Epilepsy Detection from Multi-Channel EEG Using Cross-Recurrence Quantification Analysis and Machine Learning

Nikolaos Mouzakitis

November 2, 2025

Introduction

Epilepsy is a neurological disorder, which generates recurrent, unprovoked seizures. Seizures are the result of abnormal neuronal brain activity which subsequently lead into disturbances in the behavior, sensation, the consciousness and the movement of the affected subject. Information gathered by World Health Organization (WHO), report that epilepsy affects around 50 million people worldwide, placing it as one of the most common neurological conditions globally.

Epilepsy has its origin in diverse and various causes depending on many factors. The generation of epileptic activity can be an impact of identifiable structural, genetic, infectious, metabolic, or immune-related abnormalities, while for almost 50% of the cases, their trigger still remains unknown[53]. Depending on the brain regions involved, a seizure can have several classifications. It can be classified as focal (if its origin is a specific area) or a generalized one (involving both hemispheres). Clinically, seizures are characterized by high variability, from short term lapses in awareness, up to convulsive episodes, and the fact of their unpredictable occurrence has severe impact in patient's quality of life.

For the diagnosis and monitoring of epilepsy, electroencephalography (EEG) is widely utilized, as is offers a non invasive technique for recording brain's electrical activity using surface electrodes. EEG signals contain temporal information that reflects the dynamic interactions of neuronal populations. During the seizure events, the presence of characteristic patterns such as spikes, sharp waves, or rhythmic discharges often appear, distinguishing this kind of activity from the normal background rhythms. As a result, EEG analysis holds a central role both in clinical diagnosis and for research related to this application domain.

Recent advances in signal processing combined with machine learning have greatly improved the ability EEG data analysis. Techniques like time-frequency decomposition, nonlinear dynamics, recurrence analysis, and deep neural networks can offer new solutions for automated extraction of complex spatial and temporal features from EEG recordings. The ultimate goal of these approaches is to support clinicians by providing objective, data-driven tools for seizure detection, prediction, and classification, in order to contribute to better patient care and personal treatments.

In summary, epilepsy represents a major public health challenge due to its prevalence, variability, and social consequences. Understanding the electrophysiological mechanisms that drive seizure generation and developing reliable methods for automatic EEG analysis remains a crucial research direction in modern neuroscience and biomedical engineering.

Recurrence Quantification Analysis (RQA) and Cross-Recurrence Quantification Analysis (CRQA) are nonlinear methods for the analysis of nonstationary time series, such as EEG signals. The offer the quantification of the recurring patterns in phase space trajectories [19, 20]. Introduced by Trulla et al.[19] (directly built on quantifying recurrence plots[24]) and expanded by Webber and Zbilut[20], RQA measures metrics like recurrence rate, determinism, and laminarity to capture dynamic system behavior. Thomasson et al.[21] in their work, demonstrated RQA's applicability on EEG data, mentioning the robustness it shows in accordance to noise and nonstationarity. Marwan et al.[22] further advanced recurrence plot techniques, emphasizing on developing a confidence measure of RQA in detecting dynamic transitions. Works like these, serve as a foundation of applying RQA and CRQA on EEG studies in various conditions such as epilepsy, cognitive disorders and others. XXX

1 Related Work

Frolov et al.[1] proposed an approach to analyze frequency based multiplex brain networks using recurrence quantification analysis (RQA) on EEG data, and demonstrated the way that recurrence-based synchronization indices can effectively capture both within-frequency (intralayer) and cross-frequency (interlayer) functional connectivity during cognitive tasks. Their work showed that RQA is particularly suitable for analyzing non-stationary EEG signals and revealed important insights about the evolution of functional connectivity patterns during cognitive tasks. In addition the dataset used in this research are openly available in a Figshare repository.

Núñez et al. [16] worked with resting-state EEG recordings from subjects with mild cognitive impairment(MCI), Alzheimer's disease(AD), and healthy ground truth controls in order to detect frequency based changes into their brain dynamics. By blending wavelet based Kullback–Leibler divergence (KLD) for capturing non-stationarity, and two RQA metrics(*entropy of the recurrence point density* and the *median of the recurrence point density*) insights have been extracted related to neurodegeneration presence. Research's findings show that MCI and AD are presenting notable changes in the recurrence structure and non-stationarity of EEG signals, and more specifics on the theta and beta frequency bands. Therefore, recurrence based dynamics show a capability as potential biomarkers for monitoring and detecting early Alzheimer's disease and its progression.

MCI has also investigated by Timothy et al.[25], where researchers have focused on the classification of MCI using EEG signals and combining RQA and CRQA methods. Analysis has been performed on both resting-state (eyes closed) and task-based (short-term memory) EEG data, focusing on complexity (via RQA) and synchronization (via CRQA) features. Their results indicate that MCI patients exhibit lower complexity and higher inter- and intra-hemispheric synchronization compared to healthy controls, particularly during memory tasks. The study also proposes a novel feature space approach using RQA and CRQA measures, achieving high classification accuracy (91.7%) under task conditions.

Fan and Chou [15] have also proposed an approach for real-time epileptic seizure detection using as a method the analysis of temporal synchronization patterns of EEG signals with recurrence networks and spectral graph theory. Recurrence plots were used for the modeling of the EEG dynamics, extracting graph theory's features for quantifying the synchronization. Results showed high sensitivity of 98.48% and low latency (6 seconds) for detecting seizure on the CHB-MIT dataset, performing better than other RQA measures.

Heunis and co-authors[23] have utilized resting state EEG and RQA in order to distinguish individuals of ages 0-18 of two categories; ASD(autism spectrum disorder) and typically developing.

RQA features were extracted and tested on various linear and nonlinear classifiers achieving 92.9% classification accuracy with nonlinear SVM classifier.

Author in [6], investigated changes related to aging in brain sensorimotor systems using RQA and theta-band functional connectivity in EEG signals. In the study a VR experimental paradigm was utilized with auditory stimulus across different age groups(young and elder subjects). Key findings include that elder subjects present decreased EEG complexity during motor preparation stages as measured by RQA metrics ($\Delta RR\ and\ \Delta RTE$), and had increased theta band functional connectivity highlighting the potential of RQA in detecting age related biomarkers that were not detectable using standalone signal spectral analysis.

Guglielmo et al. [7] utilized RQA features extracted by EEG signals for the purpose of classification of cognitive performance during mental arithmetic tasks. They used frontal and parietal EEG signals and analyzed them, from 36 participants by extracting six RQA metrics (*recurrence rate, determinism, laminarity, entropy, maximum diagonal line length and average diagonal line length*) from four electrodes (F7, Pz, P4, Fp1). Afterwards by applying machine learning classifiers (SVM, Random Forest, and Gradient Boosting) and they reached accuracy of classification above 0.85, showing the potential that RQA holds for generalizing on nonlinear dynamics.

Mihajlović [33] studied the discriminative effiency of traditional spectral features in comparisson to RQA-derived nonlinear metrics for the cognitive effort classification purposes. Utilizing a 4-channel wearable EEG headset, data was recorded while subjects perform tasks having variable cognitive load such as relaxation, math, reading. The key finding was that while spectral features alone often yielded higher classification accuracy, RQA features such *recurrence rate,determinism ratio* were consistently ranked among the most important features for discrimination task. A conjuction of a hybrid model using both spectral and RQA features achieved the best overall performance, showing the complementary nature of the methods in brain dynamics exploration.

Yang and co-authors [17], examined stereo electroencephalography (sEEG) recordings of 10 patients with refractory focal epilepsy for analyzing dynamical differences among discreet epileptic phases/states (inter-ictal, pre-ictal, and ictal) and regions. Using recurrence plots and CRQA, they identified epileptogenic channels with longer diagonal structures in RPs, which is a sign of more deterministic and recurrent dynamics. Their findings point out that the synchronization among the epileptogenic channels strengthened while seizures events occur, suggesting that these regions dominate the network's dynamics.

Lopes et al. [8] have proposed a combinatorial framework by mixing RQA with dynamic functional network (dFN) analysis, applying it to both MEG and stereo EEG data. The methodology they described is split into five steps: data segmentation, functional network inference, distance computation alongside networks, recurrence plot construction and finally RQA. The study demonstrated that functional networks in epilepsy patients recur more quickly than in healthy controls, suggesting RQA on dFNs could play the role of a potential biomarker. For the EEG dataset investigation, they have showed that the pre-ictal networks shown higher recurrence rates than post-ictal periods, with the τ-recurrence rate (RR_τ) proving particularly effective for seizure detection.

Rangaprakash [18] have proposed an application of RQA for the study of brain connectivity using multichannel EEG signals. In its work, a new CRQA-based feature was proposed (Correlation between Probabilities of Recurrence (CPR)), a nonlinear and non-parametric phase synchronization technique. Afterwards it was utilized for the analysis of functional connectivity in epilepsy subjects during eyes-open/eyes-closed conditions. The results demonstrated that CPR outperformed other known traditional linear methods on distinguishing seizure and pre-seizure states, identifying epileptic foci, and differentiating alongside eyes-open and eyes-closed conditions.

In another study which demonstrates the effectiveness of RQA in analyzing EEG signals for epilepsy detection, Gruszczyńska et al.[11] applied RQA on such signals in order to distinguish epileptic from healthy patients using recordings from frontal and temporal lobe electrodes (Fp1, Fp2, T3, T4). In their findings they have showed that the epileptic signals present more periodic dynamics in comparison to healthy controls, by producing higher values of RQA parameters such as determinism, laminarity, and longest diagonal line. The combination of RQA with Principal Component Analysis for dimensionality reduction and visualization, achieved 86.8% classification accuracy with SVM. Authors also demonstrated RQA's capability to identify pathological patterns in EEG signals without the requirement of seizure events during recording which have bad impact on the subject's health.

