LUISS Guido Carli University

MSc in Data Science and Management

Course: Data Science in Action

ANTICIPATING EPILEPTIC SEIZURES THROUGH EARLY WARNING SIGNALS: DATA-DRIVEN METHODS FOR DETECTING PRE-ICTAL CHANGES AND TIPPING POINTS IN EEG

Supervisor:

Prof. Alessio Martino

Co-supervisor:

Prof. Paolo Spagnoletti

Candidate:

Aurora Menegatto

Student ID: 783301

Academic Year 2024–2025

Acknowledgements

Contents

Chapter 1

Introduction

The early detection of *tipping points*, critical transitions that can trigger abrupt and often irreversible changes in complex systems, has become an increasingly important topic in both theoretical and applied research. In biomedical and neurological contexts, the ability to anticipate such transitions is particularly crucial, as it can improve patient outcomes, reduce intervention costs, and enhance the overall resilience of healthcare systems.

As a system approaches a tipping point, characteristic patterns known as early-warning signals (EWS) may emerge in observational data. These signals often manifest as subtle but measurable changes in statistical and dynamical properties, including increased autocorrelation, altered variability, shifts in higher-order moments such as skewness and kurtosis, and modifications in temporal structure. Although not deterministic, EWS are commonly interpreted as indicators of declining system resilience and increased susceptibility to abrupt transitions. Their timely detection may therefore provide a window of opportunity for early intervention, potentially preventing or mitigating systemic failure.

Despite their conceptual appeal, the identification of EWS in real-world biomedical data remains a major challenge. Physiological systems such as the brain or gene regulatory networks consist of many interacting components and exhibit nonlinear, non-stationary, and multiscale dynamics that cannot be adequately described by closed-form equations. These systems often display state-dependent fluctuations and emergent behaviors that cannot be inferred from the analysis of individual com-

ponents alone (Trulla et al. 1996). As a result, traditional statistical approaches based on linearity or stationarity assumptions frequently fail to capture the complex dynamics preceding critical transitions.

This challenge is particularly evident in the context of epileptic seizures. A seizure, rather than representing the tipping point itself, can be interpreted as the observable outcome of one or more critical transitions in brain dynamics. During these transitions, neural activity progressively shifts from a relatively stable regime to a highly synchronized pathological state.

In this thesis, the problem of early seizure prediction is addressed by integrating theoretical concepts from complexity science with data-driven signal processing techniques. The focus is on developing interpretable methodologies capable of identifying systemic failure EWS in physiological time series, with a specific emphasis on scalp electroencephalography (EEG) recordings. The experimental analysis is conducted on neurological datasets from Siena Hospital and Boston Children's Hospital, obtained from the PhysioNet open-access biomedical database.

The underlying hypothesis of this work is that specific dynamical signatures– such as loss of variability, critical slowing down, rapid upward trends, and structural changes in the signal–may act as detectable precursors to seizure onset. Because these signatures are nonlinear, multiscale, and often weak, they frequently escape standard analysis techniques. To address this limitation, this thesis systematically investigates multiple feature extraction pipelines designed to capture complementary aspects of EEG dynamics.

The first pipeline is based on Recurrence Quantification Analysis (RQA), a nonlinear time-series analysis framework that characterizes recurrence patterns in phase space and provides insight into the determinism, complexity, and stability of the underlying dynamical system. The second pipeline relies on the Discrete Wavelet Transform (DWT), which decomposes EEG signals into frequency components across multiple time scales and enables the extraction of time – frequency features sensitive to transient changes in neural activity. Beyond the independent evaluation of these approaches, a third strategy based on feature fusion is explored, integrating RQA and DWT descriptors into a unified representation in order to

exploit their complementary information content.

This research is guided by two main objectives. The first is to empirically assess whether precursors of tipping points can be consistently observed in biomedical time series using advanced signal processing methodologies. The second is to develop an efficient and modular framework for crisis anticipation that balances predictive accuracy, interpretability, and computational cost. These objectives are pursued through both supervised and unsupervised analyses: in the former, extracted features are used to train a gradient boosting classifier to distinguish between inter-ictal and pre-ictal states within a predefined prediction horizon; in the latter, the temporal evolution of selected features is examined to identify persistent precursory trends preceding seizure onset.

The remainder of this thesis is structured as follows. Chapter 2 introduces the scientific and practical motivations underlying this work. It situates the problem within the broader context of biomedical crisis management, highlighting how the failure to detect early-warning signals (EWS) can lead to delayed intervention and severe clinical consequences. The chapter then formulates the research questions and defines the scope of the study, including assumptions related to data quality, temporal resolution, and labeling task.

Chapter 3 reviews the current state of the art in crisis prediction and tipping point estimation, with particular emphasis on biomedical signal processing applications. It discusses the limitations of existing approaches in noisy and nonlinear settings such as EEG analysis and introduces in detail the two datasets used throughout the thesis, describing their origin, structure, and relevance to the research problem.

Chapter 4 introduces the first proposed methodology, based on Recurrence Quantification Analysis, and details how recurrence-based features can be used to characterize hidden dynamical changes associated with proximity to critical transitions. Chapter 5 presents an alternative pipeline based on the Discrete Wavelet Transform for multi-resolution feature extraction. In addition to the standalone DWT-based approach, this chapter also describes the merged methodology, where recurrence-based and wavelet-based features are integrated within nearest- and context-based configurations to investigate whether complementary dynamical information im-

proves predictive performance.

Chapter 6 reports the experimental results and provides a unified evaluation and discussion of all proposed methodologies. Different configurations are assessed across both datasets, with performance analyzed in terms of early-warning detection capability, robustness to noise, false positive rates, generalizability across subjects, and the trade-offs between accuracy, interpretability, and computational cost. The implications of these findings are also examined with respect to their potential extension to real-time monitoring scenarios.

Finally, Chapter 7 summarizes the main contributions of the thesis and outlines possible directions for future research, including the integration of domain knowledge, generalization to multimodal data sources, and deployment within operational clinical monitoring systems. Overall, this work aims to provide a practical and scientifically grounded framework for early-warning detection in complex physiological systems, supported by both theoretical insights and validation on real-world data.

Chapter 2

Context Description and Problem Statement

Throughout the history of medicine, there have been countless cases where recognizing warning signs even moments earlier could have altered the final outcome. In complex physiological systems, these moments often take the form of *tipping points*, critical transitions in which small changes can abruptly push the system into a qualitatively different state. Anticipating such events before they occur is both a scientific challenge and a clinical necessity.

This chapter introduces the central problem that motivated this work: the early detection of tipping points in biomedical, and more specifically neurological, time series. Within clinical neuroscience, there is an emerging interest in EWS as potential biomarkers of impending pathological state transitions. Such signals could enable timely interventions, reduce the severity of acute episodes, and ultimately improve patient outcomes.

The choice to focus on tipping points is motivated by their dual relevance. From a scientific perspective, they represent a fundamental problem in the study of complex systems (Ambika and Kurths 2021), where instability may emerge from subtle and often hard-to-detect changes in system dynamics. From a clinical perspective, the ability to anticipate such transitions offers tangible advantages, enabling preventive rather than reactive care, optimizing the use of medical resources, and improving

patients' quality of life. Among biomedical domains, neurology, and epilepsy in particular, offers a clear and clinically meaningful setting in which these concepts can be tested. Despite this potential, research on epileptic seizure dynamics from a tipping point perspective remains limited. This is mainly due to the intrinsic complexity, non-stationarity, and strong inter-subject variability of EEG signals, which make the detection of early warning patterns particularly challenging.

Epileptic seizures are a condition of considerable clinical importance. The term epilepsy describes a a group of neurological disorders due to abnormal, excessive, or synchronous neuronal activity in the brain. It is characterized by a persistent predisposition that leads to recurrent seizures. Such episodes include impaired cognitive function, altered awareness, and involuntary motor activity, which often have a major impact on safety and everyday functioning.

The brain activity of individuals with epilepsy can typically be described in terms of three states derived from EEG recordings: *inter-ictal state*, *pre-ictal state*, and *ictal state*. From a dynamical systems perspective, seizures can be interpreted as sudden, nonlinear shifts in brain state—transitioning from a relatively stable interictal phase to the ictal phase—often preceded by complex and subtle pre-ictal dynamics. Detecting these early changes in real time and being able to rapidly administer antiepileptic drug treatment to patients could transform patient care.

In recent years, advances in sensing technologies, continuous monitoring systems, and open source repositories such as PhysioNet have made large-scale physiological datasets increasingly available. Within neurology, EEG remains a cornerstone for both diagnosis and long-term monitoring. Its high temporal resolution provides a unique window into brain activity, capturing transient electrical events linked to seizures. However, EEG analysis for predictive modeling faces several significant challenges:

- **Noise and artefacts**: Recordings are frequently contaminated by muscular activity, eye movements, electrode shifts, and environmental interference.
- **High dimensionality**: Multi-channel setups generate large and complex feature spaces, which can increase the risk of overfitting and demand substantial

computational resources.

- **Non-stationarity**: EEG signals vary over time due to changes in brain state, patient condition, and recording context, making it difficult to extract stable predictive patterns.
- **Rarity of events**: In long-term monitoring, seizure episodes typically occupy only a small fraction of the total recording time, creating a highly imbalanced prediction problem.

Addressing these challenges is essential for building robust predictive models capable of anticipating seizures and other neurological tipping points.

2.1 Problem Statement

Although the theory of EWS provides a promising conceptual framework, its direct application to real-world EEG data for seizure prediction faces several limitations:

1. Low reliability of classical EWS in noisy and stochastic systems, where high noise levels can distort the expected trends of statistical indicators, leading to spurious signals, false positives, or missed detections (Proverbio, Skupin, and Gonçalves 2023).
2. Limited interpretability in many high-performing machine-learning models, which hinders clinical adoption (Wu et al. 2025).
3. Insufficient integration of multi-scale and non-linear analysis methods, which are essential to capture the hierarchical and scale-dependent dynamics of complex systems, often results in models that fail to reproduce the structural changes occurring at different temporal or spatial resolutions (Pyrialakos, Kalogeris, and Papadopoulos 2023).

Most of the current limitations indicate the demand for a methodological framework that is robust yet interpretable for the derivation of meaningful indicators from such complex signals, keeping in mind their clinical relevance.

2.2 Datasets Overview

The experimental work presented in this thesis is based on two complementary datasets obtained from the PhysioNet repository. The first consists of scalp EEG recordings collected under routine clinical conditions at the University Hospital of Siena, Italy. It offers a realistic representation of the variability and noise encountered in non-invasive monitoring, making it well-suited for the development of robust, real-world detection algorithms (Detti, Vatti, and Zabalo Manrique de Lara 2021).

The second dataset is derived from the CHB-MIT Scalp EEG Database, developed in collaboration with the Boston Children's Hospital. It contains long-term EEG recordings acquired from pediatric patients with intractable epilepsy undergoing presurgical evaluation (Shoeb 2009).

These datasets were selected due to their clinical relevance, differences in recording protocols, and their suitability for evaluating both the generalizability and robustness of the proposed methodologies. A detailed description of each dataset, including acquisition protocols, preprocessing steps, and labeling strategies, is provided in Chapter 3.

2.3 Research Objective

Building on the challenges outlined above, the main goal of this thesis is to develop and validate interpretable, data-driven methodologies for the early detection of critical transitions in EEG time series, with the ultimate aim of enabling seizure prevention. Although the models operate at the level of prediction, their clinical purpose is preventive, providing actionable anticipation windows prior to seizure onset. The proposed framework integrates:

- Non-linear dynamical measures to characterize changes in system stability.
- Time–frequency decomposition via *Discrete Wavelet Transform* to capture transient features across multiple scales.
- Automated hyperparameter optimization using *Optuna*, applied to tree-based

model hyperparameters and decision thresholds within a strict LOPO validation framework.

- Comparative evaluation of gradient boosting architectures (LightGBM and CatBoost) to assess robustness across model families.

The framework is tested on the two aforementioned datasets to assess its robustness, interpretability, and predictive performance.

2.4 Concluding Remarks

The challenges, motivations, and datasets introduced in this chapter provide the conceptual and empirical foundation for the literature review presented in Chapter 3. The next chapter examines the theoretical basis of EWS, reviews key computational approaches for EEG-based seizure prediction, and offers a detailed description of the datasets used in this work, thereby connecting the problem formulation to the subsequent methodological design.

Chapter 3

Literature Review and Dataset Description

3.1 Introduction

This chapter outlines the scientific and methodological landscape relevant to the detection and anticipation of critical transitions (*tipping points*) in biomedical time series. It first introduces the theoretical background of EWS and tipping points in complex systems, highlighting their clinical relevance in neurological contexts. It then reviews representative computational approaches proposed in the literature, ranging from classical statistical indicators to methods rooted in nonlinear dynamics and network theory.

Particular attention is devoted to the methodological frameworks that directly inform this thesis—*Recurrence Quantification Analysis (RQA)* and the *Discrete Wavelet Transform (DWT)*—which constitute the core of the proposed feature extraction pipelines. Emerging network-based approaches, such as *Dynamic Network Biomarkers (DNB)*, *Sample-Specific Networks (SSN)*, and *Network Information Entropy of Edges (NIEE)*, are briefly discussed in the final part of the thesis as future possible integration of the present work within the broader research landscape, rather than as methodologies implemented in this study.

As introduced in Chapter 2, this research aims to explore whether it is possible to identify any EWS that specifically precede epileptic seizures by applying a com-

bination of data-driven and interpretable methodologies. Rather than assuming the existence of universal precursors, the thesis investigates whether specific dynamical patterns—extracted through nonlinear and multiscale analyses—can serve as warning indicators in real EEG recordings and thus impede critical transitions in these physiological systems.

The extracted features are subsequently evaluated within a machine learning framework to assess their ability to distinguish pre-ictal from inter-ictal states in a clinically meaningful setting.

3.2 Tipping Points and Early Warning Signals in Biomedicine

Complex systems, whether ecological, physical, or biological, can experience abrupt and often irreversible transitions when they approach a critical threshold, known as *tipping point*. As outlined in previous chapters, such transitions in biomedical settings may occur as acute events (e.g., epileptic seizures or cardiovascular collapse) or as gradual shifts toward pathological conditions. Anticipating such transitions remains a fundamental challenge, as early recognition can open the door to timely preventive or therapeutic interventions.

This anticipatory capacity is grounded in the identification of *early warning signals (EWS)*: statistical or dynamical patterns that reflect a decline in the system's resilience as it nears a tipping point (Scheffer et al. 2009). Canonical examples of such signals include increased variance, rising lag-1 autocorrelation, and alterations in higher-order statistical moments or recurrence structures. These indicators have been documented across a wide range of systems, from ecological environments to the human brain.

3.2.1 Theoretical and Computational Methods

Classical EWS detection strategies rely on the principle of *critical slowing down* (CSD), whereby proximity to a bifurcation reduces a systems capacity to rapidly recover from perturbations. This slowdown manifests as measurable increases in variance, autocorrelation, and skewness in the system's time series (Dakos et al.

2012).

In neuroscience, CSD has been applied to characterize the dynamics leading up to epileptic seizures, viewed as sudden transitions in brain state indicative of critical thresholds. Meisel, Storch, et al. 2012 interpreted seizures as the collapse of adaptive self-organized criticality, while Maturana et al. 2020 provided empirical validation of CSD-based indicators as potential biomarkers of seizure susceptibility.

Building on this foundation, the *critical brain hypothesis* (CBH) suggests that optimal information processing occurs near a phase transition between ordered and disordered neural activity (Beggs and Plenz 2003; Cocchi 2017). Departures from this quasi-critical regime have been linked to pathological brain dynamics. Within this framework, S. Liu, F. Li, and Wan 2023 proposed the *Distance to Tipping Point* (DTP), a network-level indicator derived from random matrix theory and dynamic network biomarkers. Their results show that DTP undergoes abrupt changes prior to seizure onset, offering a more nuanced interpretation of critical transitions in EEG.

These theoretical advances provide the foundation for a wide range of practical implementations, particularly in the analysis and classification of EEG data for seizure detection.

3.2.2 Applied Models in Literature

Numerous studies have translated theoretical insights on tipping points and early warning signals into practical EEG-based seizure detection frameworks. For example, Sharma, Pachori, and Rajendra Acharya 2017 used an analytic time-frequency flexible wavelet transform (ATFFWT) to extract relevant features from EEG signals. These features were ranked using a t-test and classified with a least squares support vector machine (LS-SVM), achieving over 99% accuracy in distinguishing between seizure and non-seizure states. Similarly, Siuly et al. (Siuly, Y. Li, and Y. Zhang 2016) applied LS-SVM with a clustering-based approach and reported an average classification accuracy of 94.18%.

Other works have explored hybrid optimization strategies. Subasi 2013 combined discrete wavelet transform (DWT) with genetic algorithms and particle swarm opti-

mization (PSO) to optimize SVM classifiers, reaching accuracies above 97%. These models often benefit from additional dimensionality reduction techniquessuch as principal component analysis (PCA), independent component analysis (ICA), or linear discriminant analysis (LDA)—which help improve generalization by filtering out redundant or irrelevant features.

Beyond SVM, other classifiers have also shown promise. Mursalin et al. (Mursalin et al. 2017) developed a hybrid framework that couples an improved correlation-based feature selection algorithm with random forest (RF), outperforming several existing benchmarks. Time-frequency-based approaches have also been successfully combined with artificial neural networks (ANN). Tzallas et al. (Tzallas, Tsipouras, and Fotiadis 2009) reported seizure detection accuracies exceeding 97% using ANN models fed with features derived from EEG time-frequency decomposition.

Despite their strong performance, many of these systems face challenges in terms of overfitting and limited generalizabilityespecially when tested across patients or under varying recording conditions. These issues underscore the need for more modular, interpretable, and robust pipelines that can adapt to different subjects and data quality while maintaining high predictive performance.

However, recent work suggests that a purely signal-based approach may be insufficient, prompting the exploration of methods that account for multiscale and network-level brain dynamics.

3.2.3 A Shift Toward Multiscale and Network-Aware Analysis

To improve the interpretability and sensitivity of EEG-based seizure prediction, recent studies have started to combine nonlinear dynamical methods with multiscale signal analysis. Tools like Recurrence Quantification Analysis (RQA) and Discrete Wavelet Transform (DWT) offer complementary perspectives: RQA captures temporal patterns and transitions in phase space, while DWT isolates features across multiple frequency bands. For instance, Sharmila and Geethanjali 2017 applied DWT to extract sub-band features and used statistical tests to identify those most relevant to seizure activity. However, their approach lacked a thorough strategy for addressing feature redundancy and generalization to new data.

These developments point toward a more integrated strategy for seizure forecastingone that bridges local, signal-level features with large-scale network dynamics. The pipeline proposed in this thesis follows that direction by combining RQA and DWT in a modular framework. Although advanced network-based methods are not implemented here, they represent a clear path for future extensions aimed at improving both predictive accuracy and physiological interpretability.

3.3 Datasets and Data Retrieval

As anticipated in the previous chapter, the effectiveness of the experimental method proposed in the next chapters has been assessed using two complementary EEG databases. These were selected to span both controlled research settings and heterogeneous clinical scenarios. Both datasets are publicly available on *PhysioNet*, a well-established open-access platform for physiological data sharing (Goldberger et al. 2000).

3.3.1 Dataset 1: Siena Scalp EEG (PhysioNet)

The first dataset comprises scalp EEG recordings from 14 patients with clinically diagnosed epilepsy, collected at the Neurology and Neurophysiology Unit of the University of Siena (*Siena Scalp EEG Database* 2025). The cohort includes 9 male subjects (aged 25–71 years) and 5 female subjects (aged 20–58 years). Data were acquired during video-EEG monitoring at a sampling rate of 512 Hz, using electrodes placed according to the international 10–20 system. In most recording sessions, one or two electrocardiogram (ECG) channels were also included.

EEG signals were recorded using EB Neuro and Natus Quantum LTM amplifiers with reusable silver/gold cup electrodes. During the recordings, subjects were instructed to remain in bed and could be either awake or asleep.

All procedures were approved by the Ethics Committee of the University of Siena and conducted in accordance with the Declaration of Helsinki. EEG signals are stored in European Data Format (EDF). The dataset is organized into 14 subject-specific directories, each containing between 1 and 5 EDF files along with

a text file reporting seizure-related metadata. Recordings within each directory are ordered chronologically, and all dates have been anonymized. A summary CSV file is provided and reports, for each subject, the following information:

- sex;
- age;
- seizure classification;
- number of EEG channels;
- number of recorded seizures;
- total recording duration (in minutes).

Overall, the dataset includes 47 seizures recorded over approximately 128 hours of EEG data.

Each subject directory also contains a `Seizures-list-PNxx.txt` file, which specifies the sampling rate, the list of valid EEG and EKG channels, recording start and end times, and seizure onset and offset times (formatted as hh.mm.ss). These annotations enable precise identification of interictal and pre-ictal intervals and are used to define supervised labels for model training and evaluation, as well as to investigate seizure-related tipping points.

3.3.2 Dataset 2: CHB-MIT Scalp EEG Database (PhysioNet)

The second dataset includes long-term scalp EEG recordings from pediatric patients with intractable epilepsy, collected at Boston Children's Hospital. It comprises 23 cases from 22 subjects (5 males, ages 3–22; 17 females, ages 1.5–19). Case `chb21` was recorded 1.5 years after `chb01` from the same subject (*CHB-MIT Scalp EEG Database* 2025). Most patients were monitored for several days, typically after withdrawal of anti-seizure medication, to assess seizure patterns and evaluate surgical options.

Each case (e.g., `chb01`, `chb02`, etc.) includes 9-42 consecutive EDF files. Short gaps (usually $\leq 10\,\mathrm{s}$, sometimes longer) may occur between sequential files due to

hardware limitations. Protected health information has been replaced with surrogate data; while absolute dates are anonymized, relative timing within each case is preserved.

Recordings are sampled at 256 Hz with 16-bit resolution. Most files contain 23 EEG channels arranged according to the international 10–20 system, although some recordings include 24 or 26 channels. Additional signals may also be present, such as electrocardiogram (ECG) channels (e.g., in the last 36 files of `chb04`) or a vagal nerve stimulator channel (e.g., in the last 18 files of `chb09`). Dummy channels, marked as "`-`" and used solely for display purposes, are excluded from the analysis.

Overall, the dataset comprises 664 EDF files corresponding to approximately 916 hours of EEG recordings, with a total of 198 annotated seizures (182 of which belong to the original 23 subjects). Seizure onset and offset times are provided in per-file `.seizure` annotation files, which are listed in the `RECORDS-WITH-SEIZURES` index. Additional metadata, including electrode montages and seizure timing expressed in seconds from the beginning of each recording, are reported in the `chbnn-summary.txt` files. Demographic information, such as age and sex, is available in the `SUBJECT-INFO` file.

For the purposes of this thesis, only recording sessions containing a single seizure event are considered, in order to simplify evaluation and comparison across experiments. Sampling rates vary across subjects, and the dataset provides extensive metadata-including electrode configurations, anonymized demographic information, and clinical notesmaking it a realistic and challenging benchmark for the investigation of early-warning signals preceding epileptic seizures.

Both datasets comply with PhysioNet's data-use policies, ensuring full de-identification and ethical standards. Their combined use enables algorithm evaluation under both controlled and real-world conditions, supporting robust and generalizable findings. The features and analytical methods described in the following section were applied to these datasets to extract EWS aimed at detecting pre-ictal tipping points and characterizing seizure crisis transitions.

3.4 Algorithms Implemented in This Thesis

3.4.1 Recurrence Quantification Analysis (RQA)

Recurrence Quantification Analysis (RQA) characterizes the recurrence of dynamical states in a reconstructed phase space derived from a univariate time series. By embedding the signal in a higher-dimensional space, RQA captures the geometric structure of the underlying dynamics and provides quantitative measures of its temporal organization. In this thesis, RQA is adopted as the primary analytical framework due to its sensitivity to nonlinear regime transitions and its ability to reveal subtle changes in EEG dynamics preceding seizure onset.

