMT3_Abs_Alpha, LAT4_Rel_Theta, LAT5_Rel_Delta
Sz3	LAT2_Rel_Alpha, LAT3_Ratio_(Beta/Theta), LMT1_Abs_Delta, LMT2_Ratio_(Alpha/Gamma1), LMT3_Abs_Alpha, LAT4_Rel_Theta
Sz4	LAT2_Rel_Alpha, LAT3_Ratio_(Beta/Theta), LAT4_Rel_Theta, LMT1_Abs_Delta, LMT2_Ratio_(Alpha/Gamma1), LMT3_Rel_Gamma2
Sz5	RPT1_Abs_Gamma2, RPT2_Rel_Gamma4, RPT3_Ratio_(Gamma2/Theta), RPT4_Abs_Alpha, RPT2_Rel_Gamma3, RPT1_Rel_Beta, RPT3_Abs_Theta
Sz6	RPT1_Abs_Gamma3, RPT2_Rel_Gamma4, RPT3_Ratio_(Gamma2/Theta), RPT3_Abs_Theta, RPT1_Rel_Beta, RPT2_Abs_Theta, RPT4_Rel_Beta

B.2.3 Clustering and Model Assignment

Pairwise feature overlap is computed between different seizures, and a threshold of 80% is applied to define distinct clusters, as shown in Table B.6 and Table B.7, respectively.

Table B.6: Pairwise feature set overlap between seizures

	Sz2	Sz3	Sz4	Sz5	Sz6
Sz2	1.00	0.71	0.57	0.00	0.00
Sz3	0.71	1.00	0.83	0.00	0.00
Sz4	0.57	0.83	1.00	0.00	0.00
Sz5	0.00	0.00	0.00	1.00	0.57
Sz6	0.00	0.00	0.00	0.57	1.00

B.2.4 Final Feature Sets with and without Seizure Clustering

Following the clustering of seizures described earlier for Subject SID 22, the final feature sets corresponding to each seizure model are determined. In the clustered case, representative features are first identified for each seizure using the MUSE algorithm, then merged across seizures within the same cluster. This subsection summarizes the resulting model-specific

Table B.7: Clustering of seizures from training data based on feature overlap

Cluster	**Seizures**	**Description**
A	Sz3, Sz4	SPS cluster with consistent LAT/LMT feature signatures
B	Sz2	CPS with distinct LAT + LMT pattern
C	Sz5	CPS-GTC dominated by gamma and high-frequency power
D	Sz6	CPS-GTC with distinct theta and alpha feature pattern

feature sets and compares them to those obtained under a non-clustered approach, in which MUSE is applied to the pooled training data from all seizures.

Clustered MUSE Feature Selection: Seizures are grouped into four models based on the overlap of their MUSE-selected feature sets. For each model, a combined set of features is formed by taking the union of representative features from its constituent seizures. A total of 28 features are used across all models, of which 19 are unique, as depicted in Table B.8. Between 7 and 8 features are retained per model.

Table B.8: Feature sets assigned to seizure-specific models under the clustered approach

Model	**Seizures**	**# Feats.**	**Feature Set (Abbreviated)**
A	Sz3, Sz4	7	LAT2_Rel_Alpha, LAT3_Ratio_(Beta/Theta), LAT4_Rel_Theta, LMT1_Abs_Delta, LMT2_Ratio_(Alpha/Gamma1), LMT3_Abs_Alpha, LMT3_Rel_Gamma2
B	Sz2	7	LAT1_Abs_Theta, LAT2_Rel_Alpha, LAT3_Ratio_(Beta/Theta), LMT1_Abs_Delta, LMT3_Abs_Alpha, LAT4_Rel_Theta, LAT5_Rel_Delta
C	Sz5	7	RPT1_Abs_Gamma2, RPT2_Rel_Gamma4, RPT3_Ratio_(Gamma2/Theta), RPT4_Abs_Alpha, RPT2_Rel_Gamma3, RPT1_Rel_Beta, RPT3_Abs_Theta
D	Sz6	7	RPT1_Abs_Gamma3, RPT2_Rel_Gamma4, RPT3_Ratio_(Gamma2/Theta), RPT3_Abs_Theta, RPT1_Rel_Beta, RPT2_Abs_Theta, RPT4_Rel_Beta

Non-Clustered MUSE Feature Selection: To establish a baseline, MUSE is also applied to the entire set of preictal segments pooled across all training seizures, without clustering. A total of 38 features are selected using the same strategy. This higher number of features reflects the increased complexity required to model seizure variability in a unified framework. The selected features are:

LAT1_Abs_Theta, LAT1_Rel_Delta, LAT2_Rel_Alpha, LAT2_Rel_Beta, LAT2_Rel_Gamma4, LAT3_Ratio_(Beta/Theta), LAT4_Rel_Theta, LAT5_Rel_Gamma1, LAT6_Ratio_(Theta/Delta), LAT3_Rel_Alpha, LMT1_Rel_Gamma1, LMT2_Ratio_(Alpha/Gamma3), LMT2_Rel_Gamma3, LMT3_Abs_Alpha, LMT4_Rel_Beta, LMT3_Rel_Delta, LMT4_Ratio_(Gamma3/Beta), RPT1_Abs_Gamma1, RPT2_Rel_Gamma4, RPT3_Ratio_(Gamma3/Theta), RPT4_Abs_Alpha, RPT1_Rel_Beta, RPT2_Rel_Delta, RPT3_Abs_Theta, RPT2_AbsTheta, RPT4_Rel_Beta, RPT1_Rel_Alpha, LPT1_Rel_Delta, LPT3_Abs_Theta, RAP1_Abs_Theta, RAP2_Rel_Gamma2, RAP1_Ratio_(Alpha/Theta), RMT1_Abs_Beta, RMT2_Ratio_(Alpha/Beta), RPP1_Rel_Gamma3, RMT3_Abs_Theta, and LAT2_Ratio_(Beta/Delta)

Table B.9: Comparison of feature dimensions and model design between clustered and non-clustered approaches

Strategy	**# Models**	**# Total Features**	**# Unique Features**
Clustered	4	28	19
Non-Clustered	1	38	38

Table B.9 provides a summary comparison of both approaches, highlighting the number of models, total and unique features, and representative examples. This comparison demonstrates that simpler and seizure-specific models are obtained using the clustering-based approach, with fewer features required to distinguish between preictal and interictal states. In contrast, the non-clustered model required a larger and more diverse feature set to generalize across seizure types.