Another study utilizing advanced nonlinear analysis techniques for neural correlation investigation to cognitive functions [12] used *stereoelectroencephalography (sEEG)* combined alongside RQA for the examination of the relationship of the DMN and empathy. Correlations have been detected relating specific RQA metrics (mean diagonal line length, entropy of diagonal line lengths, trapping time) and empathy scores, particularly within DMN subsystems.

Regarding epilepsy diagnosis, authors in [13] proposed a new framework utilizing the combintation of RQA with genetic algorithms and Bayesian classifiers for identifying corresponding biomarkers for seizure detection. They utilized five distance norms (e.g., Euclidean, Mahalanobis) and multiple thresholds for extracting recurrence features from EEG signals, achieving 100% classification accuracy. More specific, the *transitivity* feature has shown capability of a highly discriminative biomarker, performing better compared to traditional linear methods.

Ngamga et al.[14] studied the performance achieved of RQA and Recurrence Network (RN) measures in identifying pre-seizure states from multi-day, multi-channel intracranial EEG (iEEG) recordings of epilepsy patients. Results highlighted the correlation among RQA measures (determinism, laminarity, and mean recurrence time) in detecting seizure precursors, while RN measures (average shortest path length and network transitivity) provided complementary but not so consistent insights than using the application of RQA measures alone.

Gao et al.[45] examined the application of RQA in the domain of automated epilepsy detection. Authors utilized a hybrid scheme combining nonlinear features(related to Approximate Entropy(ApEn) and RQA metrics) from the publicly available Bonn EEG dataset[46] with a deep learning classifier. Their key finding was that while ApEn and RQA features alone could achieve good classification accuracy, their performance was increased when used as input features for a Convolutional Neural Network (CNN). By constructing this hybrid approach, classification accuracy rised on 99.26% for distinguishing ictal from inter-ictal and healthy EEG signals, demonstrating the potential of the synergy among traditional metrics and modern deep learning architectures.

Researchers in [3], have applied RQA on resting-state fMRI data from TgF344-AD rats(a transgenic rat model which will eventually develop Alzheimer's disease) and their healthy-control counterparts wild-type rats(WT), in order to detect early stage biomarkers for the disease. By analyzing Default Mode-Like Network (DMLN) using RQA metrics(*entropy, recurrence rate, determinism and average diagonal line length*) changes have been detected in regions of the basal forebrain, hippocampal fields (CA1, CA3), and visual cortices (V1, V2). Also on the study's findings include reduced predictability in WT rats with aging, while AD rats exhibited less decline in predictability, suggesting some unknown yet countereacting mechanisms. This study highlights RQA's sensitivity for nonlinear dynamics in preclinical AD and the code used is also publicly available.

Lombardi et al.[5] investigated the nonlinear properties in fMRI BOLD signals during a working memory task in schizophrenic patients and healthy controls. They have attempted by using RQA, to analyze recurrence plots for quantifying determinism, trapping time, and maximal vertical line length in functionally relevant brain clusters. Outcome revealed differences in the dynamics between the two groups, and more specific in working memory and DMN areas. While their work have focused on fMRI, the methodology can be adapted also into EEG signals, which can offer a higher resolution for capturing rapid neural dynamics.

Kang et al. [2], in their study explore the dynamics and functional connectivity of the Default Mode Network (DMN) in schizophrenia, applying RQA-CRQA on resting-state fMRI data. Findings include decreased *determinism* between specific DMN regions (vMPFC-posterios cingulate and vMPFC-precuneus) in first-episode schizophrenia patients, as a signal of disturbed predictability of functional interactions. Moreover, their results achieve to correctly classify using SVM(support vector machine) schizophrenia patients from healthy controls with 77% classification accuracy.

In their research, Pentari et al.[9] have applied CRQA to resting-state fMRI data for examining the dynamic functional connectivity on patients with neuropsychiatric systemic lupus erythematosus (NPSLE). Results contain the fact that CRQA metrics, such as determinism, appear more sensitive than conventional static functional connectivity methods in order to identify aberrant connectivity patterns that correlated with visuomotor performance. The study focused on 16 frontoparietal regions and found that CRQA could detect both increased and decreased connectivity in NPSLE patients compared against the healthy controls. Building on these findings, Pentari et al.[10] subsequently expanded the investigation to whole brain network analysis in a larger cohort. In this study they demonstrate the capability of CRQA to integrate multiple recurrence metrics for revealing both hyperconnectivity in parietal regions (angular gyrus and superior parietal lobule) and hypoconnectivity in medial temporal structures (hippocampus and amygdala).

In addition there have been works where simulated data have been used in conjunction with RQA. Lameu et al.[4], investigated burst phase synchronization in neural networks using RQA. They analyzed two network types; a small-world network and a network of networks (to mimic better the real human brain), using coupled Rulkov maps to model bursting neurons. By applying RQA, they identified synchronized neuron groups and quantified their sizes during synchronization transitions. The study showed that RQA measures (*recurrence rate, laminarity inspired*(custom feature)*, and average structure size*) complement traditional order parameters by revealing localized synchronization patterns, such as the formation and growth of synchronized clusters. Kashyap and Keilholz[26] conducted a comprehensive comparison between simulated brain network models (BNMs) and real rs-fMRI data using dynamic analysis techniques, including Recurrence Quantification Analysis (RQA). In the study they employed two BNMs, the Kuramoto oscillator model and the Firing Rate model, for simulating the whole-brain activity, which was then compared to human rs-fMRI data. Among the compared dynamic analysis methods, RQA was proved particularly effective in distinguishing between the models and empirical data, demonstrating that RQA metrics (*recurrence rate, entropy, and average diagonal length*) could robustly separate the empirical data from simulations.

Shalbaf et al. [28] investigated the synchronization of EEG signals between frontal and temporal regions during propofol anesthesia using *Order Patterns Cross Recurrence Analysis* (OPCR). Their study introduced a novel index, *Order Pattern Laminarity* (OPL), for the quantification of neuronal synchronization and compared its performance with the traditional Bispectral Index (BIS). The results demonstrated that OPL correlated more strongly with propofol concentration ($P_k = 0.9$) and exhibited faster response times to transient changes in consciousness compared to BIS. Additionally, OPL showed lower variability at the point of loss of consciousness (LOC), suggesting its robustness as a measure of anesthetic depth. This work highlights the potential of recurrence-based methods (e.g., CRQA) for analyzing brain network dynamics under anesthesia, particularly in noisy, non-stationary EEG data.

Table 1: Comparison among the retrieved studies using recurrence analysis

#	Reference	Modality	Analysis Methods	Network Type
1	Frolov et al. (2020)	EEG	RQA, CRQA	Multiplex functional networks
2	Kang el al. (2023)	fMRI	RQA, CRQA	DMN, schizophrenia
3	Rezaei el al. (2023)	fMRI	RQA	Default model-like network, AD
4	Lameu et al. (2018)	—	RQA	Small-world & cluster network
5	Lombardi et al. (2014)	fMRI	RQA	schizophrenia,working memory
6	Pitsik E. (2025)	EEG	RQA	aging
7	Guglielmo et al. (2022)	EEG	RQA	cognitive tasks
8	Lopes et al. (2020)	sEEG, MEG	RQA	epilepsy
9	Pentari et al. (2022)	fMRI	RQA, CRQA	NPSLE
10	Pentari et al. (2023)	fMRI	CRQA	NPSLE
11	Gruszczyńska et al. (2019)	EEG	RQA	epilepsy
12	Mo et al. (2022)	sEEG	RQA	DMN, epilepsy
13	Palanisamy et al. (2024)	EEG	RQA	epilepsy
14	Ngamga et al. (2016)	EEG	RQA,RN	epilepsy
15	Fan and Chou (2019)	EEG	RQA,RN	epilepsy, seizure detection
16	Nunez et al. (2020)	EEG	RQA	AD
17	Yang et al. (2019)	sEEG	RQA,CRQA	epilepsy
18	Rangaprakash (2014)	EEG	CPR(CRQA-based)	epilepsy
19	Heunis et al. (2018)	rsEEG	RQA	autism spectrum disorder
20	Timothy et al. (2017)	EEG	RQA-CRQA	MCI
21	Kashyap et al. (2019)	fMRI	RQA	distinguish BNMs
22	Shalbaf et al. (2014)	EEG	CRQA(OPL)	Anesthesia depth monitoring
23	Mihajlović. (2019)	EEG	RQA	cognitive tasks

1.1 RQA relevant patents utilizing EEG modality

The application of RQA utilizing EEG modality on the biomedical field, shows an increasing interest not only in academic research, but also in commercial and clinical applications, as we can inspect on recent patent filings. Reviewing these documents can reveal industrial viable solutions being developed for real-time, embedded systems.

Becker et al.[42] describes on patent (US20080234597A1) a monitoring device and method for creating an assessment of the depth of anesthesia or coma characterizing an individual subject. Authors analyzes neuronal EEG data and uses RQA to compute a complexity parameter that quantitatively reflects the level of consciousness. In the device's core, a buffer is utilized for storing time-series data and an analysis circuit performs RQA by reconstructing phase-space trajectories, calculating recurrence plots, and extracting determinism-based complexity measures. This makes possible monitoring the depth of anesthesia level in real-time and can be utilized in clinical applications for anesthesia control during surgery or even in long term coma assessment.

Patent US20250195894A1 [39], entitled “Systems and Methods for Seizure Detection and Closed-Loop Neurostimulation,” Inventors proceed in an alternative calculation of RQA measures which entirely bypass the construction of the recurrence plot matrix(RP). They achieve this, by not creating the traditional RP in order to extract certain metrics from, but by calculating them, *on-the-fly*; dynamically accumulating the lengths of diagonal lines as each new data point is processed. This method offers the following advantages:

1. **Memory Efficiency:** It does not require the large $N \times N$ comparison matrix, reducing memory usage by approximately 88%.