The phase space is reconstructed using time-delay embedding, defined by an embedding dimension and time delay, while recurrences are identified by thresholding pairwise distances between state vectors. The resulting recurrence plot encodes the temporal evolution of the system and forms the basis for computing a set of descriptive metrics. The central measure considered in this work is the *Recurrence Rate (RR)*, which quantifies the density of recurrent states and reflects the overall stability of the system. To enhance sensitivity to gradual transitions, normalized and derivative forms of RR are also analyzed.

Additional RQA measures provide complementary information on the structure of the dynamics. *Determinism (DET)* captures the prevalence of diagonal line structures and is associated with predictability and deterministic behavior.*trapping time (TT)* characterizes vertical line structures and is indicative of intermittent phases in the signal (Webber Jr and Zbilut 2015). Variations in these measures have been previously linked to state transitions in complex physiological systems.

To obtain time-resolved features, RQA is applied within a sliding-window framework, allowing continuous tracking of dynamical changes over the course of long EEG recordings. This approach enables direct comparison of RQA features across interictal and pre-ictal periods and supports the identification of early-warning patterns associated with seizure generation.

3.4.2 Discrete Wavelet Transform (DWT)

The Discrete Wavelet Transform (DWT) decomposes a signal into multiple resolution levels by iteratively applying pairs of low-pass and high-pass filters, allowing analysis across multiple time-frequency scales. In EEG analysis, this multi-resolution representation allows both spectral and statistical information to be extracted from inherently non-stationary signals, making DWT particularly well suited for capturing transient phenomena such as pre-ictal dynamics (Sharmila and Geethanjali 2017).

When applied to EEG recordings, DWT enables the analysis of activity within specific frequency sub-bands commonly associated with physiologically meaningful rhythms. From these sub-bands, a range of features can be derived, including energy-based, amplitude-related, and variability measures, which have been reported to be sensitive to changes in brain dynamics preceding seizure onset.

DWT-based feature extraction is typically combined with sliding-window segmentation, whereby EEG signals are divided into short, partially overlapping temporal segments. This strategy supports modeling of temporal evolution while maintaining statistical reliability within each window. Such a representation is particularly suitable for early-warning scenarios, in which transitions from interictal to pre-ictal states are expected to unfold progressively rather than occur abruptly.

In contrast to nonlinear dynamical approaches, such as RQA, which emphasize temporal structure and state-space recurrence properties, DWT offers a complementary perspective by decomposing the signal into localized time-frequency components. Through successive filtering and downsampling operations, DWT separates the EEG signal into frequency-specific sub-bands while preserving temporal localization. This multiscale representation enables the characterization of oscillatory activity at different resolutions, capturing changes in spectral energy distribution and transient dynamics that may precede critical transitions. For this reason, DWT-based methods are often employed as a benchmark or alternative approach in seizure prediction studies, enabling comparative evaluation of frequency-based and dynamical descriptors under consistent experimental conditions.

3.5 Network-Based Critical Transitions Across Neurological Disorders

More broadly, the concept of *multisystem failure as a tipping point* has gained traction in the study of neurodegenerative diseases. Notably, recent analyses using extended Cox regression on UK Biobank data have shown that modifiable risk factors-such as hypertension and obesity-when present before the age of 62, can induce tipping dynamics in Alzheimer's disease progression that outweigh even genetic risk associated with the APOE genotype (Merlini et al. 2025). These findings support the view that critical transitions are not merely localized phenomena in biomedicine but reflect systemic shifts embedded in networked pathophysiology.

3.6 Summary and Research Gaps

The literature demonstrates that EWS of critical transitions can be investigated through statistical indicators, nonlinear dynamical measures, and network-based formulations. In the context of EEG-based seizure prediction, numerous approaches have been proposed, yet no consensus has emerged regarding the most reliable dynamical markers or the most interpretable modeling strategy.

Two main gaps can be identified. First, many studies focus either on spectral features or on nonlinear dynamical descriptors, without systematically comparing their complementary contributions. Second, the integration of interpretable feature extraction methods within modular and computationally efficient prediction pipelines remains an open challenge, particularly when aiming to balance predictive accuracy with physiological interpretability.

This thesis addresses these gaps by systematically comparing RQA-based and DWT-based representations of EEG dynamics, as well as their merged configurations, within a unified experimental framework. The goal is not to assume universal precursors, but to empirically assess whether specific dynamical patterns can consistently anticipate seizure onset across real-world datasets.

The next chapter introduces the methodological framework developed in this thesis,

detailing the feature construction process, temporal labeling strategy, and classification setup used to evaluate early-warning detection performance.

Chapter 4

First Methodology: Dynamical Network Modelling via Recurrence Quantification Analysis

This chapter outlines the EEG analysis approach in epilepsy, aimed at discovering tipping points—the unique transitions of dynamic behavior in the brain, which are prior to seizure initiation.

Overview of the Pipeline. Starting from raw EDF recordings, the workflow performs the following steps:

1. Signal conditioning and channel harmonization.
2. Sliding-window segmentation and phase labeling (interictal / pre-ictal / ictal).
3. Construction of dynamic functional networks via window-wise connectivity matrices.
4. Recurrence plot generation over the resulting connectivity-state trajectory.
5. Extraction of RQA descriptors (RR, DET, TT) and derived instability features (smoothing, variability measures, and temporal derivatives).

6. Downstream evaluation through supervised modeling under strict patient-level separation, complemented by an *a priori* derivative-based tipping point criterion.

The pipeline is designed as a reproducible end-to-end workflow that takes long-term multi-channel EEG recordings as input and produces both supervised pre-ictal window classifications and unsupervised dynamical transition indicators (tipping points).

It was first verified in a controlled single-patient setting to validate feature dynamics and temporal alignment, and was subsequently scaled to a multi-patient setting to assess cross-subject consistency under strict patient-level separation.

4.1 Signal Preprocessing

4.1.1 Custom Preprocessing Scripts

The EEG data, provided in European Data Format (EDF), were processed using a standardized preprocessing pipeline to ensure signal quality, inter-session consistency, and suitability for functional connectivity analysis. The pipeline was adapted to account for structural differences between the two datasets while preserving a common methodological framework.

4.1.2 Channel Filtering and Exclusion

The pre-processing pipeline began with the removal of non-cerebral channels to ensure that subsequent analyzes focused exclusively on EEG activity. Physiological signals unrelated to cortical dynamics, such as electrocardiogram (ECG), oxygen saturation, electromyogram (EMG), auxiliary markers, any channels with numeric or non-standard labels and non-standard channels, were removed.

Given the higher heterogeneity in channel naming and annotation, the ScalpEEG dataset, required a more comprehensive filtering procedure and exclusion strategy. In contrast, the CHB-MIT dataset presented a more standardized channel structure and cleaner signal acquisition setup, allowing for a simplified approach in which only

channels explicitly labeled as "EKG" were removed without affecting downstream analyses.

In addition, to ensure alignment across patients and recording sessions, channel labels were subsequently normalized by removing specific prefixes and suffixes, harmonizing case and spacing, and preserving the structure of bipolar derivations. This normalization step enabled reliable matching of homologous EEG channels across longitudinal recordings and ensured consistency in subsequent connectivity analysis.

Following label normalization, only bipolar EEG channels consistently present across all recordings for each patient were retained, ensuring longitudinal comparability within subjects. Depending on the experimental configuration, analyses were performed using either the full available channel set or a reduced common subset.Analyses were restricted to channels associated with canonical cortical regions (frontal, central, parietal, occipital, and temporal) in order to reduce the hypothetical variability introduced by peripheral or non-standard electrodes,

To assess the quality of the signal, channels with zero variance, indicative of flat or corrupted signals were identified. In ScalpEEG dataset, such channels were removed before the segmentation task, whereas in the CHB-MIT Dataset, channel removal was performed on a window-by-window basis to ensure data reliability and to account for any temporary issues that might be found in the signals.

4.1.3 Band-Pass filtering and seizure annotation parsing

All EEG signals were band-pass filtered between 0.5 Hz and 40 Hz, which is a frequency range commonly adopted in seizure analysis to isolate physiologically relevant oscillatory activity while attenuating slow drifts and high-frequency noise (Malekzadeh et al. 2021). In cases where recordings already included appropriate filtering metadata, this step was omitted to avoid redundant processing.

Seizure onset and offset annotations were extracted from dataset-specific metadata files and mapped onto a unified temporal reference aligned with the start of each EEG recording. While ScalpEEG dataset provided centralized and time-synchronized annotations, CHB-MIT dataset required additional temporal alignment due to fact that seizure times were reported relative to hospital registration or

external reference points. Therefore, dataset-specific offsets were applied to ensure accurate correspondence between EEG samples and seizure events.

4.1.4 Sliding Window Segmentation

To segment preprocessed EEG recordings, it was decided to divide them into overlapping sliding windows of 2 seconds duration with an 80% overlap, corresponding to a step size of 0.4 seconds at a sampling rate of 256 Hz, following the approach proposed in (Lopes et al. 2021). This configuration provides a compromise between temporal resolution and statistical stability. These sliding windows will be crucial to later build a functional connectivity matrix, as explained in later Section 4.1.6.

Window duration and stride are design-dependent and were selected to match the requirements of each representation: short, highly overlapping windows are used for connectivity-state tracking and recurrence dynamics, whereas longer windows are adopted for stable wavelet-domain estimates.

4.1.5 Seizure Labeling and Temporal Context

To contextualize each time window relative to the seizure event, labels were assigned as follows:

- **0 Interictal:** baseline activity far from seizure-related intervals;
- **1 Pre-ictal:** within 600 seconds preceding seizure onset;
- **2 Ictal:** from seizure onset to seizure offset.

This labeling scheme is fundamental for the subsequent analysis, as it allows the detection of early signs of instability during the pre-ictal period, interpreted as potential tipping points in neural dynamics.

To achieve reliable phase annotation, seizure onset and offset times were extracted from the summary text files available using a custom parser specifically developed for the task, which allows to ensure precise alignment between EEG segments and clinical events, providing a solid basis for phase-specific analysis.

4.1.6 Functional Network Construction

For each valid sliding window, the preprocessed EEG recordings were transformed into dynamic functional networks (dFNs). Let $X \in \mathbb{R}^{N \times T}$ denote the windowed EEG signal, where N is the number of bipolar channels and T the number of time samples. The resulting correlation matrix $C \in \mathbb{R}^{N \times N}$ captures the instantaneous functional organization of the brain during the corresponding time interval. This matrix represents the linear statistical dependence between all pairs of EEG channels within the 2-second time window and provides a time-local snapshot of cortical functional organization.

Windows containing channels with zero variance or yielding undefined connectivity estimates were excluded to prevent corrupted or non-informative data from influencing downstream analyses. This quality control step ensured that only reliable connectivity patterns were retained.

By computing connectivity matrices across successive overlapping windows, a temporally ordered sequence of functional networks was obtained, capturing the evolution of inter-regional interactions throughout the recording. Each network snapshot was associated with its corresponding temporal index, seizure-phase label (interictal, pre-ictal, or ictal), and anonymized subject identifier.

In CHB-MIT, seizure annotations were occasionally reported relative to external reference times rather than the start of the EEG recording. To ensure accurate temporal alignment, it was applied a dataset-specific temporal offsets, so that all seizure-phase labels were consistently referenced to the EEG time base, while in ScalpEEG, annotations were already synchronized and required no such adjustment.

This sequence of time-resolved functional connectivity matrices forms the structural input for subsequent RQA. Preserving temporal order, seizure context, and subject-level structure ensures that downstream analyses, including patient-wise validation procedures, remain both physiologically interpretable and reproducible across heterogeneous datasets.

4.2 Recurrence Quantification Analysis of Dynamic Networks

4.2.1 Recurrence Plot Generation and RQA Feature Extraction

Once the dynamic functional networks (dFNs) were obtained, each $N \times N$ functional connectivity matrix—representing the brain's connectivity state within a given time window—was further analyzed using recurrence plots (RPs). While correlation matrices provide static snapshots of functional organization, they do not explicitly characterize how connectivity states evolve, persist, or reoccur over time. Recurrence plots address this limitation by encoding the temporal recurrence of similar network states, thereby enabling the investigation of dynamical properties underlying brain activity.

In the RQA setting, each functional connectivity matrix was interpreted as a point in a high-dimensional phase space describing the evolution of brain network dynamics. Pairwise similarities between connectivity states across time were quantified using a matrix norm, and a fixed threshold was applied to determine whether two states were considered recurrent, following established approaches in the literature (Lopes et al. 2021).

Matrix Vectorization and Dimensional Handling

Prior to recurrence analysis, each functional connectivity matrix was transformed into a vector representation to enable direct comparison between successive time windows within a common feature space. This transformation preserves the full information content of the connectivity patterns while allowing the construction of recurrence plots across time.

To accommodate dataset-specific characteristics, two complementary vectorization strategies were adopted. For ScalpEEG, only the upper triangular elements of the symmetric correlation matrix (excluding the diagonal) were retained thereby removing redundant information and reducing dimensionality and computational burden, which is particularly advantageous when analyzing recordings with a larger

number of EEG channels. Given the relatively small and homogeneous nature of this dataset, fixed recurrence plot parameters were used across patients to ensure methodological consistency and facilitate direct comparison.

For what regards CHB-MIT, the full correlation matrix was vectorized, retaining all entries. This choice was feasible due to the lower channel dimensionality and was motivated by the greater heterogeneity of the dataset in terms of recording duration, noise characteristics, and channel configurations. In this case, recurrence plot parameters were allowed to vary within predefined ranges to maintain robustness across diverse recording conditions.

Despite these differences, both strategies preserve the same conceptual representation of brain network states and yield recurrence-based features that remain comparable across datasets in subsequent multi-patient analyses.

Recurrence Plot Computation and Primary RQA Features Extraction

Recurrence plots were generated by evaluating pairwise similarities between vectorized connectivity states across time. Each recurrence plot is a binary, square matrix in which an entry equal to one indicates that the similarity between two connectivity states exceeds a predefined threshold, while zero indicates non-recurrence. The resulting structure provides a compact representation of how often and in what temporal patterns the brain revisits similar functional configurations.

For ScalpEEG, recurrence plot parameters were held constant across patients and sessions, prioritizing longitudinal consistency and minimizing potential bias introduced by subject-specific tuning, while for CHB-MIT, a more flexible parameterization was employed to account for substantial inter-subject and inter-session variability. This balance between consistency and adaptability ensured that recurrence plots reliably captured meaningful dynamical patterns in both datasets.

From each recurrence plot, the following classical RQA measures were extracted to characterize temporal organization in brain network dynamics:

- **Recurrence Rate (RR):** proportion of time windows in which the functional connectivity pattern is similar to a previously observed pattern. Higher values

indicate more repetitive network dynamics, whereas lower values reflect greater variability over time

- **Determinism (DET):** proportion of recurrent connectivity patterns that evolve in a similar way across consecutive time windows. Higher values indicate more predictable and structured network dynamics.

- **Trapping Time (TT):** average duration for which the brain remains in the same or very similar functional connectivity configuration across consecutive time windows, reflecting persistent or laminar dynamics.

Together, these measures provide complementary insights into the stability, predictability, and temporal persistence of functional brain networks.

4.2.2 Feature Normalization and Derivation

To enhance comparability across patients and emphasize peri-ictal transitions, additional processing steps were applied to each primary recurrence feature (RR, DET, TT). These included:

Phase-wise normalization. To reduce inter-individual variability and account for baseline differences between physiological states, each recurrence feature (RR, DET, TT) was standardized separately within each seizure phase (interictal, preictal, ictal) using a z-score transformation.

For illustration, the transformation applied to the Recurrence Rate (RR) is given by:

$$\mathrm{RR\_phase\_norm}_{p,t} = \frac{\mathrm{RR}_{p,t} - \mu_{p,\phi(t)}}{\sigma_{p,\phi(t)} + \varepsilon}$$

where p denotes the patient index, t the time window, and $\phi(t)$ the phase label associated with time t. The terms $\mu_{p,\phi(t)}$ and $\sigma_{p,\phi(t)}$ represent, respectively, the mean and standard deviation of the feature computed for patient p within the corresponding phase. A small constant $\varepsilon = 10^{-6}$ was added to the denominator to prevent numerical instability in cases of near-zero variance.

This normalization strategy ensures that the feature values are interpreted relative to the patient- and phase-specific baseline, highlighting deviations occurring within each physiological state rather than absolute magnitude differences between individuals.

Temporal smoothing. To attenuate high-frequency noise and emphasize slower trends in the evolution of the features, a centered rolling average was applied. Specifically, each feature value was replaced by the mean of its surrounding 25 samples, corresponding to a temporal window of approximately 10 seconds (with a step size of ≈ 0.4 s). This smoothing procedure reduces short-lived fluctuations that are unlikely to carry physiological meaning, while preserving the broader temporal patterns relevant for seizure prediction.

Derived dynamics. Several secondary indicators were computed to capture short-term variations and local trends in the evolution of RQA features. These additional measures were designed to emphasize rapid fluctuations, directional changes, and local variability that may not be apparent from the raw features alone:

- **Instantaneous delta:**

$$\Delta \mathrm{RR}_{p,t} = \mathrm{RR}_{p,t} - \mathrm{RR}_{p,t-1}$$

  Quantifies the immediate change from one sample to the next, highlighting abrupt shifts in recurrence rate that may precede seizure onset.

- **Smoothed derivative:**

$$\mathrm{RR\_derivative}_{p,t} = \mathrm{RR\_smooth}_{p,t} - \mathrm{RR\_smooth}_{p,t-1}$$

  Computed on temporally smoothed features, this captures gradual directional trends while reducing sensitivity to high-frequency noise.

- **Local slope:** Estimated as the regression coefficient of a first-order polynomial fitted over a 5-sample rolling window. This provides a more robust estimate

of short-term trend direction compared to simple differences, by incorporating information from a local neighborhood.

- **Rolling variability:** Defined as the standard deviation of feature values within a sliding window of 10 samples. This measures local volatility, allowing detection of unstable periods where brain dynamics fluctuate more strongly than usual.

Composite indicators. In addition to the base recurrence measures (RR, DET, TT) and their derived temporal features (such as rolling means, standard deviations, and first-order derivatives) described above, composite metrics to capture potential non-linear interactions between different aspects of recurrence structure were defined. These indicators emphasize regimes where the relationship between recurrence rate and determinism reveals distinct dynamical signatures:

- **DET-to-RR ratio:**

$$\frac{\mathrm{DET}}{\mathrm{RR}+\varepsilon}, \quad \varepsilon = 10^{-6}$$

  This ratio quantifies the balance between predictable (deterministic) structure and overall recurrence density. A high DET/RR ratio indicates that, even when recurrent states are sparse, those that do occur tend to organize into predictable diagonal patterns. To prevent instability due to very low RR values, the ratio was clipped at 10.

- **LowRR-HighDET flag:** A binary indicator defined as

$$\mathbf{1}\{\mathrm{RR} < 0.1 \wedge \mathrm{DET} > 0.9\}$$

  used to detect extreme configurations where the system exhibits very low recurrence but almost entirely deterministic trajectories. Such conditions may correspond to strongly structured but highly unstable dynamics, and were hypothesized to act as early-warning markers of critical transitions.

These composite indicators, though simple, were designed to highlight boundary conditions that primary RQA metrics alone may miss, and can serve as potential

early-warning markers of critical transitions. Similar approaches have been explored in previous studies, where RQA-based features successfully differentiated pre-ictal from inter-ictal states and revealed pathological determinism in epileptic brain dynamics (Ngamga et al. 2017; Ouyang et al. 2008).

4.2.3 Tipping Point Definition and Dynamic Phase Definition

Abrupt transitions in brain dynamics were identified on the Recurrence Rate (RR) trajectory. For each recording, RR was first smoothed (as described above) using a centered moving average with window length equal to 25 samples:

$$\mathrm{RR}_{\mathrm{smooth},p,t} = \frac{1}{25} \sum_{k=-12}^{12} \mathrm{RR}_{p,t+k}.$$

The discrete first derivative was then computed as:

$$\mathrm{RR\_derivative}_{p,t} = \mathrm{RR}_{\mathrm{smooth},p,t} - \mathrm{RR}_{\mathrm{smooth},p,t-1}.$$

To account for inter-recording variability, the derivative was z-scored within each recording:

$$\mathrm{RR\_deriv\_zscore}_{p,t} = \frac{\mathrm{RR\_derivative}_{p,t} - \mu_p(\mathrm{RR\_derivative})}{\sigma_p(\mathrm{RR\_derivative}) + \varepsilon}.$$

A tipping point was flagged whenever $|\mathrm{RR\_deriv\_zscore}_{p,t}| > 3$, corresponding to a three-standard-deviation rule commonly adopted for outlier detection.

Dynamic phases were then redefined as follows:

- **Phase 0:** interictal baseline;
- **Phase 1:** interval between the last tipping point occurring before seizure onset and the first ictal window;
- **Phase 2:** ictal.

Based on this segmentation, the *lead time* was defined as the temporal interval between the last identified pre-ictal tipping point and the clinically annotated seizure

onset. This metric quantifies the anticipatory capacity of the method by estimating how far in advance a critical transition is detected prior to seizure occurrence.

4.2.4 Cross-Patient Aggregation and Statistical Testing

To assess group-level patterns, the following procedures were applied:

- Per-patient descriptive statistics (mean, median, standard deviation, quartiles) were computed for each phase.
- Peri-ictal trajectories were constructed by aligning patients' timelines to seizure onset and computing the median across patients at each relative time step.
- Statistical comparisons between phases were performed using the Wilcoxon signed-rank test on per-patient medians. The analysis framework reports paired differences and associated p-values.

4.2.5 Exploratory Analysis of Tipping Points

Before formalizing the tipping point detection framework, an exploratory analysis was performed to define how abrupt transitions in brain dynamics could be operationalized within the recurrence-based feature space. The focus was on identifying whether changes in features (particularly RR and its derivatives) could systematically anticipate seizure onset.

As anticipated, tipping points were defined as time windows where the absolute z-scored derivative of smoothed RR exceeded a fixed threshold ($|z| > 3$). This criterion flagged windows of abrupt change in network dynamics. To incorporate patient-specific variability, a dynamic phase definition was introduced: for each patient, the last tipping point prior to seizure onset was used to delimit the start of a "dynamic pre-ictal" interval, extending until the first ictal window. This dynamic definition was introduced to operationally delimit the pre-ictal interval based on patient-specific temporal markers, instead of relying on a fixed 600-second horizon.

Finally, to quantify the local impact of each detected event, we computed the change in recurrence rate (ΔRR) around the tipping point. For each event, the mean

RR in a short interval before the tipping point was compared with the mean RR in an equally sized interval immediately after it. Events were categorized as upward (RR > 0) or downward (RR < 0) transitions based on the sign of the variation in the recurrence rate computed around each detected change-point.
At this stage, it is important to clarify that the proposed features will be tested with two analytically distinct but complementary strategies.

First, a **supervised modeling track**, designed to assess the statistical separability between pre-ictal and non-pre-ictal windows under strict patient-level validation. This component evaluates performance through distribution-level metrics and focuses on cross-subject generalization.

Second, an **unsupervised dynamical instability track**, aimed at identifying localized transitions in recurrence dynamics through the derivative-based tipping point criterion. This component operates independently of class labels and evaluates seizure-level anticipation and temporal proximity to ictal onset.

In the evaluation section, these two components are evaluated, enabling an integrated interpretation of their complementary predictive and dynamical contributions.

4.3 Summary

This chapter introduced a dynamical-network framework for early seizure prediction based on Recurrence Quantification Analysis (RQA) applied to time-resolved functional connectivity. Starting from raw multi-channel EEG recordings, the pipeline performs signal conditioning and channel harmonisation, sliding-window segmentation, construction of dynamic functional networks, recurrence plot generation over the resulting connectivity-state trajectory, and extraction of recurrence descriptors that characterise stability, predictability, and persistence of network dynamics.