2. **Computational Efficiency:** It avoids the expensive read/write cycles associated with managing the large matrix, reducing processing time by approximately 30% per channel.

In addition, the patent by [40], titled ”An EEG signal classification model based on genetic algorithm and random forest”, presents a framework for EEG signal classification. The inventors propose a hybrid model consisting of three key stages:

1. **Feature Extraction:** The method employs a multi-modal feature extraction strategy. Among other features, it explicitly includes *RQA* metrics from the EEG signal, alongside other traditional time-domain/frequency features.

2. **Feature Optimization:** A genetic algorithm (GA) is then utilized for feature selection. Inventors use binary encoding for representing chromosomes, where each bit corresponds to the selection (1) or rejection (0) of a specific feature from the large extracted pool. The aim of this procedure is to optimize the feature subset in order to have maximum discriminative power.

3. **Classification:** The optimized feature subset is fed into a Random Forest classifier for a final prediction.

The patent claims that this integrated approach, validated on a public dataset, yields a superior classification accuracy compared to existing methods at the time of filing, while also demonstrating robustness through cross-validation.

Another patent[41] (CN106512206B) describes an implantable closed-loop deep brain stimulation (DBS) system that uses electrophysiological signals (specific deep brain local field potentials (LFPs) and ECG signals) for monitoring the states of a human sleeping and adjusting various stimulation parameters in real time. A device acquires ECG and deep brain signals for feature extraction in the domains of time and frequency, while calculating complexity measures. RQA metrics such recurrence rate, determinism, entropy and laminarity are utilized alongside other complexity and spectral features to classify sleep states and trigger appropriate stimulation responses. Features are then utilized for detection of sleep stages and for emergency detection/alerts (f.e, cardiac arrest or abnormal excitation).

2 The CHB-MIT EEG Database

The **CHB-MIT Scalp EEG Database**[47] is a public collection of EEG recordings from pediatric subjects that includes intractable seizures. All the recordings of the database were collected at the Children’s Hospital Boston. The dataset contains multiple cases (patients), each with long-term scalp EEG signals recorded using the international 10–20 system. Table 2 provides a summary of the demographic information of the database's subjects. This particular database is used in many studies for testing algorithms on epileptic seizure detection and epileptic research in general.

3 Dataset Description

Table 2: Demographic information for patients in the CHB-MIT Scalp EEG Database.

Case	Gender	Age (years)
chb01	F	11
chb02	M	11
chb03	F	14
chb04	M	22
chb05	F	7
chb06	F	1.5
chb07	F	14.5
chb08	M	3.5
chb09	F	10
chb10	M	3
chb11	F	12
chb12	F	2
chb13	F	3
chb14	F	9
chb15	M	16
chb16	F	7
chb17	F	12
chb18	F	18
chb19	F	19
chb20	F	6
chb21	F	13
chb22	F	9
chb23	F	6

The database includes recordings from male and female patients, with their ages in the range of 1.5 to 22 years old. Most subjects are children, a fact reflecting the pediatric nature of the dataset. This demographic diversity provides a representative sample for studying epileptic activity across different developmental stages.

The total recording duration spans approximately 982 hours, segmented into 664 different EDF files, where each recording contains 1 hour of data (though durations vary from 1 to over 4 hours). EEG signals were recorded using 23 scalp electrodes (according to the international 10–20 system), but in some files there are extra channels records, such as ECG and EMG references. The sampling rate of the recordings is 256 Hz with 16-bit resolution. Seizure events are annotated by domain experts, providing a time-stamp for seizure onsets and offsets.

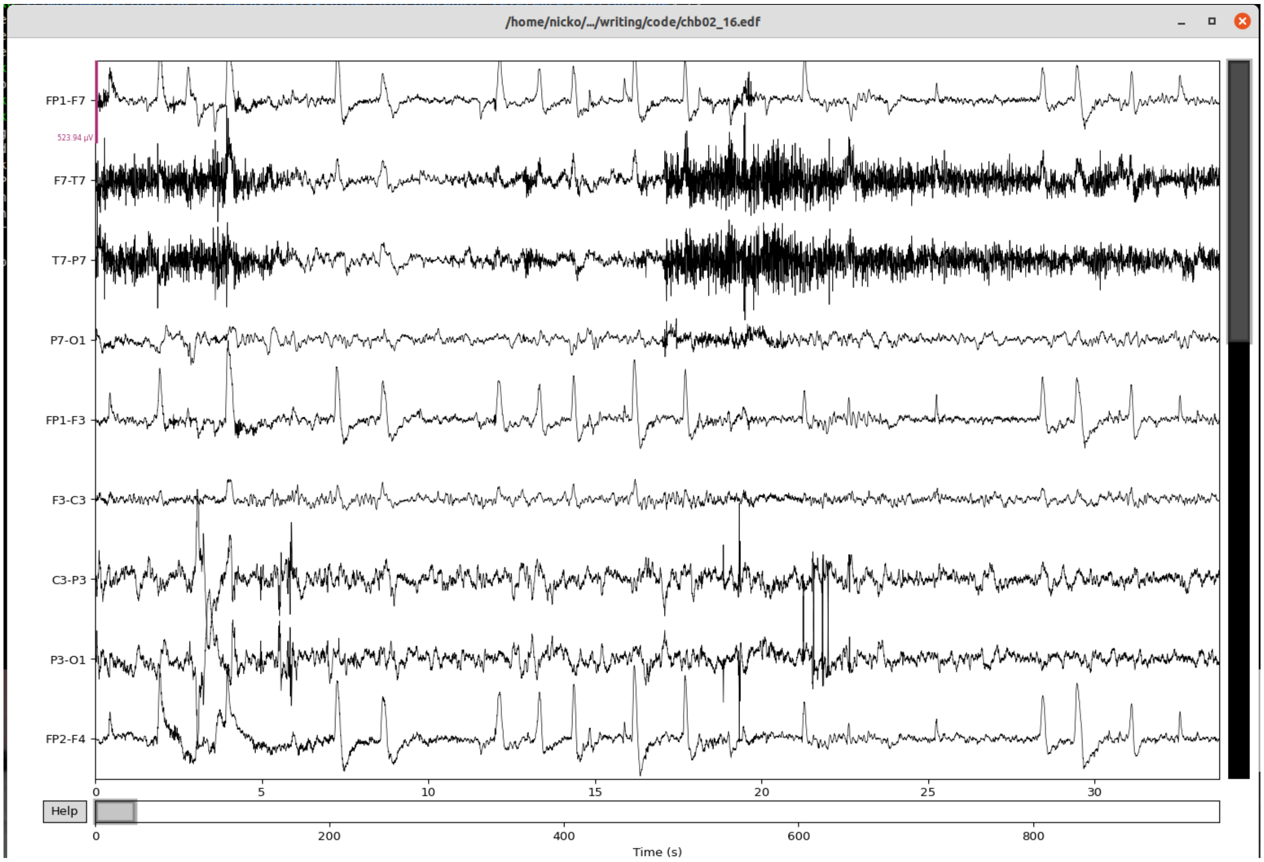


Figure 1: EEG recording visualized utilizing Python-MNE, for a patient of the dataset.

4 Filtering

Signals of electroencephalogram, in most cases carry not only the desired signal but also added noise and artifacts from physiological (eye blinks, muscle or cardiac activity) and non-physiological sources (f.e, powerline interference). In general, artifacts in a recording consist of all the non-neural signals that are mixed with the pure EEG data.

Preprocessing is required in order to increase the quality of the signal(signal-to-noise ratio), to boost the performance of further examintations in later pipelines in domains like brain-computer interfaces (BCIs) or clinical diagnostics[29, 30, 31, 32].

Some of the common preprocessing techniques are:

- **Filtering** (f.e, Butterworth, Chebyshev) to remove unwanted frequency bands.
- **Regression methods** used for removing ocular artifacts with help of reference channels.
- **Blind Source Separation (BSS)** (f.e, ICA, CCA), decompose and isolate neural activity from artifacts.

- **Wavelet/EMD-based methods** for non-stationary artifact removal.

Also hybrid approaches (f.e, wavelet-ICA) exist, combining multiple techniques for improved artifact rejection. The choice of method is related on computational constraints, artifact type, and the satisfaction of real-time processing needs, if any.

Effective preprocessing is a key, that ensures reliable feature extraction for later analysis.

Among these, wavelet-based methods have have been proposed and used for effective EEG denoising based on the non stationary nature of brain signals. Transient artifacts appearing in EEGs in varying frequency patterns are the reason that traditional linear filtering methods struggle to handle them without introducing distortions. Wavelet transforms on the other hand, decompose the signal into time-frequency representations using scalable, localized basis functions called wavelets, enabling the analysis: the signal is broken down into approximation (low-frequency) and detail (high-frequency) coefficients across decomposition levels. Artifacts, such as ocular blinks or EMG bursts, manifest as sparse, high-amplitude coefficients that can be selectively silenced using thresholds (f.e, hard or soft), followed by reconstruction—thus removing noise while preserving neural transients like epileptic spikes [55].

4.1 Evaluation metrics for EEG denoising in the CHB-MIT dataset

In order to evaluate the performance of different wavelet-based filters for EEG denoising, quantitative metrics have been computed over entire recordings per channel and filter configuration. Metrics used were:

- **Signal-to-Noise Ratio (SNR, dB)**: Measure of the ratio between the power of the clean signal and the power of the noise. Higher values indicate better noise suppression while preserving the structure of the signal.
- **Root Mean Square Error (RMSE, μV)**: This metric quantifies the average deviation between the denoised and reference signal in microvolts. Lower values indicate closer similarity to the original signal.
- **Normalized RMSE (NRMSE, %)**: RMSE normalized by the dynamic range of the reference signal expressed as a percentage. Lower values represent better performance.
- **Correlation Coefficient**: Pearson's correlation among the denoised and reference signal, providing an assessment of similarity of the two waveforms. Observed values close to 1 indicate high waveform preservation.
- **Percent Root-mean-square Difference (PRD, %)**: Quantification of the relative distortion introduced by the denoising process. Lower values indicate less distortion.