In terms of outputs, the methodology produces two complementary forms of information. First, it generates a supervised learning feature space composed of classical RQA measures (RR, DET, TT) and derived descriptors designed to capture pre-ictal instability, including phase-normalised features, temporal smoothing, short-

term variability measures, temporal derivatives, and composite indicators. Second, it defines an unsupervised tipping point criterion based on statistically anomalous accelerations in the smoothed RR trajectory, enabling the detection of dynamical transition markers without relying on absolute recurrence thresholds.

The predictive utility of these recurrence-based representations is assessed in the experimental chapter under a unified modelling protocol shared across all methodologies. In particular, supervised classification performance is evaluated using gradient boosting models within a strict validation scheme, with nested threshold selection and imbalance-aware metrics, that will be better explained in the discussion and results chapter. In parallel, the tipping point analysis is evaluated through patient-level detection rates and lead time statistics relative to clinically annotated seizure onset. A structured comparative evaluation across the wavelet-based and hybrid configurations is presented in Chapter 6.

Chapter 5

Second Methodology: Time-Frequency Modeling via Discrete Wavelet Transform

This chapter outlines the methodological framework used to implement the second line of inquiry pursued in this thesis. The main aim is to assess whether a classical time–frequency feature extraction approach can be used to reach competitive performance levels for early seizure prediction, thus offering a conceptually different solution to the recurrence-based framework presented in Chapter 4.

The proposed framework is developed as a fully reproducible, end-to-end processing pipeline that takes long-term, multi-channel EEG signals as input and produces calibrated early warning predictions. The pipeline begins with raw EEG signals stored in EDF files and comprises signal preprocessing, temporal segmentation, wavelet-based feature extraction, and supervised learning, all of which are conducted under strict patient-level data separation to prevent information leakage.

The framework draws inspiration from the work of Tran, T. T. Nguyen, and H. T. Nguyen 2022, who explored discrete wavelet transform feature sets and binary particle swarm optimisation for seizure detection on the Bonn EEG dataset. Nevertheless, here the original framework has been substantially extended and reformulated to address the more demanding scenario of real-world scalp EEG recordings

and seizure prediction, indeed the original study focused on seizure detection, i.e., classifying EEG segments as seizure or non-seizure at the time the ictal activity was already present. In contrast, the present thesis targets pre-ictal seizure prediction, which requires identifying subtle dynamical changes preceding clinically annotated onset.

Furthermore, the original pipeline processed single-channel EEG segments under controlled experimental conditions. The proposed framework introduces several essential extensions: it operates on multi-channel long-term recordings, explicitly addresses high inter-patient variability, and is designed to generalize under strict validation. These additions fundamentally shift the problem from segment-level seizure detection to patient-independent early warning modeling.

Lastly, a model-agnostic temporal decision logic based on persistence is introduced to stabilize early-warning predictions over time. Rather than relying on isolated window-level decisions, the framework enforces temporal continuity by requiring repeated evidence of pre-ictal activity before raising an alarm. This design explicitly addresses the high variability and noise of window-level classifiers in long-term EEG recordings and aligns the decision process with the sustained nature of pre-ictal dynamics, while preserving sensitivity under clinically realistic false alarm constraints.

5.1 Overview of the Pipeline

The wavelet-based approach adopts a structured, end-to-end pipeline architecture to transform raw EEG recordings into calibrated early-warning decisions.

The pipeline starts with signal conditioning and channel harmonization to ensure consistency across heterogeneous EEG channel recordings. Raw EDF files are loaded, filtered, resampled to a common sampling rate, and non-EEG channels are removed. Channel names are normalized and synchronized to facilitate consistency across recordings.

The preprocessed signals are then segmented into overlapping temporal windows and annotated using a dataset-specific pre-ictal annotation scheme, which specifies

the temporal context of each window with respect to seizure onset. Although annotation schemes vary across datasets, a standardized representation in seconds from the start of the recording is mandated, facilitating a unified annotation logic across datasets.

Each window is then converted into a multi-scale time-frequency representation using the Discrete Wavelet Transform (DWT) which decomposes the signal into localized frequency components at multiple temporal resolutions. Wavelet coefficients are then used to compute a compact set of statistical, spectral, and dynamical features that capture both amplitude distributions describing how signal energy is distributed within each band, and temporal evolution, characterizing how these components vary over time within specific frequency bands. To complement the wavelet-based features, additional time-domain and spectral features are extracted and aggregated across channels, resulting in a fixed-dimensional representation for each window.

The resulting feature vectors are normalized and rebalanced to compensate for strong class imbalance and inter-patient variability.

5.2 Signal Preprocessing

5.2.1 EDF Import and Signal Conditioning

The raw EEG signals were imported from EDF files using the MNE Python toolbox and underwent a preprocessing procedure that was consistent across cross-validation folds. Each recording was first restricted to EEG channels only (i.e., removing non-EEG channels such as markers or auxiliary signals when present), and then resampled to a common target sampling rate of 256 Hz when necessary. This resampling step ensures that all subsequent windowing and wavelet decomposition operations are performed on a consistent temporal support.

To suppress low-frequency drifts and high-frequency noise while retaining physiologically relevant EEG dynamics, each EEG channel was band-pass filtered between 0.5 Hz and 40 Hz. This filtering step targets the standard frequency range of scalp EEG and prevents contamination of wavelet coefficients by slow baseline fluctuations

or high-frequency artefacts.

5.2.2 Channel Normalisation and Harmonisation

Channel variability poses a major challenge for patient-level evaluation, as the classifier requires a fixed-dimensional input structure. To ensure that extracted features remain directly comparable across patients and recording sessions, channel names were first normalised to a common naming convention, and channel selection was carried out using an intersection-based harmonisation strategy.

Specifically, only those EEG channels consistently present across all retained recordings were included for feature extraction. Channels missing from any subset of recordings were excluded to prevent the construction of session-specific feature spaces. This approach enforces a stable and reproducible input representation throughout the dataset and aligns with the methodological requirements of the supervised validation, where the test subject must be evaluated under the same feature schema as those learned from the training subjects.

Moreover, this harmonisation strategy reflects the channel-consistency principle adopted in the recurrence-based framework introduced earlier in the thesis, thereby supporting a fair comparison between the two pipelines and mitigating confounding factors introduced by montage variability.

5.3 Sliding Window Segmentation and Temporal Labelling

5.3.1 Window Extraction

Following preprocessing, EEG recordings were segmented into overlapping temporal windows using a sliding-window strategy. This approach provides a trade-off between temporal resolution and statistical reliability of the extracted features. Each window had a duration of 10 s and was shifted by 5 s, resulting in a 50% overlap between consecutive segments. Such a configuration allows the extraction of sufficiently long segments for stable time–frequency representations while preserving a

dense temporal sampling suitable for early-warning prediction.

Each window is represented as a multivariate time-series matrix $X \in \mathbb{R}^{C \times T}$, where C denotes the number of harmonized EEG channels and T the number of samples per window. After resampling to 256 Hz, this yields $T = 10 \cdot 256$ samples per segment.

Window duration and stride are design-dependent and were selected to match the requirements of each representation: short, highly overlapping windows are used for connectivity-state tracking and recurrence dynamics, whereas longer windows are adopted for stable wavelet-domain estimates.

5.3.2 Seizure Timeline Parsing and Temporal Alignment

Seizure annotations differ across datasets, ranging from file-level seizure lists to recording-level metadata. To ensure consistent temporal labelling, all seizure onset and offset times were mapped to a unified reference system defined in seconds from the start of each EDF recording. This standardization enables an unambiguous temporal alignment between extracted windows and seizure events, independent of the original annotation format adopted by each dataset.

5.3.3 Pre-Ictal Labelling Scheme

Window labels were derived from seizure onset times through a predefined definition of the pre-ictal period. At the sample level, as previously stated for RQA Modeling, three temporal states were considered:

- **Interictal (0):** baseline EEG activity occurring far from seizure events;
- **Pre-ictal (1):** the period preceding seizure onset, defined by a fixed prediction horizon;
- **Ictal (2):** seizure activity between seizure onset and termination.

To preserve dataset-specific physiological relevance and to avoid excessively broad pre-ictal definitions in shorter recordings, the duration of the pre-ictal period was set to 30 minutes for the CHB-MIT dataset and 10 minutes for the ScalpEEG

dataset. This choice reflects structural differences between the datasets: CHB-MIT recordings typically contain long interictal intervals that support wider pre-ictal contexts, whereas Siena recordings are generally shorter and more concentrated around peri-ictal events, requiring a narrower horizon to limit label overlap and temporal ambiguity.

Window-level labels were obtained by averaging the sample-level labels within each window and rounding the resulting value to obtain a single representative class. Windows overlapping with ictal segments were excluded from both training and evaluation to prevent contamination by seizure-period dynamics and to maintain a strictly predictive setting.

Finally, the learning problem was formulated as a binary early-warning classification task that will be explained in Chapter 6. This formulation encourages the classifier to detect pre-ictal transitions preceding seizure onset rather than merely identifying ongoing seizure activity. These windowed and temporally aligned segments form the basic units on which feature extraction and classification are subsequently performed.

5.4 Discrete Wavelet Transform Feature Extraction

5.4.1 Rationale for Wavelet-Domain Modelling

Given the window-based formulation discussed above, the next step consists in transforming each segment into a compact and informative representation suitable for supervised learning. Scalp EEG recordings from both the CHB-MIT and Siena datasets are inherently non-stationary: clinically relevant changes often emerge as gradual spectral shifts and transient oscillatory events whose temporal evolution carries predictive information. This challenge is further amplified by heterogeneous electrode montages and recording conditions, which introduce substantial cross-subject variability even before modelling. For this reason, the proposed methodology adopts a wavelet-domain representation, which provides multiresolution analysis while preserving temporal localisation, unlike Fourier-based approaches that implicitly assume stationarity within the analysed segment. In an early-warning setting, this

is particularly important because pre-ictal dynamics are not expected to appear as instantaneous events but rather as sustained changes across multiple frequency components over tens of seconds or minutes. Wavelet coefficients offer a compact and structured representation of such cross-scale dynamics within each short window, making them suitable inputs for window-level classifiers under strict patient-level evaluation.

5.4.2 Wavelet Decomposition Setup

For each extracted EEG window $X \in \mathbb{R}^{C \times T}$, where C denotes the number of harmonised EEG channels and T the number of temporal samples, the Discrete Wavelet Transform (DWT) is applied independently to each channel. The Daubechies-4 wavelet is adopted as the mother wavelet due to its compact support, orthogonality, and its effectiveness in representing EEG-like oscillatory patterns, including sharp transitions, without excessive temporal smearing.

Signals are decomposed up to level five using symmetric signal extension to mitigate boundary artifacts at window edges. For each channel, the decomposition produces one approximation component and five detail components:

$$\{A_5, D_5, D_4, D_3, D_2, D_1\}.$$

This multiscale representation separates slower oscillatory content (captured by A_5) from progressively higher-frequency components (captured by D_5 to D_1). Although precise frequency boundaries depend on the sampling frequency and the discrete filterbank, this hierarchy provides a physiologically meaningful partition of the EEG spectrum into coarse-to-fine temporal resolutions, allowing the model to capture both background rhythms and transient dynamics associated with pre-ictal transitions.

5.4.3 Feature Computation per Sub-band and Channel

From each wavelet sub-band and for each EEG channel, a compact set of descriptors is computed to characterise complementary aspects of the wavelet coefficients. Let

c denote the coefficient vector associated with a given sub-band and channel. The extracted feature families include:

- **Magnitude and variability:** mean, standard deviation, and root mean square (RMS) of c, summarising central tendency and dispersion.
- **Energy:** $E = \sum_i c_i^2$, capturing the overall oscillatory power contained in the sub-band.
- **Distributional shape:** skewness and kurtosis (Fisher definition), quantifying deviations from Gaussianity that may reflect abnormal neural dynamics.
- **Signal complexity:** Shannon entropy computed on the normalised absolute coefficients, $p_i = |c_i| / \sum_j |c_j|$, measuring irregularity and information content within the sub-band.
- **Temporal dynamics:** Hjorth activity and mobility computed on the coefficient sequence. Activity corresponds to the variance of c, while mobility is estimated as $\sqrt{\mathrm{var}(\Delta c)/(\mathrm{var}(c) + \varepsilon)}$, providing an interpretable proxy of effective frequency variation.

These descriptors are computed for all six wavelet components $(A_5, D_5, \ldots, D_1)$ and for each EEG channel, yielding a structured multiscale representation that captures amplitude distributions, oscillatory power, non-Gaussian behaviour, complexity, and temporal dynamics within each window.

5.4.4 Cross-Channel Aggregation and Fixed-Dimensional Representation

To enable learning with tabular classifiers and ensure compatibility with strict evaluation planned, channel-wise features are mapped to a fixed-dimensional representation. The feature extraction module supports two complementary configurations:

- **Channel-resolved representation (`aggregate="none"`):** features are retained separately for each channel, producing channel-specific columns of the

form `Band_Feature_chXX`. This representation preserves spatial information at the cost of higher dimensionality.

- **Aggregated representation (`aggregate="mean"` or `"median"`):** for each feature and sub-band, channel-wise values are summarised across channels using the mean or the median, producing a single column `Band_Feature_mean` (or `_median`) per descriptor. This reduces dimensionality and increases robustness to channel-specific amplitude scaling.

This design ensures that each analysis window is mapped to a consistent feature schema regardless of the subject, while still retaining multichannel information through cross-channel summary statistics. As a result, feature dimensionality remains identical across all evaluation folds, including those in which the test patient is completely unseen during training.

5.4.5 Integration with Windowing and Label Alignment

Feature extraction is performed directly on the window tensors produced by the segmentation module, which outputs $X \in \mathbb{R}^{N \times C \times T}$ (with N windows), the corresponding window-level labels, and metadata including window onset times expressed in seconds from the start of the recording. Window-level labels are derived by averaging the sample-level temporal labels within each window and rounding to the nearest discrete state, ensuring a consistent and reproducible mapping between the continuous seizure timeline and the discrete learning problem.

Windows labelled as *ictal* were systematically excluded from downstream training and evaluation in order to enforce a strict early-warning formulation of the task, preventing contamination of the model with seizure-period dynamics and ensuring that predictions are made exclusively before seizure onset.

5.5 Dataset-Specific Adaptations: CHB-MIT vs Siena

Although the wavelet-based pipeline was designed as a unified methodology, minor dataset-specific adaptations were required to account for structural differences between the CHB-MIT and ScalpEEG datasets. These adaptations were introduced

to preserve methodological consistency while ensuring that the extracted features and labels remained physiologically meaningful in both contexts.

5.5.1 Adaptation to the CHB-MIT Dataset

The CHB-MIT dataset consists of long-term scalp EEG recordings with heterogeneous montages and variable inter-seizure intervals. For this reason, the pipeline was configured to support multi-hour recordings and explicit pre-ictal segmentation. A pre-ictal horizon of 30 minutes was adopted, and windows overlapping ictal segments were excluded from training to avoid contamination of the early-warning model.

Channel harmonisation was enforced by retaining only channels common to all recordings of a given patient, ensuring consistent feature dimensionality across sessions. Given the strong class imbalance between interictal and pre-ictal windows, downsampling and class weighting strategies were enabled, and decision thresholds were calibrated to prioritise high sensitivity while controlling the false alarm rate.

5.5.2 Adaptation to the ScalpEEG Dataset

ScalpEEG dataset differs substantially from CHB-MIT in terms of recording duration, electrode configuration, and seizure density. Recordings are shorter and more homogeneous, with fewer hours of interictal activity and more compact peri-ictal segments. As a consequence, the pre-ictal horizon was reduced to 10 minutes, in order to avoid excessive overlap between pre-ictal and interictal states and to ensure sufficient temporal resolution around seizure onset.

Because recordings are shorter and montages more consistent, channel harmonisation required fewer exclusions, and no aggressive downsampling was applied. However, the same windowing and wavelet decomposition parameters were preserved to maintain comparability between datasets. Feature scaling and model training were performed independently for each dataset, but the same LightGBM-based classification and evaluation protocol was applied in both cases.

These controlled adaptations allow the wavelet-based methodology to be evaluated across two datasets with markedly different characteristics, providing a more

robust assessment of its generalisation capabilities and its sensitivity to dataset-specific constraints.

5.6 Feature Scaling and Class Imbalance Handling

To reduce the influence of outliers and inter-patient amplitude variability, features were normalised using a `RobustScaler` fitted exclusively on the training fold of each evaluation iteration.

Class imbalance was addressed through a combination of negative-class down-sampling and class weighting. Specifically, the number of interictal windows was limited to at most twice the number of pre-ictal windows, and the positive class weight was defined as:

$$\text{scale\_pos\_weight} = \min\left(\frac{N_{\text{neg}}}{N_{\text{pos}}}, 5\right)$$

This strategy ensured stable training while preserving clinically meaningful sensitivity.

At this stage, the proposed transformed features are assessed using the same modeling and validation framework applied to the RQA-based representations, in order to preserve methodological coherence and enable a structured comparison.

5.7 Hybrid Integration of Spectral and Dynamical Features

Despite both approaches targeting the same challenge of early seizure prediction, they are based on different modelling paradigms and levels of abstraction. The RQA framework is based on functional brain networks and aims at dynamical transitions between different regimes using recurrence-related features, while the wavelet-based approach models pre-ictal activity at the signal level by leveraging time-frequency representations of the EEG.

Instead of being alternative approaches, the two methods offer complementary views on pre-ictal activity, focusing on different but related aspects of the underlying

neurophysiological mechanisms. This conceptual complementarity encourages the investigation of combined scenarios where the signal-level temporal information and the network-level dynamical information can be combined, as will be discussed in the next section.

5.8 Context-Aware Wavelet Modeling via RQA

Based on the complementary nature of the RQA- and wavelet-based frameworks described above, this section proposes a hybrid setup that combines recurrence-based dynamical context information with the wavelet-domain modeling process. In this context, wavelet-extracted representations are employed to establish the temporal structure and ground truth of pre-ictal dynamics, while RQA features convey supplementary information regarding the dynamical regimes of the associated EEG segments.

The proposed hybrid configuration is introduced as an additional methodological evaluation setting aimed at assessing whether the joint representation of wavelet-domain and recurrence-based features provides a measurable improvement in discriminative performance. Rather than replacing the standalone frameworks, it serves to quantify the incremental contribution of dynamical contextual information when combined with signal-level descriptors within the same classification pipeline.

5.8.1 Time-Aligned Integration of Wavelet and RQA Representations

A fundamental requirement in the integration of the two methodologies is the establishment of a temporally coherent mapping between wavelet-domain features and recurrence-based dynamical descriptors. In practical terms, this requires that feature vectors generated by the two independent pipelines be associated only when they refer to comparable EEG segments in time.

Since the DWT and RQA pipelines are computed independently, their respective window sequences may differ in temporal stride, segmentation scheme, and indexing conventions. Consequently, a direct one-to-one correspondence between feature

vectors cannot be assumed. If left unaddressed, such discrepancies would produce temporally inconsistent feature associations, thereby compromising the validity and reliability of the subsequent classification analyses. To prevent this issue, a time-aligned matching strategy was explicitly imposed, translating the requirement of temporal coherence into a formally defined integration procedure.

The integration was therefore carried out using a multi-step workflow aimed at ensuring temporal consistency and unambiguous window matching.

Step 1: Independent temporal alignment. Temporal alignment was performed separately for each subject and recording file. This precaution prevented erroneous matches across recordings, since EEG time indices reset to zero for each file and are therefore not globally comparable.

Step 2: Identifier normalization. To establish a consistent grouping structure across the two datasets, subject identifiers and recording labels were standardized. This involved converting identifiers to a uniform string format (e.g., consistent case usage), removing extraneous characters or file extensions, and harmonizing naming conventions across pipelines. This normalization step ensured that windows derived from the same physiological recording were uniquely and correctly associated prior to temporal alignment.

Step 3: Label consistency filtering. Only RQA windows linked to valid binary class labels were retained, in accordance with the supervised learning framework adopted in both standalone approaches.

Step 4: Duplicate removal. Duplicate windows sharing the same temporal reference within a recording were removed. Retaining such duplicates would create ambiguous correspondences, allowing multiple RQA windows to compete for the same wavelet window during temporal alignment.

After these structural constraints were imposed, the alignment process itself was cast as a nearest-neighbor matching problem in the time domain. For each wavelet

window at time t, the RQA window with start time $\hat{t}$ closest to t was chosen, as long as the absolute difference in times $|t - \hat{t}|$ was within a specified tolerance. The tolerance level was defined relative to the wavelet window stride according to a half-stride rule, ensuring that matched windows corresponded to overlapping or nearly overlapping EEG segments.

Only window pairs that were successfully aligned were retained for subsequent modelling. The resulting integrated representation preserves the wavelet-derived temporal structure and class label as the reference, while RQA descriptors contribute complementary information about the dynamical regime of the same EEG segment.

Finally, the quality of the alignment procedure was systematically evaluated through quantitative diagnostics. In particular, the distribution of absolute temporal differences $|t - \hat{t}|$ between matched window pairs was analysed to verify that the majority of associations occurred within the predefined tolerance and reflected genuinely overlapping EEG segments. Further, coverage metrics were calculated to determine the proportion of wavelet windows that could be successfully matched with a valid RQA counterpart, as well as the fraction of RQA windows effectively incorporated into the merged representation.This verification step confirmed that the alignment procedure preserved temporal coherence while limiting data loss and avoiding systematic matching distortions.

5.8.2 Feature sets and hybrid configurations

The hybrid feature representation obtained through time-aligned integration was evaluated through a two-level protocol designed to capture both local discriminative performance and clinically relevant predictive behaviour. Within this evaluation framework, different hybrid configurations were considered in order to quantify the impact of distinct integration strategies.

Specifically, two merged feature sets were defined:

- **MERGED_NEAREST**: direct concatenation of DWT features with the temporally nearest RQA descriptors associated with the same wavelet window.
- **MERGED_CONTEXT**: concatenation of DWT features with context-aggregated

RQA descriptors computed over a symmetric temporal neighbourhood around each wavelet window.

These configurations enable a controlled comparison between a strictly local temporal integration strategy and a context-aware dynamical aggregation scheme.

Nearest RQA integration. In the **nearest** configuration, each wavelet window starting at time t is augmented with a single RQA feature vector corresponding to the RQA window whose start time $\hat{t}$ is temporally closest to t, subject to the predefined tolerance constraint described in Section 5.8.1. This integration strategy preserves the original temporal sampling structure of the wavelet pipeline and introduces a strictly local dynamical descriptor associated with the same EEG epoch.

Context-aware RQA integration. While the nearest strategy enforces temporal consistency, it remains sensitive to small stride mismatches or local phase shifts between independently generated window streams. To mitigate this limitation and to capture short-term dynamical trends, a **context-aware** aggregation scheme was introduced.

Let t denote the start time of a wavelet window of duration W seconds. Assume that a higher-resolution RQA feature stream is available, computed at a finer temporal stride. For each wavelet window at time t, all RQA samples with start times contained within the symmetric neighbourhood

$$[t - \Delta,\ t + \Delta]$$

are collected, where Δ is defined relative to the wavelet window length (typically $\Delta = \frac{W}{2}$).

For each base RQA descriptor (e.g., RR, DET, TT, and their normalized or smoothed variants), summary statistics including the median, mean, and standard deviation are computed over the neighbourhood. The resulting aggregated features encode the local dynamical context surrounding the wavelet window rather than a single temporally nearest estimate.

In cases where no RQA samples fall within the specified neighbourhood (e.g., due to temporal gaps or boundary effects), a nearest-sample fallback is applied to ensure continuity and avoid missing values while maintaining temporal consistency.

A comprehensive quantitative comparison between the standalone wavelet framework, the recurrence-based approach, and the proposed hybrid configurations is presented in Chapter 6.

5.9 Summary

This chapter has presented a wavelet-based framework for patient-independent early seizure prediction on multi-channel EEG recordings, grounded in time–frequency feature extraction, relying on localized spectral energy modulations across frequency bands. By integrating the discrete wavelet transform with group-aware model optimization and a temporally consistent decision logic, the proposed approach enables a rigorous and clinically informed assessment of pre-ictal dynamics.

The predictive utility of these representations is assessed in the experimental chapter under a unified modelling protocol shared across all methodologies. In particular, supervised classification is conducted using gradient boosting ensemble classifiers (LightGBM and CatBoost), selected for their robustness in high-dimensional feature spaces, effective handling of class imbalance, and capacity to model complex non-linear relationships. Hyperparameters and decision thresholds are optimized using the Optuna framework within a group-aware cross-validation scheme, ensuring strict patient-level separation and preventing information leakage. Final evaluation is performed under a Leave-One-Patient-Out (LOPO) protocol to provide a clinically realistic estimate of cross-subject generalization.

In addition to the standalone wavelet formulation, the chapter introduces a time-aligned integration with recurrence quantification analysis (RQA), allowing spectral descriptors to be complemented with dynamical contextual information. This hybrid modeling strategy provides a structured framework to examine whether combining signal-level and dynamical representations enhances predictive performance.

In the broader scope of this thesis, wavelet and hybrid configurations serve as a signal-based complement to the previously introduced recurrence-based network analysis, contributing to a unified modeling perspective where spectral and dynamic characteristics are evaluated according to a consistent optimization and validation protocol.

Chapter 6

Modeling, Experimental Results, Evaluation and Discussion

This chapter reports the experimental evaluation of the methodological frameworks described in the previous chapters. After defining the feature extraction pipelines for recurrence-based (RQA) and wavelet-based (DWT) representations, the focus now shifts to their quantitative evaluation under controlled and reproducible modeling conditions.

The chapter is structured to clearly separate feature configuration, modeling strategy, and empirical results. Because the modeling and validation framework is shared across the RQA, DWT, and Hybrid feature sets, it is introduced only once in a dedicated section for clarity and consistency. Section 6.2 describes the common experimental protocol, including the Leave-One-Patient-Out (LOPO) validation scheme, performance metrics, imbalance handling, and hyperparameter optimization procedures. All models are assessed using controlled validation procedures to ensure statistical robustness, reproducibility, and clinical relevance.

Subsequently, the experimental analysis is organized into four main methodological branches. Sections 6.3.1, 6.3.2, and 6.3.3 report the results obtained for the recurrence-based framework (RQA), the wavelet-based time–frequency model-

ing approach (DWT), and their hybrid integration, respectively. Finally, Section 6.4 provides a comparative analysis of the methodologies, highlighting statistical behavior, robustness across datasets, and biomedical findings and implications.

The first part presents the results of the recurrence-based framework (RQA), beginning with a controlled single-patient validation and subsequently extending to a multi-patient change-point setting. The second part evaluates the wavelet-based early-warning pipeline (DWT). The third part examines the hybrid configurations obtained through the time-aligned integration of wavelet and recurrence descriptors and the fourth part provides a comparative discussion of the results, highlighting the methodological, statistical, and biomedical implications that emerge from the previously presented analyses. All methodologies were tested on two independent EEG datasets under a strict Leave-One-Patient-Out (LOPO) protocol, enabling a direct comparison between standalone and integrated approaches.

For all supervised classification experiments, hyperparameters are optimized using Optuna with repeated runs to account for stochastic variability. Each configuration is evaluated multiple times with different random seeds to obtain stable performance estimates.

At the window level, performance is quantified using Precision–Recall Area Under the Curve (PR-AUC) and Adjusted Balanced Accuracy (ABA), where BA denotes balanced accuracy. ABA (equivalent to Youdens J) explicitly captures the balance between sensitivity and specificity and is particularly informative under severe class imbalance, as commonly observed in seizure prediction tasks.

When applicable, performance is additionally evaluated at the event level using Event Sensitivity and Time-to-Warning (TTW), thereby aligning the quantitative analysis with the clinical objective of timely seizure anticipation.

6.1 Feature Configurations Summary

This chapter evaluates the proposed frameworks under three feature configurations derived from the methodological chapters.

- **Feature Set A:** feature configurations derived from the RQA pipeline, de-

scribed in Chapter 4.

- **Feature Set B (DWT):** timefrequency features obtained through Discrete Wavelet Transform, representing localized spectral power variations across frequency bands, described in Chapter 5.
- **Combined Set (RQA+DWT):** hybrid configuration integrating dynamical recurrence descriptors with timefrequency spectral features, described at the end of the Chapter 5.

The rationale behind this comparison is to investigate whether dynamical network descriptors (RQA) and timefrequency spectral features (DWT) capture distinct aspects of the pre-ictal transition, and whether their combination enhances predictive performance in a patient-independent setting.

6.2 Modeling and Evaluation Strategy

After completing the preprocessing and feature extraction stages, the study enters its central phase: supervised modeling for pre-ictal state detection. The objective of this stage is to determine, on a window-by-window basis, whether a patient is transitioning into a pre-ictal condition.

To ensure a fair comparison between feature representations, a unified modeling framework was adopted across all configurations, including Recurrence Quantification Analysis (RQA), Discrete Wavelet Transform (DWT), and their hybrid integration. As already stated in previous chapter's summary sections, the same data partition strategy, validation protocol, hyperparameter optimization procedure, and decision threshold selection mechanism were applied in every methodology considered.

By enforcing strict methodological consistency across feature configurations, any observed performance differences can be interpreted as arising from the intrinsic representational properties of the features rather than from procedural variability.

This standardized framework provides a controlled and reproducible basis for assessing cross-patient generalization, robustness to dataset heterogeneity, and clinical

relevance of the proposed methodologies.

6.2.1 Binary Classification Formulation

The modeling pipeline was formulated as a binary classification task, consistently applied to both the Siena and the CHB-MIT datasets. The objective was to discriminate pre-ictal EEG segments (positive class) from all other phases (interictal and ictal combined, negative class) using the feature representations derived from each methodological pipeline. Formally, the classification target was defined as:

$$y = \begin{cases} 1 & \text{if Label} = \text{pre-ictal}, \\ 0 & \text{if Label} \in \{\text{interictal, ictal}\} \end{cases}$$

This binary formulation served as the main methodological framework for both datasets and was implemented using two gradient boosting tree models: *LightGBM* and *CatBoost*. The same modeling strategy was applied consistently to both the Siena ScalpEEG and CHB-MIT cohorts under the LOPO validation scheme, ensuring full comparability of results across datasets.

6.2.2 Leave-One-Patient-Out Validation

To rigorously assess cross-patient generalization, a *Leave-One-Patient-Out* (LOPO) validation strategy was adopted, as commonly recommended in biomedical prediction studies (Loh and Smith 2022). In each outer fold, all recordings from one patient were held out for testing, while data from the remaining patients were used for training and internal validation. This design enforces strict subject-level independence and provides a realistic estimate of model performance on previously unseen individuals.

Within each LOPO fold, hyperparameter tuning and decision threshold selection were performed exclusively on the training patients through a nested validation procedure. No information from the held-out subject was used during model selection, thereby preventing information leakage and ensuring a valid patient-independent evaluation.

Clinical usability was assessed through additional temporal metrics. *Lead time* was defined as the interval between the last detected pre-ictal decision and the annotated seizure onset. In addition, the *False Alarm Rate per hour* (FAR/h) and the proportion of time spent under alarm were computed to characterize alert burden. When specified, a k-consecutive decision rule ($k = 3$ by default) was applied to reduce isolated false positives while preserving sensitivity to sustained pre-ictal activity.

The robustness of the model was further examined by varying key methodological parameters, including window length and overlap, recurrence thresholding strategies, smoothing window size, and composite indicators. These analyses allowed us to assess performance stability across configuration choices.

The overall evaluation workflow is schematically illustrated in Figure 6.1, which summarizes the nested LOPO procedure adopted throughout this study. The diagram highlights the interaction between hyperparameter optimization, inner validation, threshold selection, and final evaluation on the held-out patient, emphasizing the strict separation between training and testing phases.

Together, these components define a rigorous and fully reproducible evaluation framework, ensuring that performance comparisons across feature representations remain methodologically controlled and clinically interpretable.

6.2.3 Model Families

Within the LOPO evaluation framework described above, two tree-based classifier families were considered to assess the robustness of the proposed methodology across different learning algorithms. All models were trained, validated, and evaluated using identical data splits, preprocessing steps, and threshold selection procedures, ensuring that differences in performance could be attributed to the model family rather than to methodological variations.

The binary classification pipeline included the following models:

- `LightGBMClassifier`, using `scale_pos_weight` and early stopping;
- `CatBoostClassifier`, using native `class_weights` and optional GPU accel-

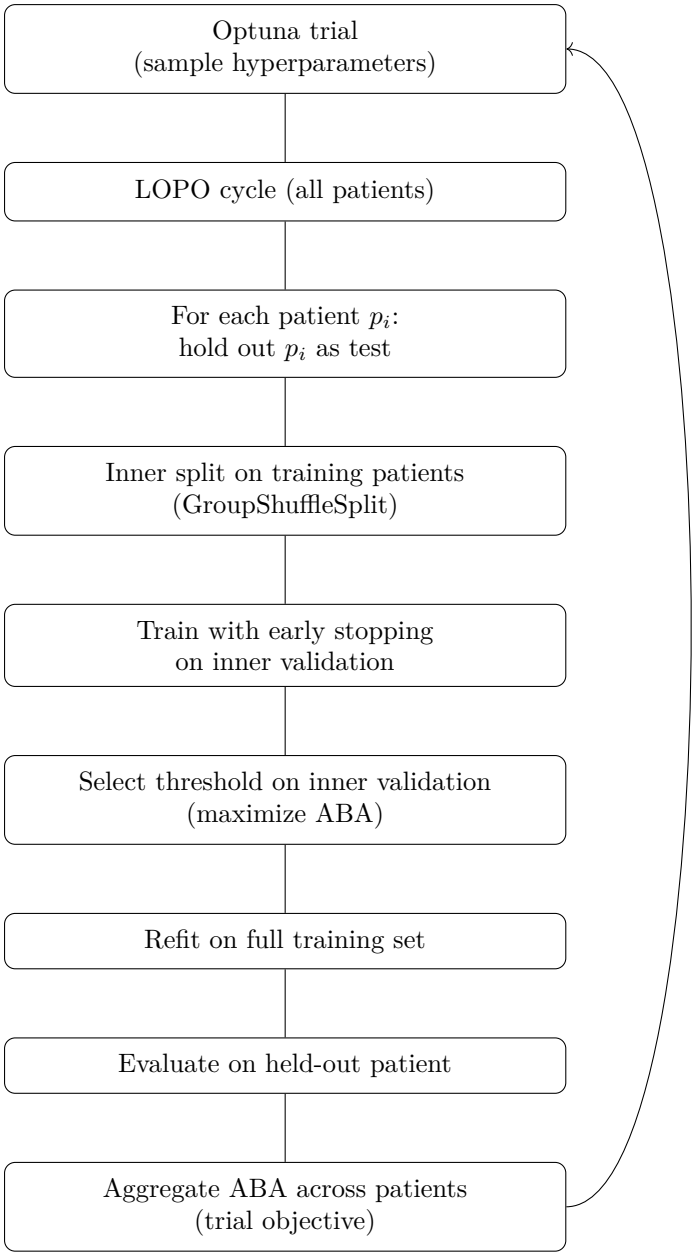


Figure 6.1: Schematic overview of the evaluation pipeline.

eration.

Both models are well suited to high-dimensional tabular features and robust to correlated inputs. All feature sets were evaluated under identical model configurations to ensure comparability.

6.2.4 Imbalance Handling and Operating Point

Given the strong imbalance between inter-ictal and pre-ictal segments, standard accuracy is not an informative performance metric, as high values can be achieved by favoring the majority class. For this reason, model selection and decision thresholding were guided by the *Adjusted Balanced Accuracy* (ABA, also known as Youden's J):

$$\mathrm{ABA} = \mathrm{TPR} + \mathrm{TNR} - 1,$$

which assigns equal importance to sensitivity (correct detection of pre-ictal segments) and specificity (correct rejection of inter-ictal segments). Since balanced accuracy is defined as $\mathrm{BA} = (\mathrm{TPR} + \mathrm{TNR})/2$, it follows that $\mathrm{ABA} = 2, \mathrm{BA} - 1$. ABA therefore represents a linearly rescaled version of balanced accuracy centered at zero, where positive values indicate performance above chance and negative values indicate performance below chance.

Within each LOPO fold, the decision threshold was selected on an inner validation split by maximizing ABA, thereby balancing false negatives and false positives. When clinically required, an additional minimum-recall constraint was imposed on the positive (pre-ictal) class to prioritize the detection of impending seizures, at the cost of an increased false positive rate. The selected threshold was then applied unchanged to the held-out test patient.

6.2.5 Hyperparameter Optimization with Optuna

Hyperparameter optimization was performed using the `Optuna` library with the Tree-structured Parzen Estimator (TPE) sampler and was restricted exclusively to tuning model hyperparameters. Feature representations were fixed prior to this stage, and no feature construction or feature selection was performed during optimization. Optuna therefore operated solely as an optimization layer between the precomputed feature sets and the classification models.

Optimization was embedded within the nested Leave-One-Patient-Out (LOPO) cross-validation framework. For each outer fold, hyperparameter trials were conducted using only the training data, which were further divided into inner training

and validation splits at the patient level to prevent information leakage. Up to 1,000 trials were executed per study, resulting in approximately 10,000 trials across all experiments. Trials yielding unstable behavior (e.g., NaN values) were discarded.

The search space included the learning rate (10^{-3}–10^{-1}), tree depth (3–8), `l2_leaf_reg` (1–10), and the number of training iterations (1000). For each trial, models were trained on the inner training subset with early stopping monitored on the validation set to reduce overfitting and computational cost.

Validation predictions were used to compute class probabilities and determine an optimal decision threshold by maximizing the Adjusted Balanced Accuracy (ABA). The resulting ABA score served as the objective function for optimization. Trials exhibiting unstable behavior or producing NaN values were discarded.

The best-performing hyperparameter configuration was subsequently retrained on the full training fold and evaluated on the held-out patient. This procedure ensured a strict separation between hyperparameter tuning, threshold selection, and final model evaluation.

Final Training with Best Parameters and Evaluation. Once the optimal hyperparameter configuration was identified, a complete LOPO evaluation was performed. For each fold, the model was re-instantiated with the selected parameters and retrained on the corresponding training set. Early stopping was again applied using the inner validation split, after which the model was refitted on the entire training fold with the optimal number of boosting rounds (or trees). The decision threshold determined during the inner validation phase was then re-applied without modification, and predictions on the held-out patient were used to generate the final evaluation metrics. These included Balanced Accuracy, Adjusted Balanced Accuracy, precision, recall, F1-score, PR-AUC, and the patient-specific decision threshold.

Post-Optimization Analysis and Interpretability Beyond performance evaluation, the trained models were analyzed to improve interpretability. For all tree-based classifiers, global feature importance scores were computed to identify which

RQA metrics contributed most consistently to classification decisions across patients. For LightGBM and CatBoost models, additional interpretability was obtained through SHAP (SHapley Additive exPlanations) analysis. SHAP values were computed both at the global level, through summary plots highlighting overall feature relevance and directionality, and at the patient level, where they explained how individual EEG-derived features contributed to specific predictions. This dual-level analysis facilitated the distinction between features that were consistently predictive across the cohort and those that were relevant only for particular patients.

6.2.6 RQA Modeling Adjustment

Although the overall methodological framework was kept consistent across datasets, selected post-processing steps were enabled or disabled to account for differences in recording conditions and seizure dynamics.

For the CHB-MIT dataset, post-processing and calibration were enabled to mitigate the higher rate of false positives observed in preliminary analyses and to stabilize temporal predictions. In contrast, for the Siena dataset, post-processing was deliberately omitted in order to preserve the raw decision behavior of the classifiers and enhance interpretability in relation to signal processing analyses. Due to the higher feature dimensionality and computational demands of this dataset, the hyperparameter search space in Optuna was also narrowed to balance model performance and computational feasibility.

Post-processing and Calibration After model training and threshold selection, optional post-processing steps were applied to the predicted probabilities to improve temporal stability and reduce spurious detections. These procedures were applied only at the inference stage and did not affect model training.

Causal smoothing was applied using a moving average over the current and preceding windows, with a fixed window length of $w = 7$. To enforce temporal consistency, a k-consecutive rule was introduced, requiring at least $k = 3$ consecutive windows exceeding the decision threshold before flagging a pre-ictal state. Finally, isotonic calibration was applied on the inner validation set to recalibrate predicted

probabilities and improve the interpretability of decision thresholds.

These mechanisms were selectively used in the CHB-MIT experiments, where seizure dynamics exhibited greater variability and false positives were more problematic. In contrast, for the Siena dataset, calibration and smoothing were deliberately omitted in order to preserve the raw behavior of the classifiers and highlight their intrinsic discriminative power without additional post-processing. These steps were applied only at inference time and are explicitly reported as dataset-specific post-processing choices, without affecting training or hyperparameter selection.

Additional Analysis: Feature Configurations Although the core methodological framework remained unchanged, for RQA modeling, an additional exploratory analysis was conducted to assess the impact of feature set composition on cross-patient performance. This analysis was not part of the primary evaluation pipeline, but was designed to investigate feature stability and redundancy across subjects. Two alternative feature configurations were considered:

- **Full feature set:** this configuration included the complete set of RQA-derived features and their transformations used throughout the main experiments. In addition to the baseline RQA metrics (RR, DET, TT), it comprised phase-normalized variants, temporally smoothed features, first-order derivatives, variability descriptors, and composite indicators derived from combinations of RR and DET.

- **Reduced feature set:** this configuration was constructed by retaining only features that consistently exhibited high relevance across preliminary analyses based on feature importance scores and SHAP values. In particular, this subset emphasized normalized and smoothed recurrence measures (e.g., phase-normalized DET and RR, RR smoothing and variability descriptors), while excluding features that showed negligible or unstable contributions across patients.

Importantly, no automated feature selection was performed during model training. Preliminary experiments were conducted to compare the performance ob-

tained using the complete (full) feature set and a reduced subset derived through interpretability-driven feature screening and redundancy analysis.

Based on these observations, and in order to reduce model complexity while enhancing interpretability, the reduced configuration was defined prior to the final optimization stage and subsequently evaluated under the same LOPO validation framework, with independent hyperparameter tuning and threshold selection.

Since, accross datasets and model configurations, the full feature set did not yield consistent or substantial performance improvements, the results obtained with the full feature set are not reported in detail later in this chapter; unless otherwise specified, all reported results refer to the reduced feature configuration.

6.2.7 DWT Modeling Adjustment - Temporal Persistence Mechanism

In the wavelet-based (DWT) modeling pipeline, window-level predictions were further refined using a temporal persistence mechanism. This procedure was introduced specifically for the DWT framework to mitigate isolated false positives observed in preliminary analyses.

Because DWT features capture localized timefrequency fluctuations within individual EEG windows, the resulting window-level predictions may appear temporally fragmented, with isolated positive detections that do not reflect sustained pre-ictal activity. To address this issue, the probabilistic outputs of the gradient boosting classifiers were post-processed using a temporal persistence mechanism. Rather than triggering an alarm based on a single positive window, an alert was generated only when pre-ictal evidence was consistently detected across multiple consecutive windows. This approach reduces the influence of sporadic threshold exceedances and ensures that alarm activation reflects temporally sustained signal alterations rather than transient fluctuations.

Formally, given a sequence of predicted probabilities p_t, a binary alarm is triggered at time t only if

$$\sum_{i=t-k+1}^{t} \mathbb{I}(p_i \geq \theta) = k,$$

where k denotes the persistence length and θ the decision threshold. This rule enforces temporal consistency by requiring repeated threshold exceedances rather than relying on a single high-confidence prediction.

In this work, persistence lengths of $k = 3$ and $k = 5$ are evaluated. Final results are reported for $k = 5$, which provides the best trade-off between sensitivity and specificity as measured by the Informedness metric (ABA). Given a window length of 10 s and a stride of 5 s, this configuration corresponds to approximately 25–30 s of continuous evidence prior to alarm generation, reflecting a deliberate balance between false-positive suppression and detection latency.

The temporal persistence mechanism introduces effective temporal smoothing that aligns with the clinical intuition that pre-ictal dynamics are sustained over time rather than instantaneous. This significantly improves robustness to noise in window-level predictions and enhances the stability of the final alarm decisions.

Importantly, the same persistence logic is applied identically to the outputs of both LightGBM and CatBoost classifiers. This ensures that the decision layer remains model-agnostic and that observed performance gains cannot be attributed solely to the choice of the underlying learning algorithm. Importantly, the procedure was not part of the RQA-based modeling pipeline, where temporal dynamics are already embedded in the recurrence-based feature representation.

6.2.8 Derivative-Based Unsupervised Tipping Point Criterion

In addition to supervised classification, an unsupervised dynamical instability analysis was performed using a derivative-based tipping point detection procedure. The criterion, that has already been explained in detail in Chapter 4 was defined *a priori* and applied identically across all subjects and datasets (Siena and CHB-MIT), ensuring methodological consistency and cross-cohort comparability.

The threshold and the patient-specific z-score were fixed across all subjects and was not tuned with respect to seizure onset timing. Consequently, detected events represent intrinsic dynamical instabilities of the recurrence trajectory rather than post-hoc alignment with ictal onset.

Within a dynamical systems perspective, seizure generation has been interpreted

as a transition between distinct brain states (Jirsa et al. 2014). In this context, abrupt and statistically rare changes in recurrence dynamics may indicate transient destabilizations of the underlying functional connectivity regime. Importantly, the proposed criterion does not assume that a seizure is imminent; rather, it identifies phases of increased dynamical instability, which in a subset of patients may precede ictal onset.

Event-Level Evaluation

While window-level classification assesses the model's ability to discriminate individual EEG segments, seizure prediction is inherently an event-oriented task, where the clinically relevant objective is the timely anticipation of seizure onset. Accordingly, model performance is additionally evaluated at the seizure level by transforming window-level probabilistic outputs into event-level indicators defined within a predefined *Seizure Prediction Horizon* (SPH).

For a seizure event with onset time t_{onset}, the SPH is defined as the interval

$$[t_{\text{onset}} - \text{SPH},\ t_{\text{onset}}),$$

where SPH is fixed across all evaluations. To ensure temporal consistency, each prediction is associated with the temporal centre of its corresponding analysis window.

A window is considered to generate a *warning* if its predicted probability exceeds a predefined decision threshold. A seizure event is considered *detected* if at least one warning occurs within its SPH. Event-level sensitivity is therefore defined as the proportion of seizure events for which at least one warning is produced within the corresponding prediction horizon.

For detected seizures, the *Time-to-Warning* (TTW) quantifies the anticipation capability of the model and is defined as

$$\text{TTW} = t_{\text{onset}} - t_{\text{first\_warning}},$$

where $t_{\text{first\_warning}}$ denotes the earliest warning time within the SPH. TTW is un-

defined for undetected events and is summarized across detected seizures using descriptive statistics.

To characterize the relationship between detection sensitivity and alarm policy, two complementary thresholding paradigms are considered. First, a **patient-calibrated threshold** is employed, in which the operating point is determined individually for each subject during model evaluation. Second, a **percentile-based thresholding** strategy is adopted, whereby the decision threshold is defined as a high percentile of the predicted probability distribution, computed either globally across subjects or separately per subject. This second strategy provides a standardized and model-agnostic operating condition for sensitivity analyses.

6.3 Results Obtained

Data Governance and Ethics All EEG data were used under the original datasets' licenses and de-identification guarantees. No re-identification was attempted. Analyses were conducted on anonymized IDs only; no clinical decisions were made and no clinical risk is implied by these analyses.

Reproducibility and Computing Environment Experiments were conducted on Ubuntu 22.04 using Python 3.13 and the following libraries: NumPy, SciPy, MNE-Python, pyts, scikit-learn, LightGBM, CatBoost, and Optuna.

6.3.1 First Methodology - RQA Modeling

This section reports the quantitative performance of the recurrence-based early warning framework described in Chapter 4, evaluated under the experimental protocol detailed in Section 6.2.

Results are first presented for a single-patient approach, followed by the multi-patient setting, ending with dataset-specific analyses.