For each one of the filters configuration, metrics were computed channel-wise and then averaged across all channels in order to calculate their global performance score.

4.2 Filter Selection Criteria

The selection of the optimal filter was based on a multi-criteria ranking strategy, where:

1. Metrics where *higher* values indicate better performance (**SNR**, **Correlation**) were ranked in descending order.

2. Metrics where *lower* values indicate better performance (**RMSE**, **NRMSE**, **PRD**) were ranked in ascending order.

3. The ranks from all metrics were averaged to obtain an overall performance rank for each filter.

The filter with the lowest average rank was considered the best compromise between noise reduction and signal fidelity. According to this evaluation, the **SYM8, level 4, threshold 0.5, hard thresholding** filter presented the highest overall performance, exhibiting:

- the highest SNR values,
- one of the lowest RMSE and the lowest NRMSE value,
- the highest correlation coefficient,
- and the lowest PRD.

This indicates that the chosen filter effectively suppressed noise while preserving the morphological features of the EEG signal, making it the most suitable choice for subsequent analysis. The results of the benchmarking those metrics in 5 EDF recording files which include seizures are presented in the following figure.

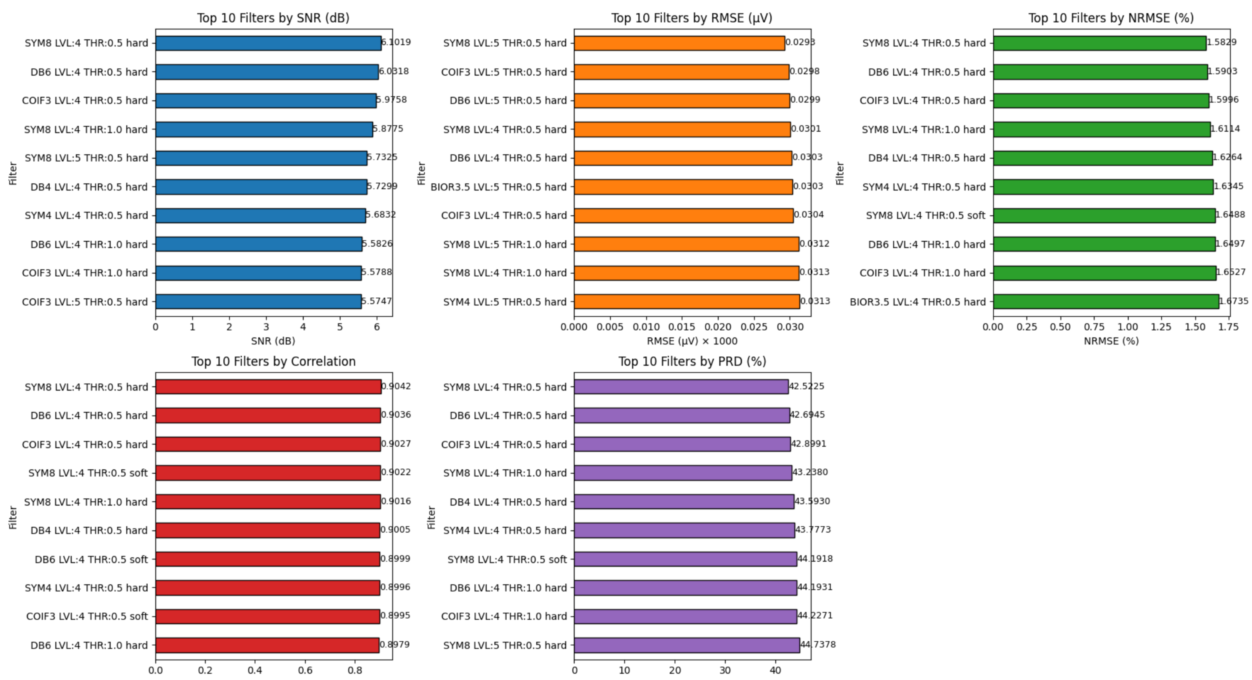


Figure 2: The top 10 filter configurations per EEG metric. RMSE values scaled.

In the following figures we present a visualization using the recording named *chb01_03.edf* comparing the original 10 first EEG channels against the filtered ones with the respective wavelets filters.

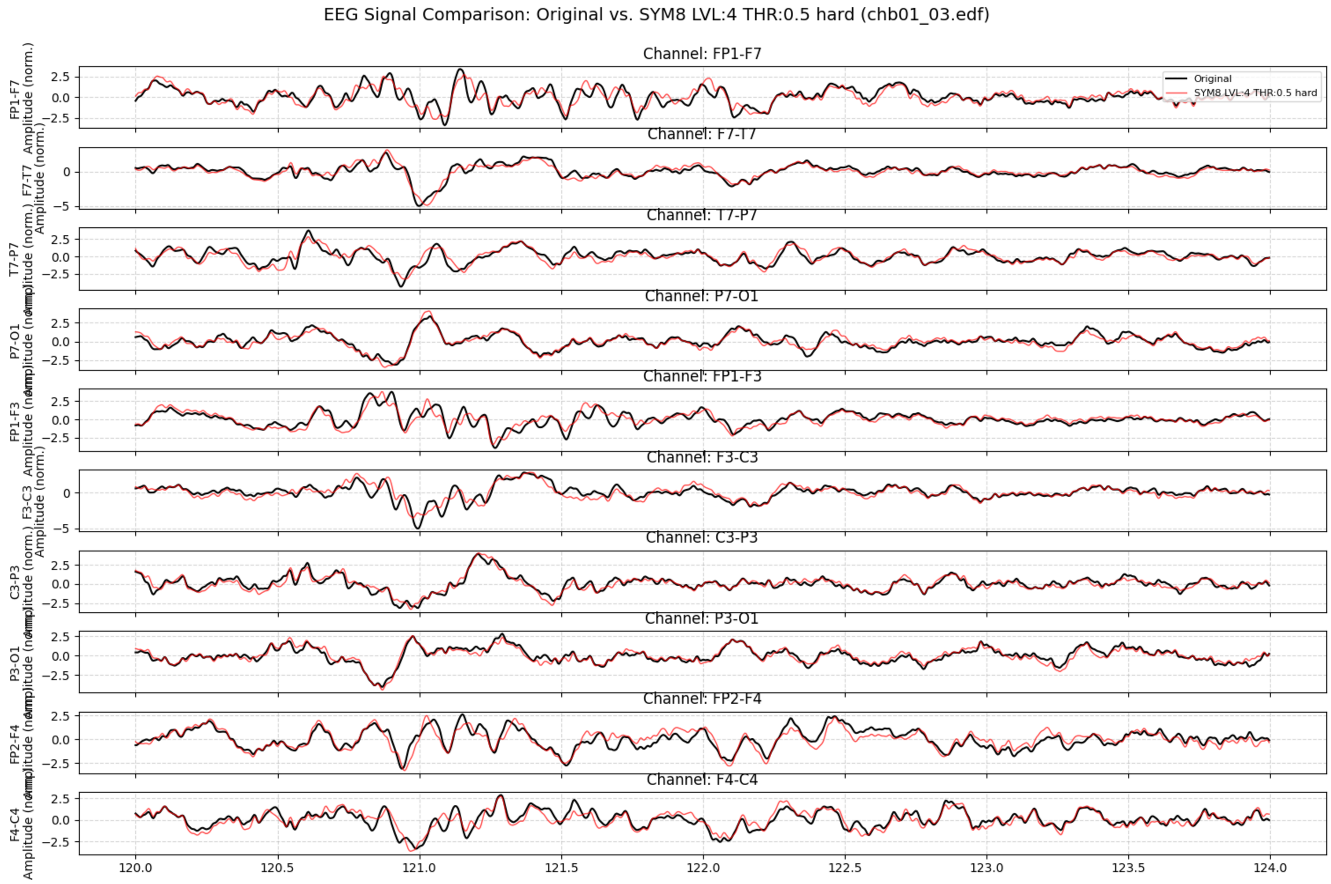

EEG Signal Comparison: Original vs. SYM8 LVL:4 THR:0.5 hard (chb01_03.edf)
Channel: FP1-F7
Channel: F7-T7
Channel: T7-P7
Channel: P7-O1
Channel: FP1-F3
Channel: F3-C3
Channel: C3-P3
Channel: P3-O1
Channel: FP2-F4
Channel: F4-C4
Original
SYM8 LVL:4 THR:0.5 hard
Amplitude (norm.)