Rationale Behind the Single-Patient Approach

Starting with a single patient enabled a controlled validation of the RQA pipeline and a direct visual inspection of the feature trajectories over time. This step ensured that the recurrence-based metrics were stable, interpretable, and consistent with clinically annotated seizure events. Moreover, it provided a foundation for extending the analysis to a multi-patient setting, where variability in seizure morphology, recording setups, and subject-specific dynamics introduces substantially higher complexity.

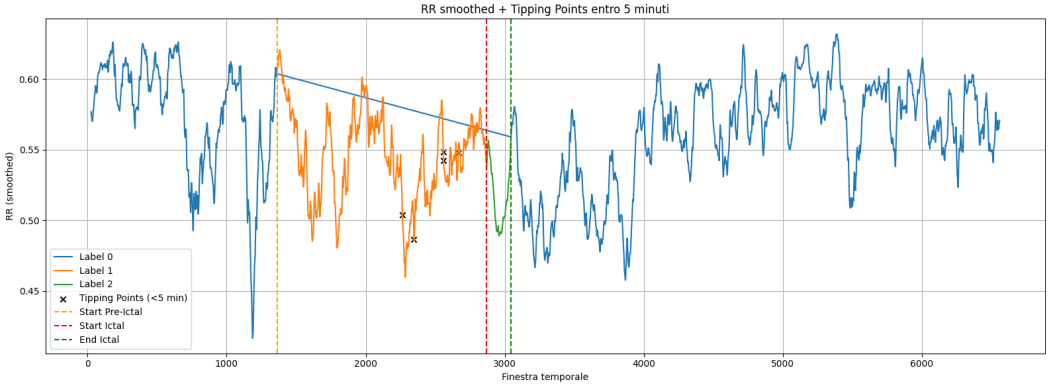


Figure 6.2: Temporal evolution of the smoothed Recurrence Rate (RR) for a representative patient. The pre-ictal and ictal phases are marked by vertical dashed lines. Black crosses denote detected tipping points occurring within 5 minutes prior to seizure onset.

Single-Patient Feature Dynamics Figure 6.2 shows the temporal evolution of the RR for one representative subject. The x-axis corresponds to consecutive EEG windows (2 s length, 80% overlap), and the y-axis reports the RR value, which quantifies the proportion of recurrent states in the reconstructed phase space.

Vertical dashed lines mark clinically annotated transitions, including the beginning of the pre-ictal phase and seizure onset. These annotations allow a direct visual comparison between changes in recurrence dynamics and seizure timing.

Across subjects, RR often displays noticeable changes during the pre-ictal phase. In several cases, a gradual decrease or increased variability is observed prior to seizure onset. This pattern suggests a progressive alteration of the underlying dynamics as the system approaches the ictal state, in line with previous recurrence-based studies of epilepsy (Le Van Quyen et al. 2001; Lopes et al. 2021).

Although RR reflects variations in recurrence density, it does not fully describe the structure of the dynamics. For this reason, additional RQA measures were examined, particularly Determinism (DET) and Trapping Time (TT). Increases in DET may indicate more predictable temporal patterns, possibly associated with transient reorganization of the system. Elevated TT values, especially during seizures, reflect longer persistence in specific states. Considered together, these metrics provide a more complete picture of seizure-related changes at the single-subject level.

Multi-Patient Approach

The multi-patient evaluation was conducted on two independent cohorts. The CHB-MIT dataset included 12 patients with at least one annotated seizure and sufficient pre-ictal data for modeling. The Siena cohort included 9 patients meeting the same inclusion criteria, with adequate pre-ictal coverage for LOPO evaluation. Two complementary analyses were performed:

- **Supervised classification**, aimed at discriminating pre-ictal windows from interictal/ictal activity using recurrence-based descriptors.
- **Unsupervised tipping point detection**, aimed at identifying abrupt dynamical transitions in the RR trajectory preceding seizure onset.

Performance was computed at the patient level and aggregated across folds. For each model family, multiple independent Optuna studies (1,000 trials each) were performed to evaluate both peak performance and run-to-run stability. All metrics are reported as mean ± standard deviation across independent runs, where each run aggregates patient-level results under the LOPO protocol.

CHB-MIT Results

Supervised Classification. **LightGBM** achieved a mean balanced accuracy of 0.544 ± 0.014, corresponding to an adjusted balanced accuracy (ABA) of 0.087 ± 0.028, and a PR-AUC of 0.248 ± 0.023. The best-performing configuration reached a balanced accuracy of 0.560 (run-level mean across patients). Decision thresholds

selected through inner validation typically ranged between 0.20 and 0.25, reflecting explicit informedness optimization under strong class imbalance.

CatBoost achieved a mean balanced accuracy of 0.554 ± 0.006, yielding an ABA of 0.107 ± 0.012 and a PR-AUC of 0.256 ± 0.024 across 10 independent runs. Variance across runs was consistently lower than for LightGBM, indicating greater robustness to hyperparameter initialization. Selected thresholds were generally higher (mean approximately 0.43), corresponding to a more conservative operating regime.

Pre-Ictal Tipping Point Flagging. Using the derivative-based recurrence rate (RR) detection framework ($|z| > 3$), pre-ictal tipping points were identified in 91.7% of analyzed patients in the CHB-MIT cohort. Detection was performed using a fixed, patient-specific z-score threshold applied to the smoothed RR derivative, without subject-wise tuning.

In CHB-MIT, lead times were shorter compared to the Siena cohort, with a median of 0.40 minutes and an interquartile range of [0.11,4.24] minutes. From a clinically oriented perspective, 45.5% of patients exhibited anticipation margins of at least 30 s, 45.5% at least 1 min, and 36.4% at least 2 min.

While detection rates were comparable across cohorts, the results for this dataset exhibited systematically shorter median lead times, suggesting either faster pre-ictal transitions or increased temporal variability in seizure dynamics. This difference may reflect the greater heterogeneity and higher noise levels typically observed in the CHB-MIT recordings.

Analysis of ΔRR values around detected tipping events revealed both upward and downward transitions, indicating that seizure onset is preceded by heterogeneous dynamical patterns rather than a single, consistent trajectory across subjects.

Tipping Density and Supervised Classification Metrics. To investigate whether dynamical instability aligned with supervised classification performance in the CHB-MIT cohort, Spearman's rank correlation coefficients (ρ) were computed between the number of detected pre-ictal tipping points per patient and Adjusted Balanced Accuracy (ABA), Balanced Accuracy (BA), and PR-AUC obtained under the LOPO

validation scheme.

Spearman correlation was selected due to the limited sample size at the patient level and to avoid assumptions of linearity or normality in the relationship between tipping density and performance metrics.

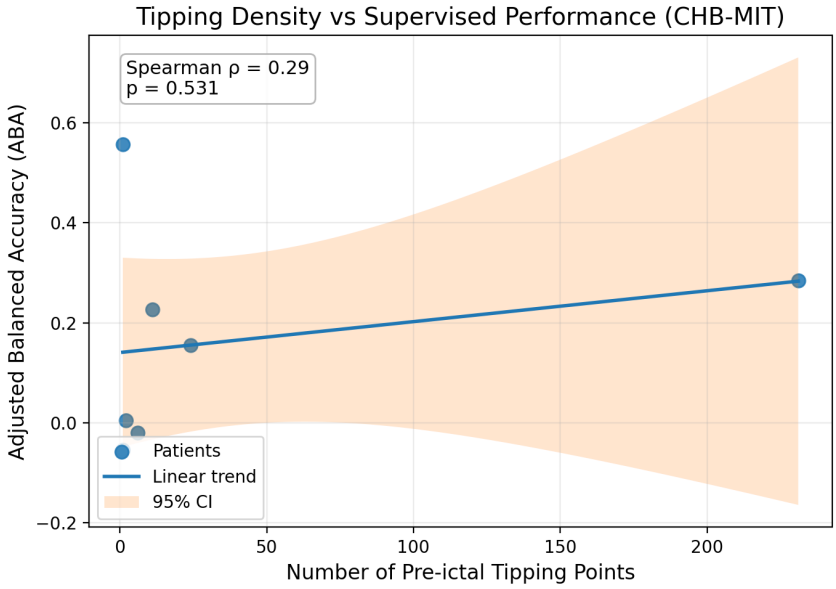


Figure 6.3: Association between pre-ictal tipping density and supervised performance (CHB-MIT cohort). Each point represents one patient. The solid line indicates a linear trend with 95% confidence band.

As illustrated in Figure 6.3, the dispersion of patient-level points and the wide confidence band further support the absence of a clear monotonic trend.

No strong monotonic association emerged between tipping density and supervised performance metrics. In particular, several patients achieving high classification performance exhibited only a limited number of detected tipping events, while some patients with pronounced dynamical instability did not necessarily correspond to the highest supervised scores.

Given the pronounced class imbalance of the task, supervised performance primarily reflects global separability between pre-ictal and interictal distributions across the recording. In contrast, tipping density captures transient dynamical instability localized in time. The absence of a strong monotonic relationship suggests that tipping density and supervised performance capture distinct aspects of the pre-ictal transition.

Visual Representation of Tipping Points (CHB-MIT). To qualitatively illustrate inter-subject variability in pre-ictal dynamical transitions within CHB-MIT, patients were stratified into three representative categories —*weak*, *medium*, and *strong*— based on the total number of detected pre-ictal tipping points.

For each patient, the number of tipping events occurring during the annotated pre-ictal phase was computed. Patients were subsequently ranked according to the density of detected pre-ictal tipping points. Representative cases were selected from the lower, intermediate, and upper portions of this empirical distribution in order to illustrate qualitatively distinct dynamical behaviors.

In this context:

- **Strong-transition patients** exhibit a high density of pre-ictal tipping points, reflecting pronounced instability in recurrence dynamics prior to seizure onset.
- **Intermediate-transition patients** show a moderate number of detectable pre-ictal transitions.
- **Weak-transition patients** display few or sparse pre-ictal tipping events, suggesting subtler or less detectable dynamical destabilization.

It is important to emphasize that this categorization is relative within the analyzed cohort and does not correspond to clinical severity or seizure intensity. Rather, it captures the extent to which recurrence-based indicators reveal measurable dynamical transitions preceding seizure onset.

The following Figures 6.4–6.6 present full-length and zoomed representations of the recurrence dynamics for each representative subject, highlighting the spatial distribution of pre-ictal tipping points relative to the annotated seizure onset.

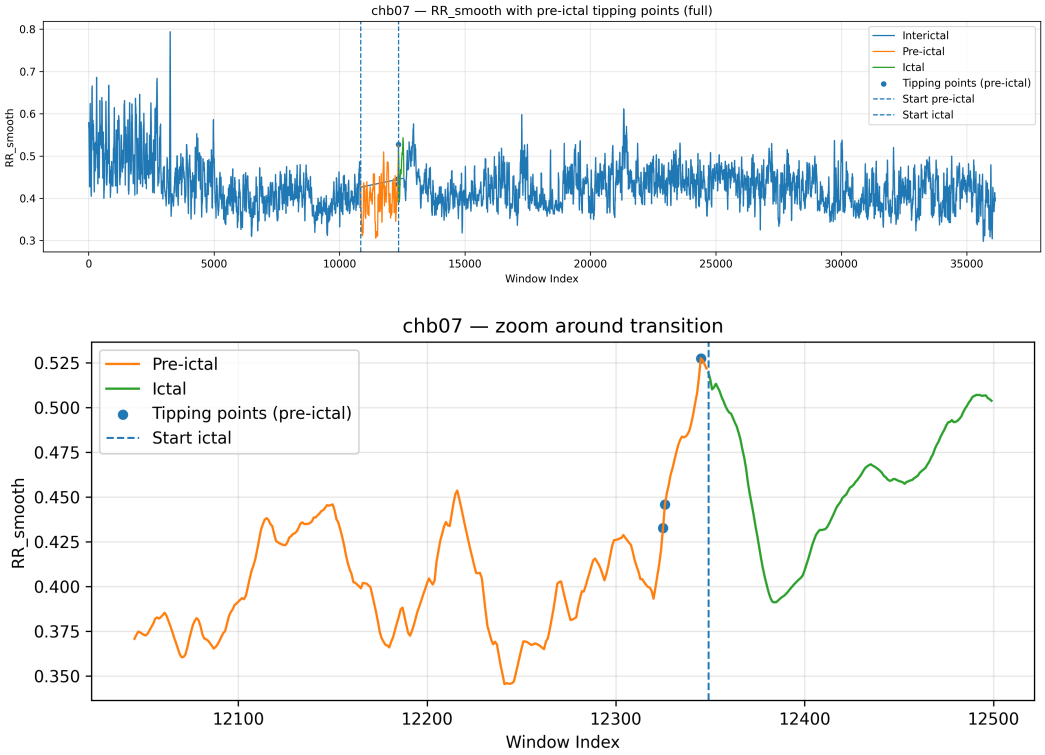


Figure 6.4: Representative weak-transition subject (patient ID: chb07). Smoothed recurrence rate (RR) trajectory with derivative-based tipping point detection ($|z| > 3$). Only limited pre-ictal dynamical instability is observed.

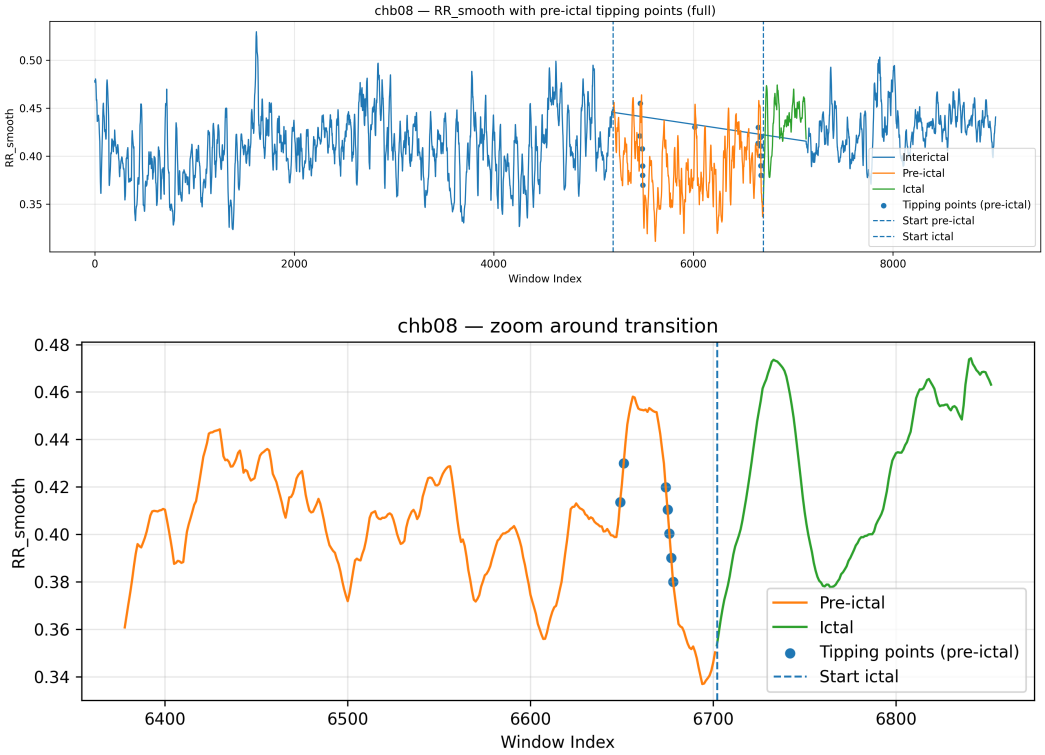


Figure 6.5: Representative intermediate-transition subject (patient ID: chb08). A moderate density of pre-ictal tipping events is observed, reflecting intermediate recurrence instability.

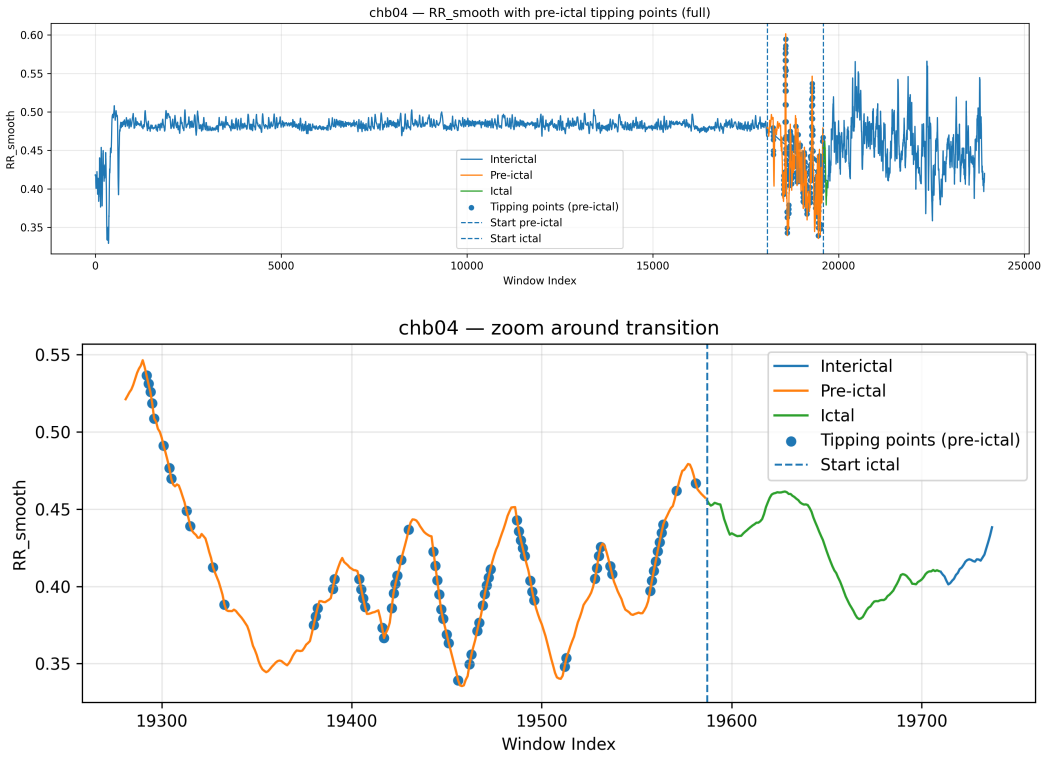


Figure 6.6: Representative strong-transition subject (patient ID: chb04). Numerous pre-ictal tipping points are detected, indicating marked recurrence instability preceding seizure onset.

Overall, these examples further confirm the heterogeneous nature of recurrence-based dynamical transitions. While some CHB-MIT patients exhibit pronounced and clearly identifiable pre-ictal destabilization patterns, others show more subtle or sparse transitions.

Importantly, although the overall detection rate in CHB-MIT (91.7%) was comparable to that observed in Siena, lead times were systematically shorter, suggesting either faster pre-ictal transitions or increased temporal variability in seizure dynamics within this cohort.

ScalpEEG Results

Supervised Classification. **LightGBM** was evaluated over 10 independent optimization runs and achieved a mean balanced accuracy of 0.570 ± 0.009, corresponding to an ABA of 0.139 ± 0.017 and a PR-AUC of 0.253 ± 0.011. The best-performing configuration reached a balanced accuracy of 0.582 (run-level mean

across patients). Decision thresholds were consistently low (mean $\approx$ 0.13–0.14), reflecting informedness-driven calibration under pronounced class imbalance.

CatBoost was evaluated over 10 independent runs and achieved a mean balanced accuracy of 0.560 ± 0.002, yielding an ABA of 0.119 ± 0.004 and a PR-AUC of 0.289 ± 0.009. The best-performing run achieved a balanced accuracy comparable to LightGBM, while selecting systematically higher decision thresholds (mean $\approx$ 0.33–0.35), indicating, even in this case, a more conservative operating regime.

Pre-Ictal Tipping Point Flagging Using the derivative-based criterion defined in Section 4, recurrence trajectories were analyzed to identify pre-ictal dynamical instabilities ($|z| > 3$).

On the Siena cohort, pre-ictal tipping points were detected in the analyzed subjects, corresponding to a detection rate of 88.9%.

The distribution of Lead time, the interval between the last detected tipping point preceding ictal onset and the clinically annotated seizure onset, was centered at the minute scale, with a median of 2.04 minutes and an interquartile range of [0.40, 4.91] minutes.

From a clinically oriented perspective, 70% of subjects exhibited anticipation margins of at least 30 s, 65% at least 1 min, and 50% at least 2 min. These results indicate that, in a substantial proportion of patients, detectable recurrence-based transitions occurred before clinically annotated seizure onset, providing a measurable pre-ictal window.

Relationship between tipping density and supervised performance. To investigate whether unsupervised dynamical instability aligned with supervised classification performance, Spearman correlations was computed between the number of detected pre-ictal tipping points per patient and adjusted balanced accuracy (ABA), balanced accuracy (BA), and PR-AUC.

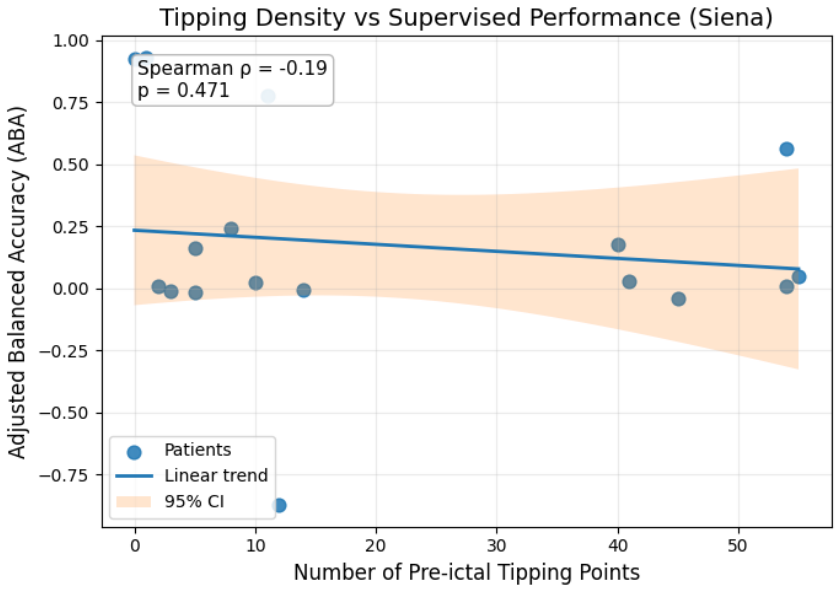


Figure 6.7: Association between pre-ictal tipping density and supervised performance (Siena cohort). Each point represents one patient. The solid line indicates a linear trend with 95% confidence band. No significant monotonic association was observed (Spearman $\rho = -0.19$, $p = 0.47$).

Consistent with the findings observed in the CHB-MIT dataset, no significant association was observed (ABA: $\rho = -0.19$, $p = 0.47$; BA: $\rho = -0.19$, $p = 0.47$; PR-AUC: $\rho = -0.39$, $p = 0.12$). Notably, several of the highest-performing patients in supervised classification exhibited few or no detected tipping points.

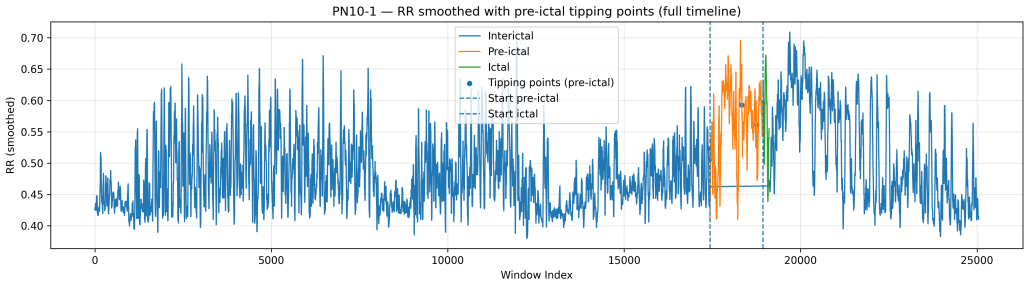
Given the pronounced class imbalance of the task, supervised performance primarily reflects global separability between pre-ictal and interictal distributions, whereas tipping density captures transient dynamical instability. Consistent with the CHB-MIT findings, no strong monotonic association was observed, indicating that tipping density and supervised performance reflect different dimensions of pre-ictal dynamics.