Figure 3

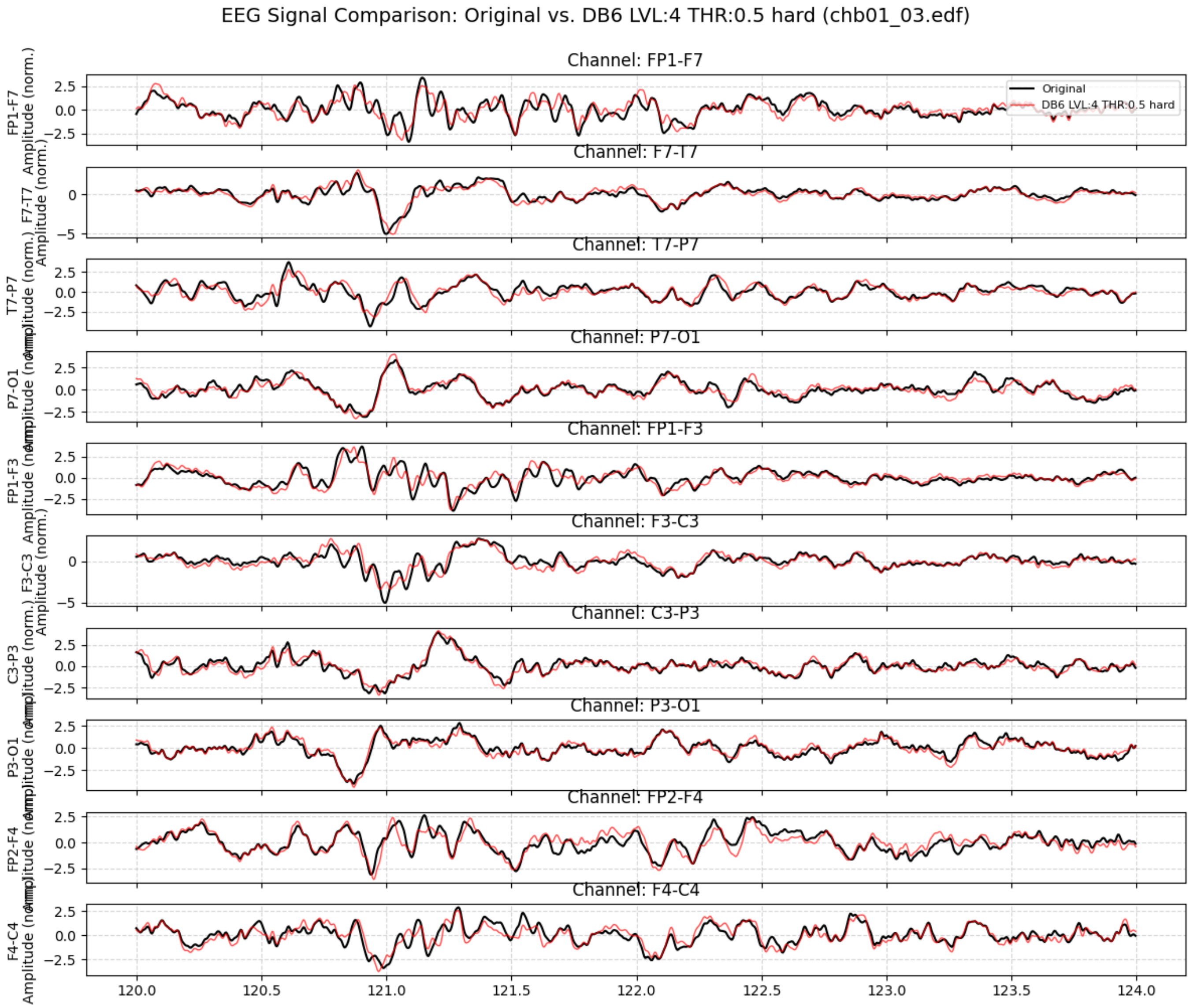

EEG Signal Comparison: Original vs. DB6 LVL:4 THR:0.5 hard (chb01_03.edf)
Channel: FP1-F7
Channel: F7-T7
Channel: T7-P7
Channel: P7-O1
Channel: FP1-F3
Channel: F3-C3
Channel: C3-P3
Channel: P3-O1
Channel: FP2-F4
Channel: F4-C4
Original
DB6 LVL:4 THR:0.5 hard
Amplitude (norm.)
120.0
120.5
121.0
121.5
122.0
122.5
123.0
123.5
124.0

Figure 4

5 Phase space reconstruction of EEG signals

In order to analyze multi-channel EEG's nonlinear dynamical system, the reconstruction of the underlying phase space is required, based on the scalar measurements of each channel. According to Takens' embedding theorem [34], a time series $x(t)$ can be embedded in an m-dimensional space using time-delay coordinates:

$$\vec{y}(t) = [x(t), x(t+\tau), x(t+2\tau), \ldots, x(t+(m-1)\tau)] \tag{1}$$

where m is the embedding dimension and τ is the time delay. The critical challenge lies in determining the appropriate values for these parameters to faithfully reconstruct the system's dynamics without distortion.

5.1 Determination of Embedding Parameters

The reconstruction of the phase space from a single time series $x(t)$ requires the specification of two parameters: the time delay τ and the embedding dimension m. These two parameters determine how the reconstruction will represent and how close it will reveal the underlying dynamics without distortion.

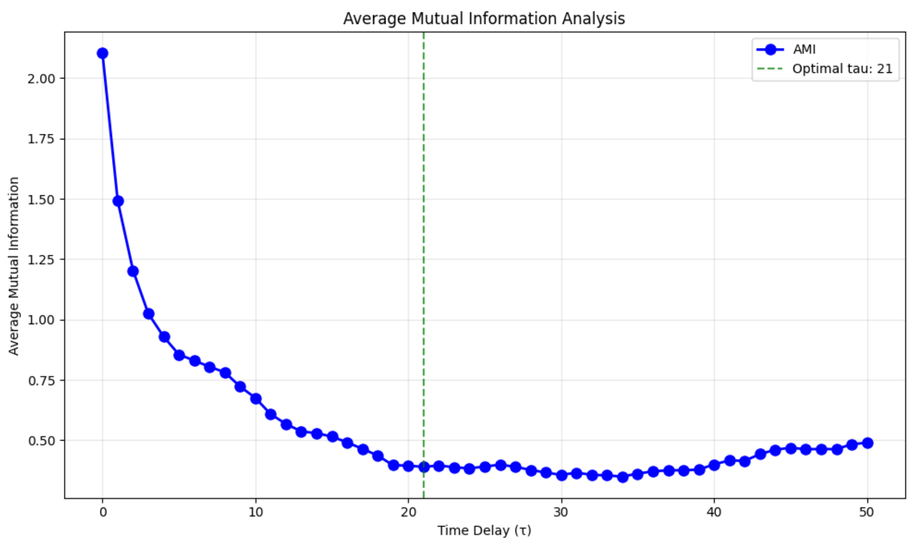


Figure 5: Calculation of τ using AMI for a sample EEG channel. The first minimum of the AMI function (green dashed line) is chosen to become the optimal τ to ensure independence between delay coordinates.

5.1.1 Calculation of time delay τ utilizing mutual information

The time delay τ can be estimated by applying the *Average Mutual Information* (AMI) method, a concept which was first introduced by Fraser and Swinney [36]. In contrast to linear autocorrelation, mutual information has the ability to capture both linear and nonlinear dependencies among the original time series $x(t)$ and its delayed version $x(t+\tau)$.

The mutual information $I(\tau)$ between $x(t)$ and $x(t+\tau)$ is defined as:

$$I(\tau) = \sum_{x(t),\, x(t+\tau)} P(x(t), x(t+\tau)) \log_2 \left(\frac{P(x(t), x(t+\tau))}{P(x(t))\, P(x(t+\tau))} \right)$$

where $P(\cdot)$ denotes probability.

The optimal time delay τ is chosen as the value at which $I(\tau)$ reaches its *first minimum*. This value indicates a good compromise between independence (too small τ) and irrelevance (too large τ) of the coordinates in the embedding vector.

5.1.2 Estimating embedding dimension m using false nearest neighbors approach

When the embedding dimension m is too small, the phase space becomes *projected* rather than properly *embedded*. This projection can create artificial neighborhoods where points appear to be close due to geometrical constraints of the space rather than their actual dynamical similarity. These are named as *false nearest neighbors*.

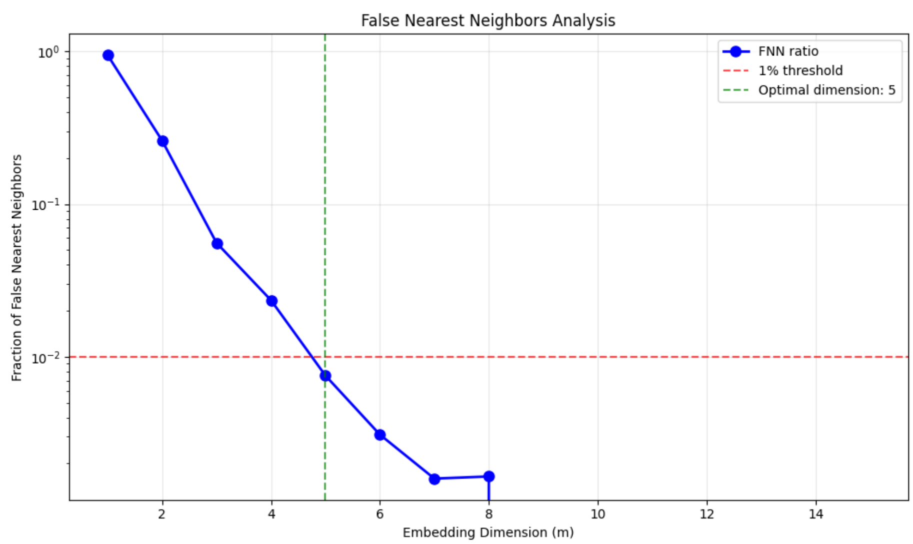


Figure 6: Calculation of the embedding dimension using the FNN scheme.

Mathematically, two points $\vec{y}_i$ and $\vec{y}_j$ are false neighbors if their distance increases significantly when embedded in higher dimension:

$$\frac{\|\vec{y}_i^{(m+1)} - \vec{y}_j^{(m+1)}\|}{\|\vec{y}_i^{(m)} - \vec{y}_j^{(m)}\|} > R_{\text{tol}} \tag{2}$$

where R_{tol} is a tolerance threshold (typically 10–15).

The False Nearest Neighbors(FNN) method [35] provides a systematic approach in order to determine the minimal sufficient embedding dimension. The method's steps are:

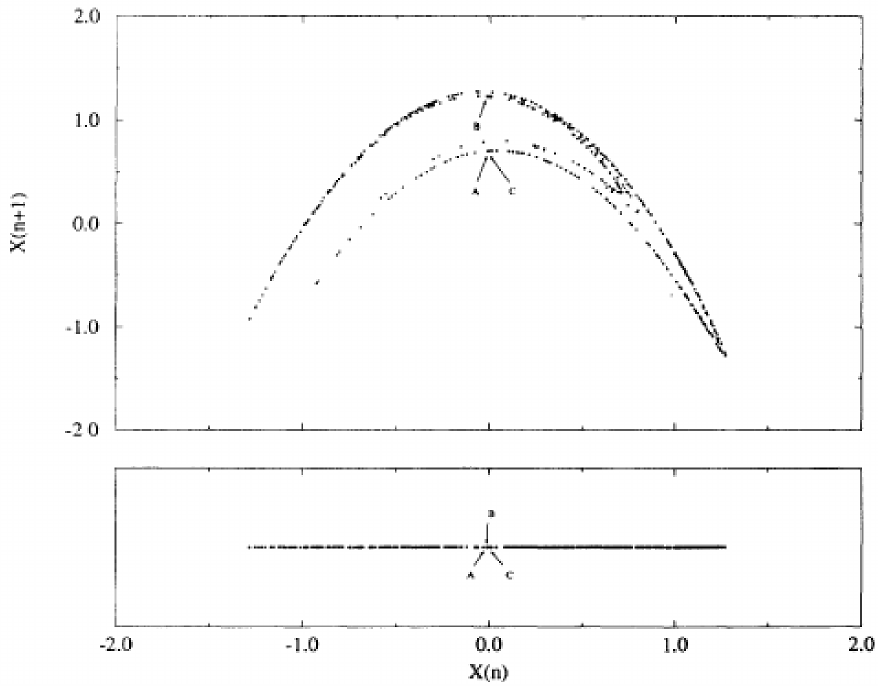


Figure 7: Schematic illustration of false neighbors. In insufficient embedding dimension (down), points A and B appear neighbors due to projection. When proper embedding is employed(up), their true separation is revealed.

1. For each point in dimension m, identify its nearest neighbor.
2. Embed the data in dimension $m + 1$.
3. Calculation the relative distance increase between each point and its former neighbor.
4. If the increase exceeds predetermined thresholds, the neighbor point is classified as a false neighbor
5. The optimal m is the smallest dimension where the fraction of false neighbors drops below an acceptable level (typically 1–5%)

Both relative and absolute criteria are included in the algorithm:

$$\text{Relative:} \quad \frac{\|\vec{y}_i^{(m+1)} - \vec{y}_j^{(m+1)}\|}{\|\vec{y}_i^{(m)} - \vec{y}_j^{(m)}\|} > R_{\text{tol}} \tag{3}$$

$$\text{Absolute:} \quad \|\vec{y}_i^{(m+1)} - \vec{y}_j^{(m+1)}\| > A_{\text{tol}} \cdot \sigma_x \tag{4}$$

where σ_x is the standard deviation of the time series.

When optimal parameters τ and m have been determined for a given signal, the phase space can be reconstructed according to Takens' theorem. This reconstruction provides the geometric picture of the underlying dynamics.

Figure 8 presents the reconstructed phase space for a normal EEG segment from channel 'Fp1-F7' from CHB-MIT 's chb24_01.edf data.

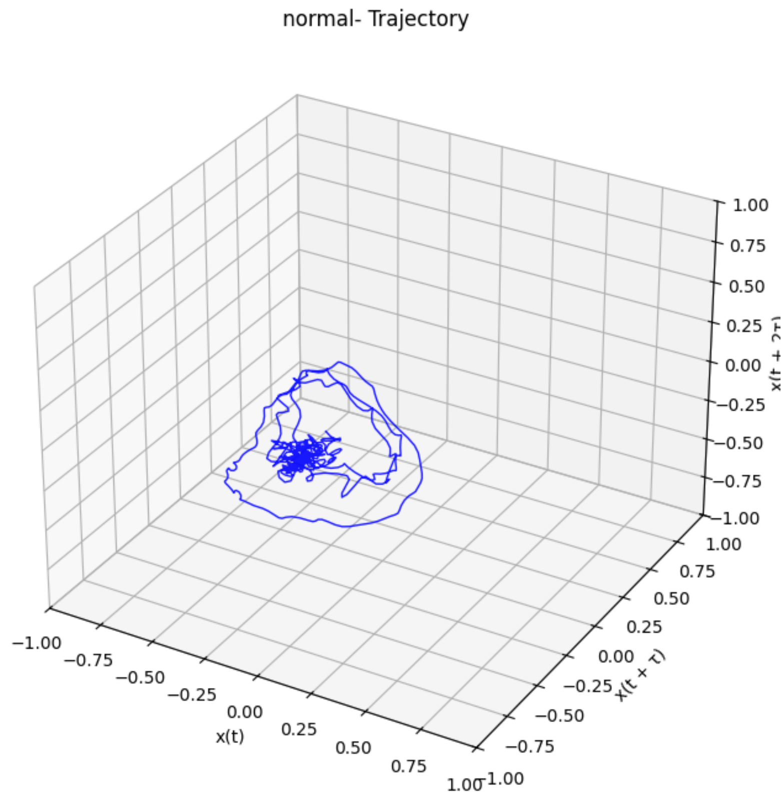


Figure 8: 3D phase space reconstruction for a 3-second EEG segment's channel.

6 Recurrence Quantification Analysis (RQA)

Having reconstructed the phase space trajectory of the EEG signals, the next step is to analyze its dynamical properties. Recurrence Quantification Analysis is a powerful nonlinear method that provides precisely this functionality by quantifying the number and duration of recurrences of a dynamical system to its previous states [37]. The core of this quantification process is the *Recurrence Plot (RP)*, a visualization which denotes the times at which the phase space trajectory revisits approximately the same area. In most non trivial cases, a phase space does not have a dimension (two or three) which allows a direct visualization, so for higher dimensional phase spaces the only solution is a projection into a two or three dimensional space. However, RP enables the examination of a higher-dimensional phase space trajectory via its two-dimensional representation of its recurrences.

6.1 The Recurrence Plot (RP)

RP is a symmetric, two-dimensional matrix that visualizes the recurrences of states. For a reconstructed trajectory $\vec{y}(t)$ of length N, the recurrence matrix $\mathbf{R}$ is defined as:

$$R_{i,j} = \Theta(\varepsilon - \|\vec{y}(i) - \vec{y}(j)\|), \quad i, j = 1, \ldots, N \tag{5}$$

where:

- $\Theta(\cdot)$ is the Heaviside step function ($\Theta(x) = 0$ if $x < 0$, and $\Theta(x) = 1$ otherwise),
- ε is a predefined distance threshold (radius),
- $\|\cdot\|$ is a norm.

By interpreting the RP, several metrics can be extracted for further analysis.

6.2 Key RQA Metrics and Their Interpretation

RQA provides a set of metrics that can quantify the number and the duration of the recurrences of a dynamical system. These metrics are categorized by those which are based on diagonal structures, which relate to the predictability and deterministic nature of the system, and those that are based on vertical structures, which can capture laminar states or chaos-chaos transitions.

The definitions of the core RQA metrics, as implemented in tools like the utilized `PyRQA`, are as follows [44]:

Recurrence Rate (RR) The recurrence rate is the simplest measure, defined as the density of recurrence points in the RP. It corresponds to the probability that a state recurs and is analogous to the correlation sum.

$$RR = \frac{1}{N^2} \sum_{i,j=1}^{N} R_{i,j}$$

Determinism (DET) Determinism quantifies the percentage of recurrence points that form diagonal lines. Diagonal lines are a signature of deterministic dynamics, where segments of the trajectory run in parallel for some time. A higher DET indicates a more predictable, deterministic system.

$$DET = \frac{\sum_{l=l_{\min}}^{N} l\,P(l)}{\sum_{l=1}^{N} l\,P(l)}$$

where $P(l)$ is the histogram of diagonal line lengths l, and $l_{\min}$ is the minimum line length (typically 2).

Laminarity (LAM) Laminarity measures the percentage of recurrence points that form vertical lines. Vertical lines indicate states that do not change or change very slowly for a period (laminar states). It can detect chaos-chaos transitions or intermittency.

$$LAM = \frac{\sum_{v=v_{\min}}^{N} v\,P(v)}{\sum_{v=1}^{N} v\,P(v)}$$

where $P(v)$ is the histogram of vertical line lengths v, and $v_{\min}$ is the minimum line length.

Ratio (RATIO) The ratio is a measure of complexity, calculated as the ratio between DET and RR. It can be sensitive to transitions between order and chaos.

$$RATIO = \frac{N^2 \sum_{l=l_{\min}}^{N} l\,P(l)}{\left(\sum_{l=1}^{N} l\,P(l)\right)^2}$$

Average Diagonal Line Length (L) This metric represents the average time that two segments of the trajectory remain close, providing an estimate of the mean prediction time.

$$L = \frac{\sum_{l=l_{\min}}^{N} l\,P(l)}{\sum_{l=l_{\min}}^{N} P(l)}$$

Trapping Time (TT) Trapping time is the average length of vertical lines, quantifying the mean time the system remains trapped in a specific state (laminarity in time).

$$TT = \frac{\sum_{v=v_{\min}}^{N} v\,P(v)}{\sum_{v=v_{\min}}^{N} P(v)}$$

Longest Diagonal Line ($\mathbf{L_{max}}$) The length of the longest diagonal line in the RP is related to the Lyapunov exponent of the system. A shorter $L_{\max}$ suggests a faster divergence of trajectories, which is a hallmark of chaos.

$$L_{\max} = \max(\{l_i \mid i = 1, \ldots, N_l\})$$

Divergence (DIV) Divergence is the inverse of $L_{\max}$. It is related to the Kolmogorov-Sinai entropy and the sum of the positive Lyapunov exponents, providing a measure of how quickly nearby trajectories diverge.

$$DIV = \frac{1}{L_{\max}}$$

Longest Vertical Line ($\mathbf{V_{max}}$) The length of the longest vertical line is another indicator of the system's laminar behavior.

$$V_{\max} = \max(\{v_i \mid i = 1, \ldots, N_v\})$$

Entropy (ENTR) The Shannon entropy of the probability distribution $p(l)$ of the diagonal line lengths. It reflects the complexity of the deterministic structure in the system. A higher ENTR indicates a more complex and less periodic dynamics.

$$ENTR = -\sum_{l=l_{\min}}^{N} p(l) \ln p(l), \quad \text{where } p(l) = \frac{P(l)}{\sum_{l=l_{\min}}^{N} P(l)}$$

Trend (TREND) Trend quantifies the paling of the RP towards its edges, which can be caused by non-stationarity in the data (e.g., a slow drift in the mean of the signal). It is calculated as the slope of the linear regression of the local recurrence rate RR_i over the distance from the main diagonal.

$$TREND = \frac{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)(RR_i - \langle RR_i \rangle)}{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)^2}$$

where $\tilde{N}$ is the number of diagonals parallel to the Line of Identity (LOI) that are considered, and RR_i is the recurrence rate in the i-th diagonal.