Visual Representation of Tipping Points. To qualitatively illustrate inter-subject variability in pre-ictal dynamical transitions, patients were stratified into three representative categories *weak*, *medium*, and *strong* based on the total number of detected pre-ictal tipping points.

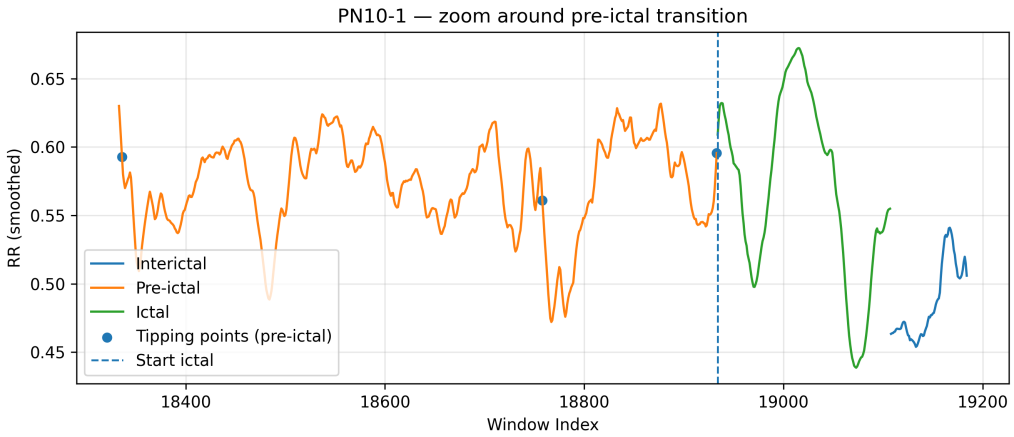
For each patient, the number of tipping events occurring during the annotated pre-ictal phase was computed. Patients were ranked accordingly, and representative

cases were selected from the lower, median, and upper portions of the distribution using quantile-based thresholds.

The following Figures 6.8–6.10 illustrate representative examples from each category:

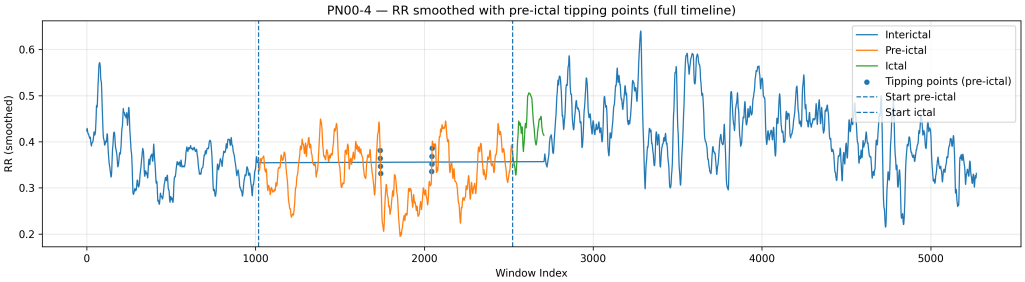


(a) Full RR trajectory across the recording.

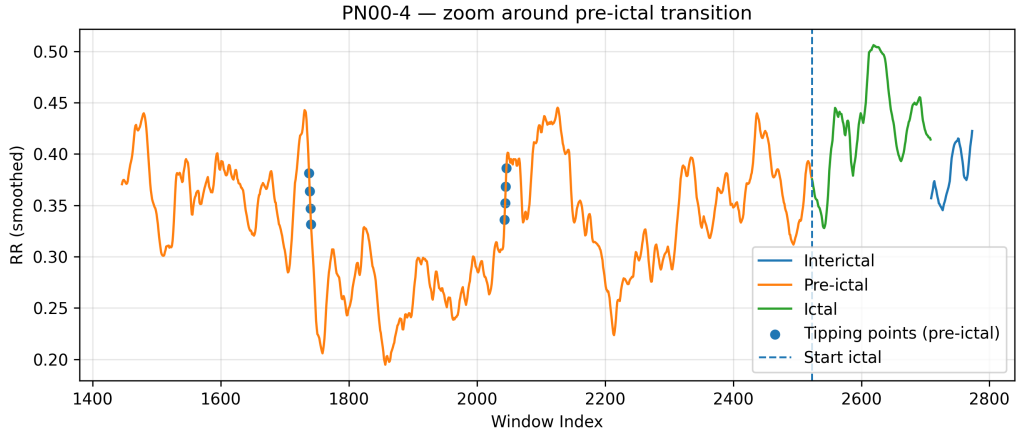


(b) Zoom on the pre-ictal region. Detected tipping points are marked.

Figure 6.8: Representative weak-transition subject (patient ID: PN10-1). Smoothed recurrence rate (RR) trajectory with derivative-based tipping point detection ($|z| > 3$). Only limited pre-ictal dynamical instability is observed.

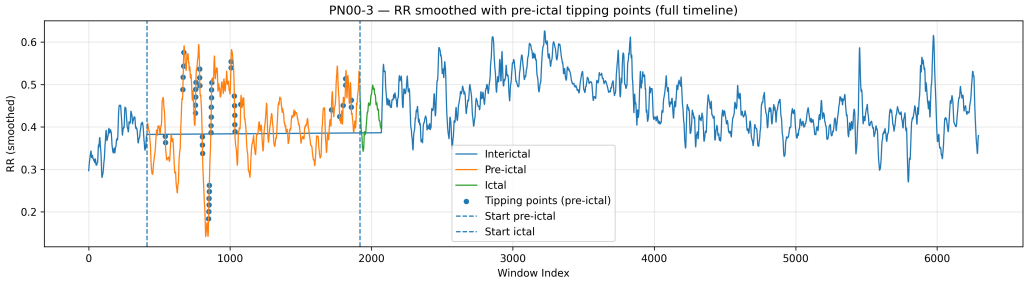


(a) Full RR trajectory across the recording.

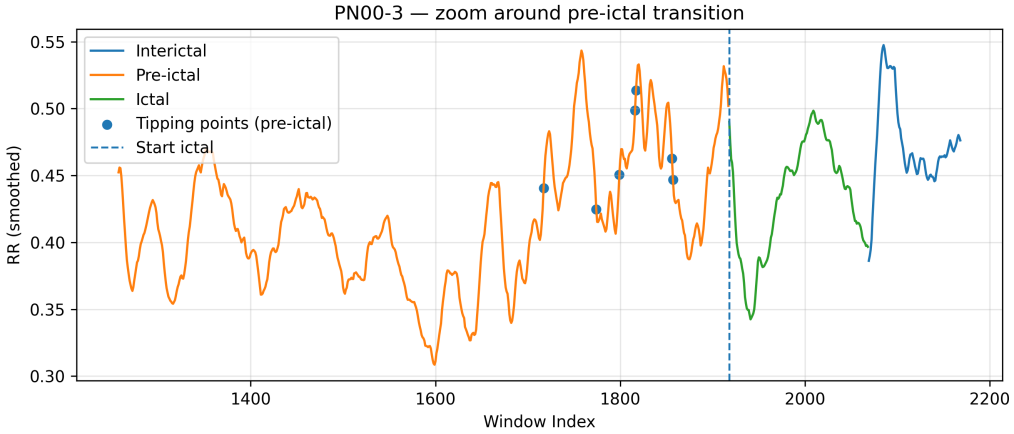


(b) Zoom on the pre-ictal region with visible recurrence fluctuations.

Figure 6.9: Representative intermediate-transition subject (patient ID: PN00-4). A moderate density of pre-ictal tipping events is observed, reflecting intermediate recurrence instability.



(a) Full RR trajectory across the recording.



(b) Zoom on the pre-ictal region showing pronounced recurrence destabilization.

Figure 6.10: Representative strong-transition subject (patient ID: PN00-3). Numerous pre-ictal tipping points are detected, indicating marked recurrence instability preceding seizure onset.

In summary, these examples are representative of the heterogeneity of recurrence-based dynamical transitions. Whereas some patients show strong and easily identifiable pre-ictal destabilization patterns, others show more subtle transitions. This is consistent with known inter-individual variability in seizure dynamics and suggests that recurrence-based instability may constitute a patient-specific dynamical biomarker of pre-ictal state transitions.

Feature Contribution Analysis

To further examine the internal decision mechanisms of the supervised models, SHAP analysis was performed for both LightGBM and CatBoost across datasets. Across independent runs and cohorts, recurrence variability measures consistently emerged as the dominant predictors, in particular RR derivatives, capturing short-

term fluctuations in recurrence rate, were systematically ranked among the most influential features and DET/RR ratios and Trapping Time (TT) variability followed in importance across models. Aggregated SHAP values across patients indicated that features associated with increased recurrence instability contributed positively to pre-ictal classification, whereas more stable recurrence patterns were more frequently associated with interictal states.

Cross-Dataset Comparison

Model	Dataset	Runs	Patients	BA	ABA	PR-AUC
LightGBM	CHB-MIT	10	12	0.544 ± 0.014	0.087 ± 0.028	0.248 ± 0.023
LightGBM	Siena	10	9	0.570 ± 0.009	0.139 ± 0.017	0.253 ± 0.011
CatBoost	CHB-MIT	10	12	0.554 ± 0.006	0.107 ± 0.012	0.256 ± 0.024
CatBoost	Siena	10	9	0.560 ± 0.002	0.119 ± 0.004	0.289 ± 0.009

Table 6.1: RQA multi-patient LOPO performance across datasets. Values are reported as mean ± standard deviation across independent runs.

Across both cohorts, balanced accuracy values remained consistently above chance level for all model families, confirming that recurrence-based descriptors preserve discriminative information across heterogeneous clinical settings.

Quantitatively, the Siena cohort exhibited systematically higher balanced accuracy and adjusted balanced accuracy than CHB-MIT for both LightGBM and CatBoost. Although the magnitude of this improvement was modest, it remained consistent across independent optimization runs. In contrast, CHB-MIT results displayed greater run-to-run variability, particularly for LightGBM, suggesting reduced stability of the optimized configurations under more heterogeneous recording conditions.

PR-AUC values followed a slightly different pattern compared to balanced accuracy. Although Siena showed higher balanced accuracy overall, the increase in PR-AUC was not uniform across models. CatBoost achieved its highest PR-AUC in the Siena cohort, whereas differences between datasets were smaller for LightGBM. This indicates that improvements observed in balanced accuracy do not necessar-

ily translate into proportional gains in PR-AUC, highlighting that the two metrics capture different aspects of performance under class imbalance.

Across both datasets, substantial inter-patient variability was observed. While several subjects exhibited clearly positive ABA values, others yielded negative informedness under the LOPO protocol. This variability remained consistent across independent optimization runs, indicating that model performance heterogeneity reflects subject-level differences rather than instability of the training procedure.

Unsupervised tipping point detection exhibited comparable behavior in both cohorts. Pre-ictal RR transitions were identified in a substantial proportion of patients in each dataset, with associated lead times of similar magnitude. Moreover, both upward and downward ΔRR transitions were observed across subjects in both cohorts, indicating that the heterogeneity of pre-ictal dynamical trajectories is not dataset-specific.

Taken together, the comparable detection rates across cohorts, combined with inter-subject variability in transition timing and tipping density, suggest that recurrence-based instability reflects patient-specific dynamical reorganization processes rather than a uniform deterministic precursor of seizure onset.

To further examine the threshold-independent discriminative behaviour of the two gradient boosting models, Precision-Recall curves were computed from LOPO out-of-fold probabilities using the best-performing configurations. Figure 6.11 illustrates the global PR trade-off on the Siena cohort.

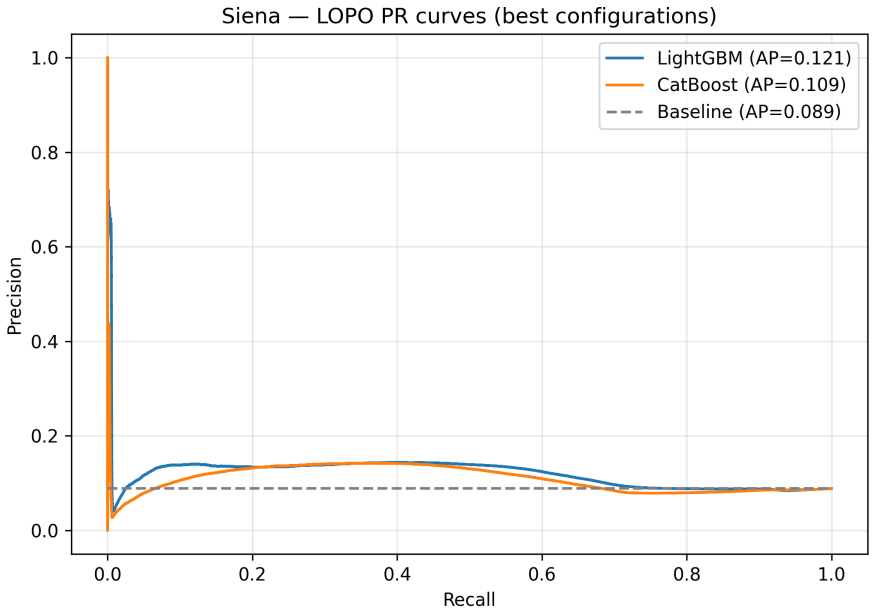


Figure 6.11: Precision-Recall (PR) curves computed from LOPO out-of-fold probabilities using the best-performing configurations for LightGBM and CatBoost on the Siena cohort. The dashed horizontal line represents the positive-class prevalence (random baseline). Average Precision (AP) values are reported in the legend.

As shown in Figure 6.11, both models operate above the random baseline across most recall levels. LightGBM consistently maintains a slightly higher precision for low-to-moderate recall values, resulting in a higher overall Average Precision. However, the overall difference between models remains limited, suggesting that discriminative capacity is largely driven by the recurrence-based feature representation rather than by algorithm-specific effects.

Overall, these findings suggest that recurrence-based features demonstrate stable behavior across independent cohorts, both in supervised classification performance and in the unsupervised detection of pre-ictal dynamical transitions.

6.3.2 Second Methodology - Wavelet-Based Multi-Patient Evaluation (DWT) Modeling

The wavelet-based framework was evaluated under the same multi-patient LOPO protocol adopted for the RQA-based methodology. The CHB-MIT and ScalpEEG datasets were tested using identical patient inclusion criteria and cross-validation

procedures. Performance was computed at the patient level and averaged across 10 independent robustness runs per dataset.

In contrast to the recurrence-based approach, the DWT pipeline relies exclusively on time-frequency features extracted from sliding EEG windows. It does not model nonlinear dynamical recurrence patterns or state-space trajectories. Consequently, evaluation was restricted to supervised classification performance, as wavelet descriptors do not naturally support derivative-based tipping point detection within the same framework.

CHB-MIT Results

Supervised Classification. Using LightGBM with DWT features, the model achieved a mean balanced accuracy of 0.503 ± 0.010, corresponding to an adjusted balanced accuracy (ABA) of 0.007 ± 0.010, and a PR-AUC of 0.100 ± 0.006 across 10 independent runs ($n = 12$ patients).

Balanced accuracy remained consistently close to 0.50, while ABA values were centered around zero with limited variability across optimization studies. Despite extensive hyperparameter tuning, performance metrics converged toward similar values, indicating that the model was unable to extract a stable cross-patient decision boundary from the DWT representation. Occasional increases in recall were typically accompanied by reductions in specificity, resulting in operating points that remained near chance level overall.

To verify that this behavior was not dependent on model architecture, the same feature set was evaluated using CatBoost under identical robustness conditions (10 independent runs; 1,000 Optuna trials per run). CatBoost achieved a mean balanced accuracy of 0.508 ± 0.004, corresponding to an adjusted balanced accuracy (ABA) of 0.016 ± 0.009, and a PR-AUC of 0.101 ± 0.014. The alignment between LightGBM and CatBoost confirms that the near-chance performance reflects limitations of the spectral feature representation rather than model-specific effects.

Siena Results

Supervised Classification. On the Siena cohort ($n = 9$ patients), LightGBM with DWT features achieved a mean balanced accuracy of 0.474 ± 0.029, corresponding to an adjusted balanced accuracy (ABA) of -0.051 ± 0.029, and a PR-AUC of 0.088 ± 0.005 across 10 independent runs.

Balanced accuracy values were consistently below 0.50, and ABA was negative in the majority of runs, indicating performance below chance level. This pattern suggests that the model systematically favored one class under strong imbalance, without establishing a stable separation between pre-ictal and non-pre-ictal states.

CatBoost evaluation under identical robustness conditions produced comparable results. The best-performing run achieved a balanced accuracy of 0.463, corresponding to an adjusted balanced accuracy (ABA) of ABA $= -0.037$ and PR-AUC $= 0.092$ across the same 10 independent runs.

As in the CHB-MIT dataset, performance remained indistinguishable from random classification, indicating that DWT-based spectral descriptors do not provide reliable cross-subject predictive structure under patient-independent validation.

Cross-Dataset Comparison

Model	Dataset	Runs	Patients	BA	ABA	PR-AUC
LightGBM	CHB-MIT	10	12	0.503 ± 0.010	0.007 ± 0.010	0.100 ± 0.006
CatBoost	CHB-MIT	10	12	0.508 ± 0.004	0.016 ± 0.009	0.101 ± 0.014
LightGBM	Siena	10	9	0.474 ± 0.029	-0.051 ± 0.029	0.088 ± 0.005
CatBoost	Siena	10	9	0.463 ± 0.007	-0.37 ± 0.013	0.092 ± 0.007

Table 6.2: DWT multi-patient LOPO performance across datasets. Values are reported as mean ± standard deviation across independent runs.

Across both cohorts and model architectures, DWT-based spectral descriptors yielded performance at or below chance level. The consistency of this pattern across datasets indicates that the limitation is not attributable to architectural choice or hyperparameter configuration, but rather to the absence of stable cross-patient discrimina-

tive structure in the wavelet representation.

While DWT features capture local timefrequency fluctuations, they do not encode reproducible dynamical transition signatures preceding seizure onset. Consequently, no robust population-level decision boundary emerged under strict patient-independent validation.

In contrast, the recurrence-based framework demonstrated consistently above-chance performance in both datasets, highlighting the importance of explicitly modeling large-scale dynamical organization. Overall, DWT-based spectral descriptors appear insufficient as a standalone representation for multi-patient seizure prediction. However, the absence of predictive structure in wavelet features does not imply that spectral information is irrelevant. Rather, it indicates that local timefrequency fluctuations may require integration within a broader dynamical framework to become discriminative at the population level.

This observation naturally motivates the exploration of a hybrid representation in the following section that combines spectral descriptors with recurrence-based dynamical features, aiming to jointly capture local oscillatory modulation and large-scale network instability preceding seizure onset.

6.3.3 Hybrid Integration of Wavelet and Recurrence Features

To assess whether combining recurrence-based descriptors with contextual information improves multi-patient generalization, hybrid feature sets were evaluated under the same LOPO protocol used for the standalone RQA and DWT pipelines. For clarity reason and to keep coherence between all the experiments, for each hybrid configuration, model selection always relied on the already-mentioned independent Optuna studies (1,000 trials each) repeated 10 times, enabling the assessment of both peak performance and run-to-run stability.

CHB-MIT Results

Merged-Context Hybrid Features For the merged-context hybrid representation on the CHB-MIT cohort ($n = 12$ patients), the supervised model achieved a mean LOPO Adjusted Balanced Accuracy of ABA $= 0.1636 \pm 0.0084$ across 10 runs,

corresponding to a balanced accuracy of BA $= 0.5818 \pm 0.0042$. The mean LOPO PR-AUC was 0.3727 ± 0.0020.

The best-performing run reached ABA $= 0.1744$ (i.e., BA $= 0.5872$) and PR-AUC $= 0.3706$ (run-level mean across patients). Overall, the best run improved over the mean ABA by approximately 6.6%, confirming that performance gains were not driven by a single favorable initialization.

Importantly, run-to-run variability across independent runs was limited (run-level standard deviation 0.0084 for ABA and 0.0020 for PR-AUC), supporting the robustness of the tuning and training procedure. In contrast, within-run dispersion across patients remained substantial (typical run-level standard deviation $\approx$ 0.14–0.15 in LOPO ABA), consistent with strong inter-subject heterogeneity under strict cross-patient validation. This pattern suggests that the hybrid representation captures a reproducible population-level discriminative signal. At the same time, subject-specific dynamics remain the main driver of fold-wise performance variability.

Merged-Nearest Hybrid Features The merged-nearest hybrid configuration (71 features) achieved a mean LOPO Adjusted Balanced Accuracy of ABA $= 0.1191 \pm 0.0100$ across 10 independent seeds, corresponding to a balanced accuracy of BA $= 0.5596 \pm 0.0050$. The mean LOPO PR-AUC was 0.3662 ± 0.0043. The best-performing run reached ABA $= 0.1319$ (i.e., BA $= 0.5659$) with PR-AUC $= 0.3648$. Run-to-run variability remained limited, indicating stable optimization behavior consistent with the merged-context configuration.

Compared to the merged-context configuration, performance differences were modest but systematic. The context-enriched hybrid representation yielded slightly higher mean ABA and PR-AUC values, suggesting that incorporating temporal contextual information provides a small but consistent gain in cross-patient separability. However, substantial inter-patient variability persisted under the LOPO protocol, confirming that subject-specific dynamical structure remains the dominant source of performance dispersion under strict cross-patient evaluation.

ScalpEEG Siena Results

Merged-Context Hybrid Features For the merged-context hybrid representation on the Siena ScalpEEG cohort ($n = 9$ patients), the supervised model achieved a mean LOPO Adjusted Balanced Accuracy of ABA $= 0.554 \pm 0.0232$ across independent runs, corresponding to a balanced accuracy of BA $= 0.1084 \pm 0.0464$.

The mean LOPO PR-AUC was 0.430 ± 0.030.

The best-performing seed reached ABA $= 0.570$ (i.e., BA $= 0.570$) and PR-AUC $= 0.451$ (run-level mean across patients). Overall, the best seed improved over the mean ABA by approximately 2.9%, confirming that performance gains were not driven by a single favorable initialization.

Importantly, run-to-run variability across independent runs also in this case, was limited (seed-level standard deviation 0.023 for ABA and 0.030 for PR-AUC), supporting the already mentioned robustness of the tuning and training procedure. In contrast, within-run dispersion across patients remained substantial (typical run-level standard deviation ≈ 0.30–0.35 in LOPO ABA), consistent with strong inter-subject heterogeneity under strict cross-patient validation.

These results confirms that merged-context hybrid configuration captures a reproducible distriminative structure between patients in the Siena cohort, achieving stable above-chance performance under strict LOPO evaluation.

Merged-Nearest Hybrid Features The merged-nearest hybrid configuration achieved a mean LOPO Adjusted Balanced Accuracy of ABA $= 0.3832 \pm 0.0207$ across seven independent seeds, corresponding to a balanced accuracy of BA $= 0.8832 \pm 0.0207$. The mean LOPO PR-AUC was 0.2872 ± 0.0216.

The best-performing seed (seed 105) reached ABA $= 0.4101$ (i.e., BA $= 0.9101$) with PR-AUC $= 0.2912$. Run-to-run variability remained limited, indicating stable optimization behavior and consistent convergence across random initializations.

Even in this cohort, the context-enriched formulation achieved marginally superior metrics, suggesting that embedding explicit temporal information enhances discriminative stability in the patient-independent setting. Nonetheless, cross-patient

heterogeneity remains the dominant factor shaping generalization limits.

Cross-Dataset Comparison - merged approach

To contextualize these results in relation to the standalone pipelines and across datasets, we report a consolidated cross-dataset comparison below.