These metrics, when applied to EEG signals, allow for the characterization of the brain's dynamic states. For example, in an epileptic seizures occurance, often higher determinism (DET), laminarity (LAM) or recurrence rate(RR) is observed compared to the more stochastic and complex inter-ictal states. This fact makes RQA metrics a viable solution for identifying pathological patterns.

6.3 Cross-Recurrence Quantification Analysis (CRQA)

While Recurrence Quantification Analysis (RQA) is powerful for analyzing the dynamics of a single system, many real-world phenomena, including brain activity, involve the interaction between multiple subsystems. Cross-Recurrence Quantification Analysis (CRQA) extends the concepts of RQA to analyze the coupling, synchronization and degree of similarity that the dynamics between two different systems present[**?**].

6.3.1 The Cross-Recurrence Plot (CRP)

The foundation of CRQA is the Cross-Recurrence Plot (CRP). For two reconstructed phase space trajectories $\vec{x}(i)$ from system X and $\vec{y}(j)$ from system Y, both of length N, the cross-recurrence matrix is defined as:

$$CR_{i,j} = \Theta(\varepsilon - \|\vec{x}(i) - \vec{y}(j)\|), \quad i, j = 1, \ldots, N \tag{6}$$

Unlike the standard RP, which is symmetric about the main diagonal (Line of Identity, LOI), the CRP is generally *not symmetric*. This asymmetry can reveal directional relationships or leader-follower dynamics alonside the two systems.

6.3.2 Interpretation of CRQA Metrics

The same quantitative measures defined for RQA (Section 6.2) can be applied to the CRP, but their interpretation shifts from describing *self-similarity* to describing *coupling* and *interaction*:

- **Cross-Recurrence Rate (CRR)**: The probability that the state of system X at time i is close to the state of system Y at time j. A high CRR indicates overall similar states between the two systems.
- **Cross-Determinism (CDET)**: The percentage of recurrent points in the CRP that form diagonal lines. Diagonal lines occur when the two systems follow a similar path in phase space for some time. **This is a crucial metric for epilepsy detection**, as it quantifies the transient synchronization between different brain regions. A seizure often manifests as increased CDET between channels in the epileptogenic zone.
- **Cross-Laminarity (CLAM)**: Measures the laminarity between the two systems, indicating when one system gets trapped in a state while the other changes.
- **Average Diagonal Line Length (L)** in the CRP estimates the mean time that the two systems remain synchronized or follow a similar trajectory.

6.3.3 Why CRQA for Multi-Channel EEG?

Applying CRQA to pairs of EEG channels is particularly well-suited for epilepsy detection for several reasons:

- **Synchronization Detection**: Epileptic seizures are characterized by abnormal, excessive synchronization of neuronal populations. CRQA directly quantifies this synchronization in the phase space.
- **Nonlinear and Non-stationary**: CRQA does not assume linearity or stationarity, making it robust for analyzing the complex, transient dynamics of EEG signals.
- **Directional Insights**: While not explored in all analyses, the potential asymmetry of the CRP can, in principle, help identify the propagation path of a seizure.
- **Focus on Interaction**: It moves beyond analyzing individual channels in isolation to directly measure the dynamic interplay between different brain regions, which is often where the pathology lies.

In this thesis, CRQA is employed to compute a set of features (Table 4) for all unique pairs of EEG channels. These features capture the complex synchronization patterns that distinguish pre-ictal, ictal, and inter-ictal states, forming the basis for the subsequent machine learning classification.

7 Methodology

In this section the methodology for processing EEG data is described in order to perform CRQA to analyze epileptic and non-epileptic brain activity. The approach consists of different parts such as loading and segmenting EEG recordings, extracting non-overlapping time windows, selecting the embedding parameters and computing CRQA features for channel pairs. The methodology is implemented in Python using libraries such as `numpy`, `torch`, `pyopencl`, and `pyrqa` [43].

7.1 Data preprocessing and windowing

Prior filtererd EEG recordings are stored in NumPy array format (`.npy`), accompanied by metadata specifying the sampling frequency (f_s) and channel information (22 channels with their respective labels as `FP1-F7`, `F7-T7`, ..., `FT10-T8`). The time axis is computed as $t = \frac{n}{f_s}$, where n is the sample index and f_s is the sampling frequency in Hertz (Hz).

Each EEG channel is segmented into continuous regions based on predefined boundaries from CHB-MIT dataset [47] annotations, distinguishing epileptic from non-epileptic segments. Then, each segment is further divided into non-overlapping time windows of fixed size (512 samples, equivalent to 2 seconds at 256 Hz). For each segment, the number of windows is calculated by performing integer division of the segment length by the window size and discarding any incomplete windows. Each window is associated with a segment index, window index within the segment, start and end sample indices, and a label (1 for epileptic, 0 for non-epileptic).

7.2 Embedding parameters selection

The embedding dimension `m` and time delay τ were set to 3 and 1, respectively, following common practice in EEG analysis. This choice was also motivated by known limitations that the FNN algorithm experiences when applied to noisy and autocorrelated signals, such as EEG. For instance, [48] have shown that FNN can falsely indicate low-dimensional determinism in autocorrelated stochastic processes, while [49] showed that the effects of noise can actually lead in overestimation of the embedding dimension. In order to mitigate these effects and keep a consist methodology across a large dataset, the presented methodology adopts fixed values for the embedding parameters rather than optimizing them per recording or window. The decision to use constant values is influenced by applied precedents in the EEG literature. McSharry et al.[50] have applied with success fixed embedding parameters in their multi-channel scalp EEG seizure research, arguing that nonlinear methods must justify their complexity over simpler linear benchmarks. In our case, fixed parameters ensure consistency across the large CHB-MIT dataset and help avoid overfitting to local dynamics that may not generalize.

7.3 Threshold selection

Radius fraction R is utilized for determination of the percentage of mean diameter of the reconstructed phase space where a recurrence can occur. In order to estimate and standardize R, an exploration of its effect on CRPs and RR/DET metrics is performed. By keeping constant $\tau = 1$ and $\boldsymbol{m} = 3$ CRPs are generated by selecting random patients and random recording windows of the dataset, while computing the mean channel-wise recurrence rate and mean channel-wise determinism from the 22x22 CRPs features.

Results of RR and DET are presented in the following table. The different explored values for radius fraction are set to be 0.1, 0.15, 0.20 and 0.30 for this experiment.

Table 3: Comparison of RR and DET values for different radius (R) values

Recording	**R**	**RR (%)**	**DET (%)**	**Window**
patient_24	0.10	9	77.36	Normal
patient_24	0.15	16.7	85.3	Normal
patient_24	0.20	23.8	89	Normal
patient_24	0.30	36.4	93.25	Normal
patient_24	0.10	16.35	97.6	Epileptic
patient_24	0.15	26	99	Epileptic
patient_24	0.20	35.16	99.44	Epileptic
patient_24	0.30	51.9	99.74	Epileptic
patient_10	0.10	10.6	83.5	Normal
patient_10	0.15	16.54	86.8	Normal
patient_10	0.20	22.13	88.97	Normal
patient_10	0.30	32.5	92.87	Normal
patient_10	0.10	14.94	93.6	Epileptic
patient_10	0.15	22.9	95.7	Epileptic
patient_10	0.20	30.3	96.41	Epileptic
patient_10	0.30	43.59	97.16	Epileptic
patient_8	0.10	7.08	59.26	Normal
patient_8	0.15	11.82	65.12	Normal
patient_8	0.20	16.33	67.59	Normal
patient_8	0.30	24.76	71.73	Normal
patient_8	0.10	16.4	98.3	Epileptic
patient_8	0.15	25	99.46	Epileptic
patient_8	0.20	33.1	99.60	Epileptic
patient_8	0.30	47.71	99.76	Epileptic
patient_1	0.10	18.31	96.02	Normal
patient_1	0.15	27.83	97.70	Normal
patient_1	0.20	36.61	97.89	Normal
patient_1	0.30	52.08	98.89	Normal
patient_1	0.10	20.89	96.11	Epileptic
patient_1	0.15	33.93	98.46	Epileptic
patient_1	0.20	45.61	99.22	Epileptic
patient_1	0.30	64.64	99.68	Epileptic

As it can be observed, both RR and DET increase while the radius fraction R increases for all patients/windows combinations, since by having a larger radius there are more points to be considered as recurrent in the phase space. Epileptic windows present consistently higher RR and DET values when compared with normal windows at the same R and same patients. It should be noted that there is inter-patient variability also, suggesting that optimal radius selection may benefit from patient-specific tuning. Additionally, DET values in epileptic windows approach saturation near 100% as R increases, a fact that suggests having a moderate radius fraction (e.g., $R = 0.15$–0.20) could provide a better balance between sensitivity and specificity in CRP analysis for the determinism metric.