Hybrid Configuration	Dataset	Runs	Patients	BA	ABA	PR-AUC
Merged-Context Hybrid	CHB-MIT	10	12	0.5818 ± 0.0042	0.1636 ± 0.0084	0.3727 ± 0.0020
Merged-Nearest Hybrid	CHB-MIT	10	12	0.5596 ± 0.0050	0.1191 ± 0.0100	0.3662 ± 0.0043
Merged-Context Hybrid	SIENA	10	9	0.5542 ± 0.0232	0.1084 ± 0.0464	0.430 ± 0.030
Merged-Nearest Hybrid	SIENA	10	9	0.6916 ± 0.0104	0.3832 ± 0.0207	0.2872 ± 0.0216

Table 6.3: Hybrid multi-patient LOPO performance across datasets. Values are reported as mean ± standard deviation across independent Optuna runs (1,000 trials per run).

6.3.4 Additional Event-Based Early Warning Analysis

In order to extend the analysis beyond window-level discrimination and recurrence-based tipping detection, an additional, clinically oriented event-based evaluation was conducted. This perspective shifts the focus from individual windows to seizure-level anticipation, emphasizing practical early warning performance.

Event-Based Early Warning Evaluation: Sensitivity and Time-to-Warning. The evaluation quantifies early warning behavior in terms of (i) *event-level sensitivity*, defined as the proportion of seizures for which at least one sustained alarm is generated prior to onset, and (ii) *time-to-warning* (TTW), defined as the temporal interval between the first sustained alarm and the clinically annotated seizure onset.

Four representations were evaluated under the same LOPO setting, below CHB-MIT Results can be found:

- **DWT_ONLY** wavelet-based spectral features only;
- **RQA_ONLY_NEAREST** recurrence-based descriptors without contextual enrichment;

- **MERGED_NEAREST** hybrid representation combining spectral and recurrence features with nearest aggregation;
- **MERGED_CONTEXT** hybrid representation integrating recurrence and spectral descriptors with contextual temporal features.

Results were aggregated across the same 10 independent seeds used in the supervised experiments. Importantly, this experiment operated on a restricted subset of seizure events, with constant coverage of 5/24 events (coverage = 0.208) across all seeds; therefore, the sensitivity estimates should be interpreted as conditional on the set of evaluable events under the adopted filtering criteria.

Representation	Runs	Events	Coverage	Sensitivity	Median TTW (min)
DWT_ONLY	10	5	0.208	1.00 ± 0.00	7.21 ± 1.62
RQA_ONLY_NEAREST	10	5	0.208	1.00 ± 0.00	7.92 ± 1.49
MERGED_NEAREST	10	5	0.208	0.84 ± 0.08	7.34 ± 0.97
MERGED_CONTEXT	10	5	0.208	0.72 ± 0.10	7.57 ± 0.89

Table 6.4: CHB-MIT experiment event-based early warning performance aggregated across 10 seeds. TTW is reported as median time-to-warning in minutes (mean ± std across seeds). Coverage denotes the fraction of total seizure events included in the experiment under the adopted criteria (5 out of 24).

Across evaluated events, sustained alarms were typically generated several minutes prior to seizure onset, with median TTW values in the ∼7–8 minute range. Event-level sensitivity was maximal for DWT_ONLY and RQA_ONLY_NEAREST within the evaluated set, while the hybrid variants showed lower sensitivity but comparable TTW. These event-level results provide a complementary perspective to window-level classification metrics by focusing on seizure-event anticipation rather than instantaneous window discrimination.

In addition, below the same representation evaluated under the same LOPO setting can be found for Siena ScalpEEG dataset:

Representation	Runs	Events	Coverage	Sensitivity	Median TTW (min)
DWT_ONLY	10	14	0.368	0.86 ± 0.00	9.96 ± 0.00
RQA_ONLY	10	14	0.368	1.00 ± 0.00	9.69 ± 0.14
MERGED_NEAREST	10	14	0.368	0.82 ± 0.25	9.92 ± 0.02
MERGED_CONTEXT	10	14	0.368	1.00 ± 0.00	9.95 ± 0.02

Table 6.5: Siena experiment event-based early warning performance aggregated across 10 seeds. TTW is reported as median time-to-warning in minutes (mean $\pm$ std across seeds). Coverage denotes the fraction of total seizure events included in Exp2 under the adopted criteria (coverage ≈ 0.368).

Notably, median TTW values in Siena were consistently close to the maximum seizure prediction horizon (approximately 10 minutes), indicating that sustained warnings were typically generated near the beginning of the SPH window. This differs from CHB-MIT, where TTW values clustered around 7–8 minutes.

6.4 Comparative Evaluation and Discussion

6.4.1 RQA-Based Framework: Discussion of Results

EEG dynamics vary substantially across patients due to differences in seizure onset zones, large-scale network organization, medication status, and recording conditions. In this context, achieving consistent cross-patient generalization is particularly challenging.

Nevertheless, the recurrence-based framework demonstrated stable above-chance performance across both independent cohorts under the strict Leave-One-Patient-Out (LOPO) validation protocol. In practical terms, above-chance performance means that the model performs better than a random classifier even when evaluated on completely unseen subjects. Given the strong class imbalance inherent to seizure prediction, where pre-ictal windows represent only a small fraction of the data, a naive classifier could achieve deceptively high overall accuracy by always predicting the majority class. For this reason, balanced accuracy and adjusted balanced accuracy were adopted to ensure that both sensitivity and specificity were properly accounted for. The systematic positivity of these metrics indicates that the model

captures meaningful structure distinguishing pre-ictal from non-pre-ictal states.

Moreover, the above-chance performance is maintained under patient-independent evaluation, suggesting that the extracted features capture, at least partially, shared dynamical structure across individuals rather than exclusively patient-specific patterns.

In other words, recurrence-based descriptors appear to encode dynamical properties of generic pre-ictal state and capture dynamical signatures of the pre-ictal transition that remain partially consistent across subjects, despite substantial inter-individual variability.

In CHB-MIT, LightGBM achieved a mean balanced accuracy of approximately 0.54 with limited inter-run variability, while CatBoost yielded slightly higher and more stable values around 0.55. Adjusted balanced accuracy remained systematically positive across runs, indicating that the observed discrimination is not driven by class imbalance artifacts. On the Siena cohort, balanced accuracy was marginally higher overall, whereas PR-AUC differences were more model-dependent.

A key finding is the strong heterogeneity observed across both datasets. While a subset of patients consistently exhibited balanced accuracy values above 0.60 and meaningful PR-AUC levels, others remained close to chance under LOPO validation. Importantly, this finding was largely classifier-independent: switching between LightGBM and CatBoost reduced run-to-run variance but preserved the same patient-level performance structure. This convergence suggests that the limiting factor is not the learning model, but the extent to which separable pre-ictal recurrence signatures are present in each subject.

The comparison between datasets further supports the stability of recurrence-based descriptors under heterogeneous recording conditions. Although the Siena cohort yielded slightly higher balanced accuracy values, the CHB-MIT dataset characterized by longer recordings and greater clinical and technical heterogeneity exhibited comparable above-chance behaviour. This indicates that recurrence-based features preserve discriminative structure across different acquisition settings and patient populations, even when signal characteristics and recording protocols differ substantially.

The unsupervised tipping point analysis provides a complementary and independent perspective on the same underlying dynamics. While detection rates were high in both cohorts, the timing, density, and distribution of detected transitions varied markedly across subjects. Importantly, no clear increasing or decreasing trend was observed between the number of detected tipping points and supervised classification performance. This indicates that local dynamical instability and global statistical separability represent related but distinct aspects of the pre-ictal transition.

At first glance, one might expect that patients exhibiting a larger number of detected dynamical instabilities would also yield higher classification accuracy. However, this assumption implicitly equates local instability with global separability, which are conceptually distinct properties. Supervised classification metrics quantify how well pre-ictal and non-pre-ictal windows can be separated across the entire recording, reflecting the overall statistical structure of the feature distributions. In contrast, tipping density captures temporally localized accelerations in recurrence dynamics that is, transient deviations from the patient's baseline dynamical regime.

A patient may therefore exhibit few but highly consistent pre-ictal shifts that produce clear distribution-level separation, resulting in good classification performance despite low tipping density. Conversely, another patient may show multiple transient instabilities that do not generate sustained separability across windows, yielding limited classification gains despite high tipping counts. This dissociation suggests that global statistical separability and localized dynamical instability reflect distinct but related dimensions of seizure-related transitions.

Jointly, these findings indicate that recurrence-based descriptors encode dynamical information that remains coherent across cohorts, while also being intrinsically modulated by patient-specific properties. The variability observed under LOPO validation therefore appears to be primarily driven by heterogeneity in the underlying EEG dynamics rather than by model instability, hyperparameter selection, or optimization artefacts.

More broadly, the complementarity between supervised separability and unsupervised dynamical instability reinforces the interpretation of epilepsy as a multiscale phenomenon, in which local fluctuations and global reorganization may not

coincide uniformly across patients. This conceptual distinction motivates the integration of feature representations capable of jointly modeling both distribution-level separability and transient dynamical instabilities within a unified predictive framework.

RQA-Based Framework: Biomedical Interpretation From a neurophysiological point of view, recurrence-based features describe how brain connectivity patterns evolve over time, rather than focusing only on changes in spectral power at individual channels. Instead of analysing isolated oscillations, RQA measures track how similar network configurations reappear and how stable or variable these configurations are across consecutive time windows.

The Recurrence Rate (RR) indicates how frequently similar connectivity patterns reoccur over time. Determinism (DET) reflects how structured and temporally organized these patterns are, while Trapping Time (TT) measures how long the system tends to remain in similar states before shifting to a different configuration. Together, these metrics provide an indication of whether the underlying brain dynamics are relatively stable or progressively becoming more irregular.

The feature importance analysis showed that derivatives of RR and measures of variability were among the most relevant predictors. This suggests that the most informative signal is not the absolute value of recurrence itself, but rather how quickly recurrence patterns change. In other words, pre-ictal states are better characterized by increasing instability or faster reorganization of connectivity patterns, rather than by a fixed recurrence level.

Such behaviour is consistent with the idea that seizures emerge from progressive dynamical changes in brain networks. Instead of a sudden isolated event, the system appears to gradually lose stability before transitioning into the ictal state. Recurrence-based descriptors therefore capture this evolving instability, rather than simply detecting static differences between pre-ictal and interictal activity.

Single-patient trajectories qualitatively support this interpretation. In several subjects, the pre-ictal phase is characterized by progressive modulation, increased variability, or subtle drift in recurrence structure, followed by a sharper transition at

seizure onset. Such behaviour is compatible with the view of epilepsy as a transition between distinct attractor states in a nonlinear dynamical system, rather than as a purely stochastic event.

The derivative-based tipping point criterion further operationalizes this perspective. By identifying statistically rare accelerations in the RR trajectory, the method detects moments of dynamical instability without imposing absolute thresholds on recurrence levels. The observation of comparable detection rates across independent cohorts, combined with variability in lead times and tipping density, suggests that recurrence-based instability reflects patient-specific reorganization processes rather than a uniform deterministic precursor of seizure onset. This nuance is essential: tipping points mark phases of increased instability, but do not imply inevitability or fixed temporal proximity to the ictal event.

The strong inter-subject variability observed in both supervised performance and tipping behaviour aligns with the clinical heterogeneity of epilepsy as a pathology. Seizure onset zone, structural connectivity constraints, pharmacological treatment, and recording duration may all influence whether the pre-ictal state manifests as gradual network destabilization or as abrupt local synchronization. In patients where pre-ictal dynamics involve progressive large-scale reorganization, recurrence measures are particularly informative. In others, where transitions are rapid or spatially confined, recurrence-based descriptors may capture only partial aspects of the evolving state.

Importantly, the stability of above-chance performance across independent datasets and classifiers indicates that the recurrence-derived signal is not an artefact of model selection, but reflects a reproducible property of the underlying EEG dynamics when such dynamical restructuring is present. This supports the interpretation of RQA features as physiologically meaningful markers of evolving network stability.

Overall, the RQA framework captures a dynamical dimension of the pre-ictal state that extends beyond local oscillatory alterations. While not universally sufficient for all patients, recurrence-based measures provide a theoretically grounded and empirically supported representation of seizure-related network transitions, reinforcing the view of epileptic seizures as critical transitions in complex brain systems.

6.4.2 DWT-Based Framework: Discussion of Results

The Discrete Wavelet Transform (DWT)-based framework had very limited cross-patient generalization capabilities under the strict LOPO setting, especially when compared to the recurrence-based representation.

On the Siena dataset, LightGBM using DWT features showed consistent adjusted balanced accuracy values close to or slightly below zero. While occasional moderate levels of recall were noted, these were countered by decreases in specificity, leading to balanced accuracy values that were indistinguishable from chance, suggesting that the classifier was mostly operating in a high-sensitivity but low-specificity region, which was indicative of highly unstable decision boundaries under extreme class imbalance.

On the CHB-MIT dataset, results were slightly more encouraging. In most trials, adjusted balanced accuracy values remained slightly positive, with balanced accuracy values hovering around 0.50-0.51. The best-performing models showed small positive ABA values (approximately +0.02 to +0.03), which suggested the presence of weak but identifiable discriminative information in the spectral domain.

One of the most important points to note is the convergence of 10 independent Optuna optimization runs to very similar levels of performance. While extensive hyperparameter search (1,000 trials per experiment) did not yield stable and significant improvements, the convergence of the optimization runs suggests that the observed limitations are very unlikely to be caused by model instability or suboptimal hyperparameter tuning. Instead, they likely represent inherent constraints of the spectral feature representation when assessed in a patient-independent fashion.

Also in this case, the high degree of inter-subject variability supports this view. While some patients showed positive values of informedness, indicating that their pre-ictal transitions are accompanied by measurable spectral modulations, these patterns were not consistent enough across subjects to enable successful LOPO generalization. The failure of DWT features to generalize across heterogeneous patients strengthens the literature belief that spectral changes alone do not provide a stable cross-population biomarker for the pre-ictal state. This limitation can be inter-

preted neurophysiologically. DWT features are computed independently for each EEG channel and primarily quantify local variations in spectral power within predefined frequency bands. As a result, they describe oscillatory activity at a regional level, without explicitly modeling interactions between brain regions or the temporal organization of network states.The model therefore struggles to identify a shared spectral signature of the pre-ictal state.

In this sense, the limitation of the DWT framework does not necessarily imply that spectral changes are absent, but rather that they are insufficiently consistent across individuals to support stable cross-population discrimination when used in isolation.

The DWT-based approach seems to be able to detect local oscillatory changes that might precede seizures in individual patients. However, when assessed in a strict cross-validation fashion across patients, these spectral changes are not consistent enough in terms of their magnitude and quality to enable successful population-level classification.

DWT-Based Framework: Biomedical Interpretation From a neurophysiological perspective, DWT-based features describe how the spectral content of the EEG signal changes over time. In practice, they measure how the energy in different frequency bands (such as low or high frequencies) varies within short time windows. These features are computed separately for each EEG channel and mainly reflect local oscillatory activity. As a result, they capture variations in power at specific electrodes but do not explicitly model how different brain regions interact with each other or how the overall network configuration evolves over time. For this reason, DWT features are well suited to detect localized spectral alterations that may precede seizures in some patients. However, they are less sensitive to global reorganization processes or complex dynamical transitions involving coordinated changes across multiple regions.

The near-chance LOPO performance observed in the Siena cohort, and the only slightly positive results obtained in CHB-MIT, suggest that pre-ictal transitions are not primarily characterized by consistent and reproducible changes in local spectral

power across patients.

In some individuals, positive adjusted balanced accuracy values indicate that spectral alterations may precede seizure onset. For example, certain patients appear to show progressive increases in low-frequency power or specific desynchronization patterns before the ictal event. However, these spectral changes are not sufficiently consistent across subjects to support reliable cross-patient generalization.

The large differences between patients also suggests that the spectral power changes alone cannot adequately model the processes involved in seizure onset. In some patients, the onset of seizure could be a process of larger reorganization of activity patterns rather than mere oscillatory changes at the individual electrodes.

Notably, the large-scale hyperparameter optimization experiments always converged to similar levels of near-chance performance. This indicates that the restriction is not due to improper model optimization but to the limited predictive information in the DWT-based spectral features when used alone.

These observations are consistent with, rather than in contradiction to, the recurrence-based results. While the DWT-based features focus on the question of which oscillatory properties change, the RQA-based features encode how the dynamical regime changes. This conceptual consistency offers a physiologically well-founded rationale for the development of hybrid methods combining spectral and dynamical representations.

6.4.3 Hybrid Framework: Discussion of Results

The hybrid framework combines recurrence-based dynamical descriptors with wavelet-derived spectral features to verify whether complementary representations can advance cross-patient generalization under strict LOPO validation. The configuration of the merged-context set in the CHB-MIT cohort was the one that had the overall best performance of all the other evaluated approaches reaching a mean balanced accuracy of about 0.58 (ABA $\approx$ 0.16). This is quite a distance from the standalone RQA framework (BA $\approx$ 0.54-0.55), which it clearly dominated. The DWT-based representation that still remained almost at the chance level performed worse than the merged-context configuration. The merged-nearest configuration was the one

that improved beyond the recurrence-only baseline at the same time, with balanced accuracy values about 0.56 (ABA ≈ 0.12), but the gain was smaller compared to the context-enriched variant. In the SIENA cohort, the hybrid integration similarly outperformed the standalone features representation. The merged-nearest configuration achieved a mean balanced accuracy of approximately 0.69 (ABA $= 0.38 \pm 0.02$), with stable behavior across independent seeds, reinforcing the assumption that hybrid formulation shows clear discriminative structure under strict patient-independent validation.

Across datasets, the systematic difference between merged-context and merged-nearest configurations indicates that explicit temporal context, when merged, is more effective in promoting cross-patient separability. Performance improvements, however, were not isolated to good optimization runs. They were observed on independent seeds, where the variability between runs was low, while the patients variability was still the overall dominant variability source. This shows that the improvement of the discriminative structure at the population level was the result of hybrid integration and not due to optimization artifacts.

From a modeling perspective, these findings clarify the role of spectral information. When used in isolation, DWT-based features do not generalize reliably across subjects. However, when combined with recurrence-based dynamical descriptors, spectral features appear to provide complementary information that strengthens the global separability of pre-ictal versus non-pre-ictal states.

Overall, the hybrid results demonstrate that integrating dynamical and spectral representations yields the most robust performance under patient-independent validation, while still preserving the intrinsic heterogeneity characteristic of seizure dynamics.

Biomedical interpretation. From a neurophysiological perspective, the hybrid results support the interpretation of the pre-ictal transition as a multi-scale process that cannot be fully characterized by either oscillatory changes or network dynamical reorganization alone. Wavelet-derived features primarily capture local time–frequency modulations at single channels, which may reflect changes in corti-

cal excitability, rhythmic synchronization, or band-limited energy shifts preceding seizures. Recurrence-based descriptors, instead, quantify how the evolving functional network revisits similar connectivity configurations and how stable these configurations remain over time, thus capturing aspects of dynamic organization and large-scale instability.

The improvement observed with hybrid integration suggests that, in at least a subset of patients, pre-ictal evolution involves concurrent (or partially coupled) changes across these two scales: local oscillatory modulation may co-occur with broader alterations in network state recurrence and stability. Under patient-independent validation, neither component is sufficiently consistent across subjects when considered alone (particularly for DWT), but their combination can strengthen the common signal shared across individuals. At the same time, the persistence of strong inter-subject variability in both datasets indicates that the relative contribution of spectral and dynamical mechanisms remains patient-specific, consistent with the clinical heterogeneity of epilepsy and differences in seizure onset zones, propagation pathways, and recording conditions.

Overall, the hybrid framework provides the most robust performance under strict LOPO evaluation, suggesting that integrating dynamical and spectral representations is a promising direction for improving cross-patient seizure prediction, while still reflecting the intrinsically heterogeneous nature of pre-ictal dynamics.

6.4.4 Additional Event-Based Early Warning Perspective.

The additional experiment provides a complementary clinical perspective by shifting the focus from window-level discrimination to event-level anticipation. Importantly, for CHBMIT dataset, event-level sensitivity reached maximal values for both the DWT_ONLY and RQA_ONLY_NEAREST representations within the evaluated subset, while hybrid variants showed slightly reduced sensitivity but comparable anticipation margins.

In the ScalpEEG dataset, a higher event-level sensitivity was observed for recurrence-based representations. Both RQA_ONLY and MERGED_CONTEXT achieved perfect event-level sensitivity within the evaluable subset, whereas DWT_ONLY

and MERGED_NEAREST showed reduced detection rates.

Across the subset of considered events, in CHBMIT, alarms were typically generated several minutes before seizure onset, with median time-to-warning (TTW) values in the 7–8 minute range. Instead, TTW values were consistently close to the maximum horizon for seizure prediction (approximately 10 minutes) in Scalp EEG Siena, indicating that sustained alarms were typically generated near the beginning of the SPH window.

This apparent dissociation between window-level separability and event-level sensitivity is noteworthy. While the DWT-based framework exhibited near-chance balanced accuracy under LOPO validation, it nevertheless generated early alarms for all evaluated seizure events in this context. This suggests that event-level anticipation does not necessarily require strong global separability across all windows, but may instead rely on localized probability surges that cross alarm thresholds in proximity to seizure onset.

In contrast, the hybrid configurations achieved superior statistical discrimination at the window level, but displayed slightly lower event-level sensitivity under the adopted alarm criteria. This indicates that optimizing global separability does not automatically maximize seizure-level detection under fixed alarm rules.

These different findings between the two datasets suggest that event-level anticipation dynamics can be dataset-dependent. While hybrid contextual integration did not maximize sensitivity in CHB-MIT, it proved highly effective in the Siena cohort, highlighting that the benefit of temporal contextualization may depend on cohort characteristics and recording modality. In addition, the results reinforce even more the multi-dimensional nature of seizure prediction. Statistical separability, dynamical instability, and event-level anticipation represent related but non-identical dimensions of pre-ictal dynamics. A representation may perform moderately in one dimension while excelling in another, depending on how information is aggregated and thresholds are applied.

Finally, the restricted coverage of the experiment (in CHBMIT only 5 out of 24 seizure events were evaluable, corresponding to a coverage of 0.208 while in the ScalpEEG dataset, coverage was higher but still limited, at approximately 0.368)

further suggests that early warning performance must be interpreted conditionally, as it depends on the subset of seizures meeting the adopted filtering criteria. They do not represent performance across the full set of recorded events. Future extensions should therefore evaluate event-level behavior on a broader seizure set and integrate false-alarm analysis to fully characterize clinical utility.

6.4.5 Comparative Discussion of 3 methods

Supervised Classification Perspective

A comparative assessment of the three methodological branches RQA, DWT, and their combined form reveals a consistent performance gradient under strict LOPO validation.

Evaluating the DWT-based framework singularly, it operates at or near chance level across datasets, indicating that in changes spectral power alone do not provide sufficiently reliable cues for good generalization across different patients. Although small positive deviations were occasionally observed, these effects were limited in magnitude and inconsistent across subjects and through optimization runs. This suggests that the oscillatory modulations that precede seizures are highly patient-specific and do not translate into a shared spectral signature at the population level.

In contrast, the recurrence-based framework was able to achieve stable better than chance performance (BA $\approx$ 0.54–0.56), demonstrating that dynamical recurrence descriptors capture partially shared structure in pre-ictal transitions across subjects. While inter-patient variability remained substantial, the consistently positive adjusted balanced accuracy across independent runs indicates the presence of reproducible, real and useful discriminative information beyond random fluctuations or imbalance artifacts. These findings suggest that alterations in dynamical stability and recurrence structure constitute a more coherent cross-subject signal than spectral power changes alone.

The hybrid framework showed definitely the strongest overall performance, systematically outperforming spectral features used in isolation while retaining the structured contribution of recurrence-based descriptors. On the CHB-MIT dataset,

the merged-context configuration achieved balanced accuracy values approaching 0.58 (ABA ≈ 0.16), clearly surpassing both stand-alone representations. The merged-nearest variant also improved upon the recurrence-only baseline, although the margin was comparatively smaller. The hybrid formulation worked well also in the ScalpEEG cohort, demonstrating robust discriminative capacity. The merged-nearest configuration attained a mean balanced accuracy of approximately 0.69, with consistent results across different random seeds and an ABA of ≈ 0.38.

The consistent advantage and better metrics observed for the context-enriched variant suggests that directly taking into account temporal context improves cross-patient separability, potentially by strengthening sustained pre-ictal dynamics and reducing the effect of random fluctuations.

Importantly, these improvements did not eliminate inter-subject heterogeneity. Even under hybrid integration, patient-level variability remained pronounced, confirming that seizure generation mechanisms are intrinsically modulated by subject-specific properties. Nevertheless, the progressive improvement observed from DWT to RQA to Hybrid reflects increasing representational richness: spectral descriptors alone are insufficient; recurrence-based dynamical features provide stable cross-patient structure; and their integration yields the most consistent population-level discrimination under strict patient-independent evaluation.

Unsupervised and Dynamical Perspective

The unsupervised analysis further clarifies the distinction between the two approaches.

The RQA framework incorporates an explicit dynamical instability criterion through derivative-based tipping point detection. Across both cohorts, a substantial proportion of patients exhibited statistically rare accelerations in recurrence dynamics preceding seizure onset. Although timing and details of these changes varied across subjects, the presence of a consistent pattern of dynamical destabilization was consistent at the cohort level.

On the other hand, the DWT pipeline does not include an explicit dynamical instability mechanism thta allows to identify the moment in which a system becomes

unstable. The spectral features it uses can tell us what's happening at a specific time and frequency, but they don't directly show when a system is about to shift suddenly or reorganize itself. As a result, DWT-based models have to rely entirely on statistical separability of spectral windows, without access to explicit indicators of dynamical shifts or transitions.

As observed in both cohorts, tipping density and supervised separability quantify related but distinct dimensions of the pre-ictal transition. The supervised metrics reflect a sense of how distributed things are on a global level, across the whole recording, whereas tipping events capture where the system is unstable in a specific, localized way, and how this dynamical transition happens over time.

Their dissociation across patients reinforces the multi-dimensional nature of transitions in seizure-related context.

Biomedical Interpretation

From a neurophysiological perspective, the comparative behavior of the DWT-based, RQA-based, and hybrid frameworks reflects complementary but not equivalent ways of characterizing the pre-ictal transition.

The DWT-based framework captures the timefrequency structure of the EEG signal at the level of individual recording channels. In biomedical terms, it quantifies how the amplitude (power) of specific frequency bands such as delta, theta, alpha, or beta rhythms changes over time within localized cortical regions. When increases in band power are registered, it typically reflects enhanced synchronization of neuronal populations oscillating at that frequency, whereas decreases may indicate desynchronization or reduced coherence of local circuits. By looking at these shifts, the framework can monitor where certain patterns of brain activity are getting stronger or weaker prior to seizure onset.

However, as already discussed in statistical terms, the current findings suggest that these spectral changes are not reliably replicable across patients in a patient-independent manner. While some patients show identifiable pre-ictal spectral changes in oscillatory power, these are highly variable in terms of magnitude, spatial extent, and frequency content. Thus, no reproducible population-level spec-

tral signature characterizes the pre-ictal state under patient-independent evaluation.

In contrast, the RQA-based framework describes how patterns of functional connectivity evolve over time. Despite strong inter-patient variability, it achieved stable above-chance performance across independent cohorts, indicating that recurrence-based descriptors capture dynamical properties that are more reproducible across subjects. At the signal level, the most informative predictors were not the absolute values of recurrence measures, but instead how they variated in the short-term. Rather than converging toward a fixed abnormal configuration, the network seems to show increased dynamical volatility prior to seizure onset: the connectivity matrices become progressively less similar across adjacent windows, the recurrence structure fluctuates more rapidly, and specific configurations do not the persist as long. This behavior is compatible with a situation in which the brain moves from a metastable interictal regime toward a pathological attractor state (Jirsa et al. 2014; Breakspear 2017). Increased short-term variability in recurrence structure may reflect weakening of inhibitory control, altered excitationinhibition balance, and reduced resilience of large-scale functional networks (Meisel, Kuehn, et al. 2012; Kramer, Kirsch, and Szeri 2012). However, such destabilization does not necessarily occur in the form of a uniform spectral change. Rather, it could appear in the form of higher sensitivity to perturbations, temporary reconfiguration of connectivity patterns, and lower persistence of pre-existing stable network configurations. In this context, the pre-ictal state can be viewed as a phase of lower dynamical stability, in which the system exhibits increased dynamical fluctuations before crossing a critical threshold into the ictal state.

Importantly, the hybrid results refine and extend this physiological interpretation. While spectral power changes alone do not generalize reliably across patients, they are not meaningless either. The improved performance of the merged-context configuration suggests that pre-ictal transitions involve coordinated changes across multiple spatial and temporal scales: local rhythmic activity evolves in parallel with large-scale alterations in connectivity stability.

In this view, spectral modulations may reflect the local electrophysiological manifestation of an underlying dynamical destabilization process. The hybrid represen-

tation captures both the micro-scale oscillatory adjustments and the macro-scale shifts in network recurrence structure. By integrating these dimensions, it provides a more complete description of the transition from a metastable interictal regime toward the ictal attractor.

Thus, pre-ictal dynamics appear neither purely oscillatory nor purely dynamical. Rather, they reflect multi-scale reorganization processes in which local rhythmic modulation and global connectivity instability interact (Lopes et al. 2021). The hybrid framework better captures this interaction, yielding stronger cross-patient separability while preserving physiological interpretability.

The unsupervised tipping point analysis further supports this view. High detection rates were observed in both cohorts, yet no strong monotonic association emerged between tipping density and supervised classification performance. This dissociation indicates that localized dynamical instabilities and global statistical separability represent distinct dimensions of the same transition process. In some patients, few but consistent dynamical shifts generate clear distribution-level separation; in others, multiple transient instabilities do not translate into stable discriminative structure.

Looking at the results together, it seems that seizure generation, in this experimental setting, is more consistently associated with alterations in the stability and organization of large-scale connectivity dynamics than with fixed spectral power changes. The DWT results reinforce this interpretation by demonstrating that oscillatory alterations alone are insufficient for reliable cross-patient prediction, whereas recurrence-based measures sensitive to temporal reorganization retain discriminative value across heterogeneous cohorts.

Overall, the convergence of supervised performance, hybrid integration gains, and unsupervised tipping dynamics supports the view of epilepsy not merely as a disorder of abnormal rhythms, but as a condition involving progressive destabilization of large-scale brain network dynamics prior to seizure onset.

Synthesis and Implications

Overall, the comparative analysis under strict patient-independent validation reveals a consistent performance hierarchy: DWT < RQA < Hybrid. Spectral descriptors alone do not provide sufficiently stable cross-patient cues, whereas recurrence-based features capture reproducible aspects of pre-ictal dynamical reorganization. Their integration within the hybrid framework yields the strongest and most consistent population-level discrimination.

The findings show that using recurrence-based descriptors gives a solid base for understanding what happens before a seizure, and spectral information becomes more useful when it's part of a bigger picture that includes how things change over time and how stable the network is. This progressive improvement across methods reflects increasing representational richness and a more complete picture, rather than differences in validation protocol or model selection.

At the same time, it can be observed that patient's brains work really differently when it comes to seizures. This means that no single approach can guarantee a perfect seizure prediction for every subject, which suggest that each patient's brain has its own unique mix of physical, functional, and physiological characteristics that affect how seizures start.

Cumulatively, supervised classification results, the reproducible improvement observed under hybrid integration, and the dynamical instability analysis point toward a multi-scale process underlying seizure onset (Lopes et al. 2021; Scheffer et al. 2009).

The results of this study show that future models of seizure prediction should focus on integrated representations that can simultaneously capture the spectral and dynamical information. Models that take into account the multi-scale organization and individual brain logics variability may provide the most promising option for future improvement in cross-patient generalization.

Chapter 7

Conclusions and Future Work

This thesis systematically investigated the prediction of epileptic seizures from scalp EEG signals, analyzing the ability of different signal representations to model the pre-ictal transition within a rigorously patient-independent framework. A Leave-One-Patient-Out (LOPO) protocol was adopted on two independent datasets to ensure rigorous cross-patient evaluation. The primary goal was not to maximize performance on individual patients, but to identify which EEG signal properties were sufficiently stable and shared to allow generalization across subjects.

Three methodological families were evaluated:

(i) a dynamical representation based on Recurrence Quantification Analysis (RQA),

(ii) a spectral representation based on the Discrete Wavelet Transform (DWT),

(iii) a hybrid configuration integrating recurrence-based descriptors with time–frequency features.

All approaches were implemented within the same modeling framework, with nested hyperparameter optimization, internal validation-based decision threshold selection, and repeated experiments across multiple random seeds. This ensured direct comparability between methods and strict control over potential sources of variability.

Summary of Results

The comparative analysis under the LOPO validation framework revealed a consistent performance hierarchy:

$$\text{DWT} < \text{RQA} < \text{Hybrid}$$

The hybrid configuration not only showed the highest score in each of the two datasets but also was declared the most solid solution in the patient-independent scenario. In particular, the merged-context variant in the CHB-MIT cohort achieved almost 0.58 balanced accuracy values, which is better than both the standalone recurrence-based and purely spectral representations. Moreover, the performance gains remained consistent across different independent random seeds, indicating that the improvement is structural rather than incidental. Based on these observations, it can be inferred that the fusion of dynamic recurrence descriptors and temporal-frequency spectral data provides complementary information about the pre-ictal transition state. In the course of the pre-ictal state, the separate representations portray only partial views of the signal, whereas together they yield a more comprehensive and generalizable explanation of the underlying dynamics, thus, making it easier to differentiate between patients.

Pathophysiological Interpretation

From a neurophysiological perspective, the results suggest that the pre-ictal phase is not characterized by a universal spectral signature, but rather by a progressive destabilization of network dynamics.

The most informative RQA features were not absolute recurrence values, but their short-term variations. This indicates that the pre-ictal state is better described as an increase in instability and rapid reorganization of connectivity patterns, rather than as a stable increase or decrease in a single index.

In contrast, DWT features capture local modifications of oscillatory activity. In some patients, such spectral changes precede seizure onset, but they do not emerge

as reproducible population-level biomarkers under LOPO validation, reinforcing the previous literature assumption suggesting that spectral descriptors alone are insufficient to capture the full complexity of pre-ictal transitions, particularly when strict cross-subject generalization is enforced. Spectral alterations appear strongly subject-specific in terms of frequency band involvement, magnitude, and spatial distribution.

The superiority of the hybrid configuration suggests that the pre-ictal transition is a multi-scale phenomenon: local oscillatory changes are intertwined with broader reorganization of connectivity dynamics. Integrating these two dimensions provides a more complete description of the destabilization process preceding seizure onset and offering a better generalization across individuals.

In the present work, the identified early warning signals primarily emerged as:

- increased short-term variability of the Recurrence Rate (RR);
- elevated temporal derivatives of recurrence indices, indicating faster reorganization of network dynamics;
- reduced stability and persistence of recurrent patterns;
- statistically anomalous accelerations in dynamic trajectories (tipping points), detected through a criterion based on the standardized derivative of RR.

These signals reflect a reduction in system resilience. From a neurophysiological standpoint, this does not necessarily correspond to overt clinical manifestations or macroscopically evident EEG alterations. Rather, it suggests that the dynamic balance between excitation and inhibition becomes progressively more fragile making the functional network increasingly sensitive to internal fluctuations.

From a theoretical perspective, the increase in variability is aligned with the phenomenon of increased variance observed near critical transitions in complex systems. Dynamic accelerations and reduced trajectory stability can be interpreted as manifestations of a loss of resilience, conceptually related to critical slowing, in which the system struggles to quickly return to equilibrium after small perturbations.

Although autocorrelation and recovery rate were not directly measured, the observed variations in recurrence descriptors indicate that the system tends to deviate more easily from the typical interictal regime and remain in alternative dynamic configurations for longer. This pattern points to a reduction in stability of the interictal state: in other words, the brain appears less capable of maintaining its stable functional configuration over time, exhibiting larger and less contained fluctuations compared to the interictal phase.

Quantitatively, these signals correspond to a reduction in the temporal stability of recurrence trajectories. Recurrence Rate and other RQA indices show greater variability and higher derivatives, indicating faster changes between successive functional configurations. This implies that the interictal state is not maintained with the same temporal coherence observed in stable phases, but is more likely to deviate towards alternative dynamic configurations associated with the seizure.

In this sense, the identified early warning signals do not correspond to an early clinical symptom directly observable in routine EEG inspection. Rather, they represent quantitative indicators of functional instability, reflecting a process of progressive destabilization that, in a subset of patients, precedes the emergence of clinically manifest seizures.

Heterogeneity and Limitations

This heterogeneity does not appear to arise from model instability or optimization problems, but more likely reflects intrinsic differences in seizure generation mechanisms, epileptic focus localization, and the structural and functional properties of individual brain networks.

In addition, the proposed binary formulation of the problem (pre-ictal vs. non pre-ictal) simplifies what is likely a continuous and multi-stage dynamical process.

A natural extension of the present work could comprehend models that explicitly incorporate the temporal progression of the transition as outlined in the following section.

Potential Integration

Although not part of the experimental pipeline implemented in this thesis, one of the most promising directions, in light of the theoretical foundations discussed, would be to apply recurrence analysis not only to raw EEG signals but also to the trajectories of dynamic functional connectivity matrices.

Recurrence analysis of time-varying network topologies has been proposed as a tool to assess global reorganization and stability in complex brain networks (Lopes et al. 2021). This approach is conceptually related to the theory of Dynamical Network Biomarkers (DNB), introduced by (Luonan Chen et al. 2012), which defines critical transitions as synchronized increases in variability and intra-network correlation within specific subsystems.

While recurrence quantification of dynamic networks and DNB arise from different methodological traditions, they share a common objective: detecting collective instabilities that emerge from interactions among network components. Integrating these measures and indicators could therefore provide a more explicit and structurally grounded representation of the collective instability that precedes seizure onset.

More recently, entropy-based early warning frameworks, such as the Network Information Entropy Early-warning (NIEE) proposed by (X. Zhang, R. Liu, and Lihua Chen 2019), have suggested that regime shifts may also be preceded by changes in network entropy, interpretable as a progressive loss of structural stability and coordination among interacting components.

Integrating entropy metrics into a connectivity- or recurrence-based framework could add an information-theoretic perspective to the characterization of pre-ictal transitions.

Together, these approaches outline a conceptual framework for multiscale modeling of early warning signals in epilepsy. Signal-level nonlinear dynamics (RQA), mesoscopic network reorganization, correlation-induced instability (DNB), and entropy-based structural uncertainty (NIEE) represent complementary perspectives. Their integration into a unified predictive framework could not only improve patient-

independent generalization but also help elucidate the dynamic mechanisms underlying seizure onset as a critical transition in a complex adaptive system.

General Conclusion

These findings suggest that recurrence-based instability indices reflect genuine predictive structure rather than dataset-specific artefacts, as their discriminative signal persists under strict cross-patient validation. While not universally sufficient, these variables offer a theoretically grounded and physiologically consistent description of the network transitions associated with epileptic seizures.

The findings are consistent with a growing body of literature that interprets epilepsy as a disorder of dynamical systems, characterized by progressive loss of stability and connectivity reorganization prior to the ictal event.

The observed progression from purely spectral representations to dynamic descriptors and finally to multiscale integration suggests that effective modeling of the pre-ictal phase requires tools capable of simultaneously capturing local instability and global reorganization.

In this sense, the primary contribution of this thesis lies not only in the quantitative performance gains, but in showing that network-level dynamical instability can be formalized as an interpretable early warning mechanism, capable of anticipating seizure onset under strict patient-independent validation.

This evidence strengthens the view of epilepsy as a critical transition in a complex dynamic system and provides a methodological basis for future developments towards multiscale, physiologically based predictive models.

Bibliography

Ambika, G. and Jürgen Kurths (2021). "Tipping in complex systems: theory, methods and applications". In: *The European Physical Journal Special Topics* 230.16-17, pp. 3177–3179. DOI: 10.1140/epjs/s11734-021-00269-2.

Beggs, John M and Dietmar Plenz (2003). "Neuronal avalanches in neocortical circuits". In: *Journal of Neuroscience* 23.35, pp. 11167–11177.

Breakspear, Michael (2017). "Dynamic models of large-scale brain activity". In: *Nature Neuroscience* 20.3, pp. 340–352. DOI: 10.1038/nn.4497.

CHB-MIT Scalp EEG Database (2025). Version 1.0.0. Boston Children's Hospital; pediatric scalp EEG, 22 subjects, 23 cases. PhysioNet. URL: https://physionet.org/content/chbmit/1.0.0/ (visited on 09/05/2025).

Chen, Luonan et al. (2012). "Detecting early-warning signals for sudden deterioration of complex diseases by dynamical network biomarkers". In: *Scientific Reports* 2, p. 342.

Cocchi, Luca et al. (2017). "Criticality in the brain: A synthesis of neurobiology, models and cognition". In: *Progress in Neurobiology* 158, pp. 132–152.

Dakos, Vasilis et al. (2012). "Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data". In: *PLoS ONE* 7.7, e41010. DOI: 10.1371/journal.pone.0041010.

Detti, Paolo, Giovanni Vatti, and Gabriele Zabalo Manrique de Lara (2021). *EEG Dataset for Epileptic Seizure Detection from Scalp EEG Recordings*. https://physionet.org/content/siena-scalp-eeg/1.0.0/. PhysioNet. DOI: 10.13026/8m5v-9828.

Goldberger, Ary L et al. (2000). "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals". In: *Circulation* 101.23, e215–e220.

Jirsa, Viktor K. et al. (2014). "On the nature of seizure dynamics". In: *Nature Communications* 5, pp. 1–10. DOI: 10.1038/ncomms3512.

Kramer, Mark A., Heidi E. Kirsch, and Andrew J. Szeri (2012). "Pathological patterns of neuronal synchrony in epilepsy". In: *The Journal of Neuroscience* 32.46, pp. 16334–16345. DOI: 10.1523/JNEUROSCI.2907-12.2012.

Le Van Quyen, Michel et al. (2001). "Anticipation of epileptic seizures from standard EEG recordings". In: *The Lancet* 357.9251, pp. 183–188. DOI: 10.1016/S0140-6736(00)03591-1.

Liu, Shun, Fali Li, and Feng Wan (2023). "Distance to criticality undergoes critical transition before epileptic seizure attacks". In: *Brain Research Bulletin* 200, p. 110684.

Loh, KJ and T Smith (2022). "A robust LOPO validation framework for cross-subject EEG analysis". In: *Biomedical Signal Processing and Control* 75, p. 103579.

Lopes, Marinho A. et al. (2021). "Recurrence quantification analysis of dynamic brain networks". In: *European Journal of Neuroscience* 53.4, pp. 1040–1059. DOI: 10.1111/ejn.14960.

Malekzadeh, Anis et al. (2021). "Epileptic Seizures Detection in EEG Signals Using Fusion of Handcrafted and Deep Learning Features". In: *Sensors* 21.22, p. 7710. DOI: 10.3390/s21227710. URL: https://www.mdpi.com/1424-8220/21/22/7710.

Maturana, Matias I. et al. (2020). "Critical slowing down as a biomarker for seizure susceptibility". In: *Nature Communications* 11.1, p. 2172. DOI: 10.1038/s41467-020-15908-3.

Meisel, Christian, Christian Kuehn, et al. (2012). "Critical slowing down governs the transition to neuron spiking". In: *PLoS Computational Biology* 8.2, e1002338. DOI: 10.1371/journal.pcbi.1002338.

Meisel, Christian, Alexander Storch, et al. (Jan. 2012). "Failure of Adaptive Self-Organized Criticality during Epileptic Seizure Attacks". In: *PLOS Computational*

Biology 8.1, pp. 1–8. DOI: 10.1371/journal.pcbi.1002312. URL: https://doi.org/10.1371/journal.pcbi.1002312.

Merlini, Simona et al. (2025). "Multisystem failure, tipping points, and risk of Alzheimer's disease". In: *Alzheimers & Dementia* 21.1, e70249. DOI: 10.1002/alz.70249.

Mursalin, Md et al. (2017). "Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier". In: *Neurocomputing* 241, pp. 204–214. DOI: 10.1016/j.neucom.2017.02.053.

Ngamga, E. J. et al. (2017). "Evaluation of recurrence quantification measures for the analysis of epileptic EEG". In: *Physica A: Statistical Mechanics and its Applications* 482, pp. 300–310. DOI: 10.1016/j.physa.2017.04.043.

Ouyang, Gaoxiang et al. (2008). "Dynamic characteristics of absence EEG recordings with multivariate empirical mode decomposition". In: *Epilepsy Research* 82.2-3, pp. 152–160. DOI: 10.1016/j.eplepsyres.2008.08.014.

Proverbio, Daniele, Alexander Skupin, and Jorge Gonçalves (2023). "Systematic analysis and optimization of early warning signals for critical transitions using distribution data". In: *iScience* 26.7, p. 107156. DOI: 10.1016/j.isci.2023.107156.

Pyrialakos, Stefanos, Ioannis Kalogeris, and Vissarion Papadopoulos (2023). "Multiscale analysis of nonlinear systems using a hierarchy of deep neural networks". In: *International Journal of Solids and Structures* 271-272, p. 112261. DOI: 10.1016/j.ijsolstr.2023.112261.

Scheffer, Marten et al. (2009). "Early-warning signals for critical transitions". In: *Nature* 461.7260, pp. 53–59. DOI: 10.1038/nature08227.

Sharma, Manish, Ram Bilas Pachori, and U. Rajendra Acharya (2017). "A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension". In: *Pattern Recognition Letters* 94, pp. 172–179. ISSN: 0167-8655. DOI: 10.1016/j.patrec.2017.03.023.

Sharmila, A and P Geethanjali (2017). "Automatic detection of epileptic seizures using wavelet transform and SVM". In: *Computer Methods and Programs in Biomedicine* 140, pp. 247–253.

Shoeb, Ali H. (2009). "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment". PhD thesis. Massachusetts Institute of Technology.

Siena Scalp EEG Database (2025). Version 1.0.0. Unit of Neurology and Neurophysiology, University of Siena. PhysioNet. URL: https://physionet.org/content/siena-scalp-eeg/1.0.0/ (visited on 09/04/2025).

Siuly, Siuly, Yan Li, and Yanchun Zhang (2016). "A novel hybrid classifier based on principal component analysis and least square support vector machine for EEG signal classification". In: *Computer Methods and Programs in Biomedicine* 124, pp. 34–46. DOI: 10.1016/j.cmpb.2015.10.004.

Subasi, Abdulhamit (2013). "Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders". In: *Computers in Biology and Medicine* 43.5, pp. 576–586. ISSN: 0010-4825. DOI: 10.1016/j.compbiomed.2013.01.020.

Tran, Duc T., Thanh T. Nguyen, and Hung T. Nguyen (2022). "Epileptic seizure detection using discrete wavelet transform and binary particle swarm optimization". In: *Diagnostics* 12.3, p. 612. DOI: 10.3390/diagnostics12030612.

Trulla, L. L. et al. (1996). "Recurrence quantification analysis of the logistic equation with transients". In: *Physics Letters A*.

Tzallas, Andreas T., Markos G. Tsipouras, and Dimitrios I. Fotiadis (2009). "Epileptic seizure detection in EEGs using time–frequency analysis". In: *IEEE Transactions on Information Technology in Biomedicine* 13.5, pp. 703–710. DOI: 10.1109/TITB.2009.2017939.

Webber Jr, Charles L and Joseph P Zbilut (2015). *Recurrence Quantification Analysis: Theory and Best Practices*. Springer.

Wu, Yitong et al. (2025). "A Review of Machine Learning and Deep Learning Trends in EEG-Based Epileptic Seizure Prediction". In: *IEEE Access* 13. Open access, pp. 159812–159841. DOI: 10.1109/ACCESS.2025.3606966. URL: https://doi.org/10.1109/ACCESS.2025.3606966.

Zhang, Xiaojie, Rui Liu, and Lihua Chen (2019). "Network information entropy reveals early-warning signals of critical transitions". In: *Entropy* 21.6, p. 587. DOI: 10.3390/e21060587.