7.4 Cross Recurrence Quantification Analysis (CRQA)

CRQA quantifies the recurrent patterns between pairs of EEG channels within each time window. The `pyrqa` library is used with OpenCL acceleration for efficient computation. The process is as follows:

1. **Time Series Length Validation**: For each window pair, the lenght's of the two time series are compared on having same length to ensure compatibility.

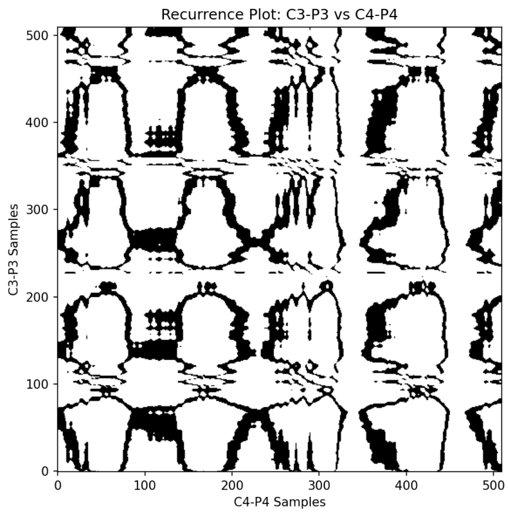
2. **Phase Space Reconstruction**: The time series are embedded into a phase space using the optimal τ and m, via the `TimeSeries` class in `pyrqa`.

3. **Radius Selection**: The radius for defining recurrence is computed using the Phase Space Separation (PSS) method (`pss` function). The maximum distances in the phase spaces of both channels are averaged to obtain a mean diameter, and the radius is set to 15% of this value (`radius_fraction=0.15`).

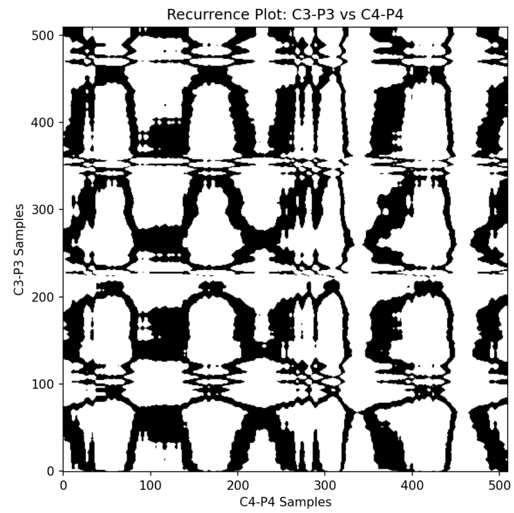
4. **CRQA Computation**: The `RQAComputation` class constructs a cross-recurrence matrix using a `FixedRadius` neighborhood, Euclidean metric, and Theiler corrector of 1. The computation yields 16 CRQA features, listed in Table 4, plus the segment label as the 17th feature.

Table 4: Quantitative measures computed by PyRQA

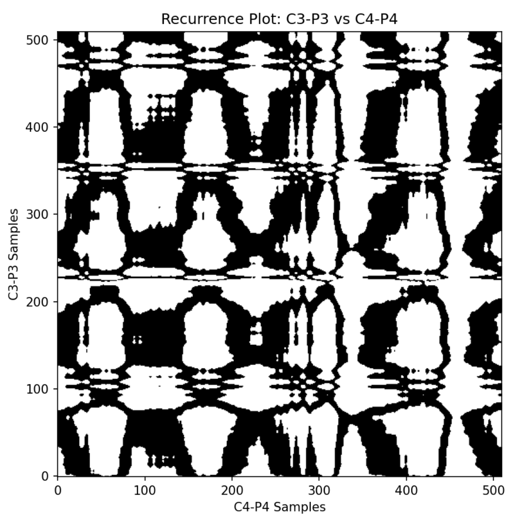
Metric	**Abbreviation**
Recurrence Rate	RR
Determinism	DET
Average Diagonal Line Length	L_{avg}
Longest Diagonal Line Length	L_{max}
Divergence	DIV
Entropy Diagonal Lines	H_{diag}
Laminarity	LAM
Trapping Time	TT
Longest Vertical Line Length	V_{max}
Average White Vertical Line Length	W_{avg}
Longest White Vertical Line Length	W_{max}
Longest White Vertical Line Divergence	W_{max}^{-1}
Entropy Vertical Lines	H_{vert}
Entropy White Vertical Lines	H_{wvert}
Ratio of Determinism to Recurrence Rate	DET/RR
Ratio of Laminarity to Determinism	LAM/DET



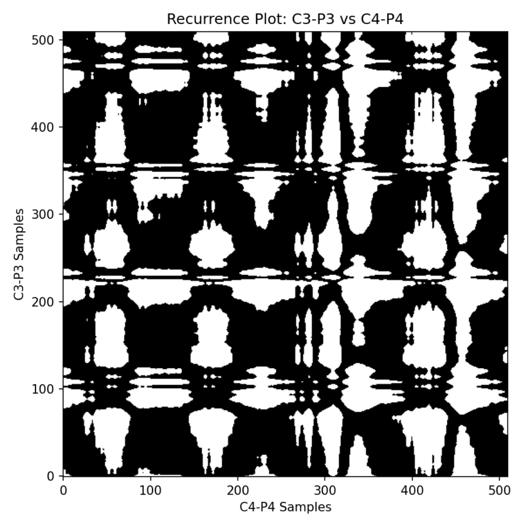
(a) radius fraction = 0.1



(b) radius fraction = 0.15



(c) radius fraction = 0.2



(d) radius fraction = 0.2

Figure 9: CRPs for the selected EEG channel pairs, for an epileptic window. (a) R = 0.1 (b) R = 0.15 (c) R = 0.2 (d) R = 0.3

7.5 Algorithm Summary

The methodology is summarized in Algorithm 1, which outlines the CRQA computation for each window and channel pair.

Algorithm 1 Cross Recurrence Quantification Analysis (CRQA) for EEG Windows

Input: EEG windows $\{X_{c,w}\}$ for channels $c \in C$, windows $w = 1, \ldots, N_w$, where N_w = number of windows, number of electrodes N_e
Output: CRQA feature matrix M of shape $(N_w, N_e, N_e, 17)$
for each window index $w = 1$ to N_w **do**
    for each channel pair $(c_1, c_2) \in C \times C$ **do**
        Time Series Preparation
        Truncate time series $X_{c_1,w}$ and $X_{c_2,w}$ to minimum length
        Set embedding parameters
        Set $\tau = 1$
        Set $m = 3$
        CRQA Computation
        Construct cross-recurrence plot for $(X_{c_1,w}, X_{c_2,w})$ using `PyRQA` with:
            Fixed radius neighborhood $r = 0.15 \times$ mean diameter, Euclidean metric, Theiler corrector = 1
        Extract 16 CRQA features: {RR, DET, L_{avg}, $L_{\max}$, DIV, H_{diag}, LAM, TT, $V_{\max}$, W_{avg}, $W_{\max}$, $W_{\max}^{-1}$, H_{vert}, H_{wvert}, DET/RR, LAM/DET}
        Feature Storage
        Append window label l_w to features
        Store features in $M[w, c_1, c_2, :]$
    end for
end for
Return RQA feature matrix M

7.6 Feature aggregation

The resulting RQA feature matrix has dimensions $[N_w, N_e, N_e, 17]$, where N_w is the number of windows, N_e is the number of channels, and 17 represents the 16 CRQA features plus the segment label(epileptic or normal). To summarize CRQA features across all channel pairs for each window, a mean feature matrix is computed by averaging the 16 CRQA features across all channel pairs, resulting in a final matrix (`mean__feature_matrix`) of shape $[N_w, 17]$, where the last column retains the window label. This matrix summarizes the average dynamical interactions within each window.

8 Class Imbalance in RQA-EEG Data

Real world EEG datasets for epileptic seizure detection are by their nature imbalanced, reflecting the transient nature of seizure's occurence in contrast to normal brain activity. In our aggregated dataset, derived from CRQA features across the multi-channel EEG recordings, the distribution is stark: 98.99% normal segments (89633 examples) against 1.01% epileptic (901 examples), yielding a $\sim$99:1 ratio. This skew is a well known challenge in machine learning which is often reffered as the "imbalanced learning problem".

As discussed by He and Garcia[54], standard classifiers, (f.e, Random Forest), can achieve misleadingly high accuracy ($\sim$99%) by over-predicting the majority class, resulting in poor recall for epileptic events. In RQA based models, this bias can minimize the effects of discriminative patterns, such as the elevated determinism (DET) or laminarity (LAM) in epileptic signals with false negatives posing ethical risks, delaying interventions, while naive oversampling (f.e, duplication) incorporates the risk of overfitting in minority RQA features.

In order to mitigate this, data-level resampling strategies can be employed, focusing on synthetic generation to enrich the minority class without discarding information gained from the majority class.

Table 5: Class distribution in the preprocessed RQA-EEG dataset, highlighting severe imbalance.

Class	Percentage (%)	Count
Normal (0)	98.99	89,633
Epileptic (1)	1.01	901

8.1 Dataset Imbalance

To address the higly imbalanced dataset, three main methods can be employed.

In data level, methods on creating more synthetic samples belonging to the minority class exist, such as Synthetic Minority Over-sampling Technique (SMOTE). With this method it is possible to generate synthetic epileptic samples by interpolating between minority instances and their k-nearest neighbors in feature space [51]. Unlike random duplication, SMOTE promotes diversity, reducing overfitting risks in low-sample regimes.

On the classification algorithm level, specific weights and costs can be used in order to favor the minority class. For instance, many classifiers, such as SVM or tree-based models, allow for the assignment of class weights. These weights are usually set to be inversely proportional to the class frequencies, pushing the model towards paying more attention on errors made on the minority class on the training phase. In the same manner, cost-sensitive learning methods define a higher penalty for misclassifying minority class samples, optimizing for a cost function that reflects the real-world imbalance.

As a third alternative, ensemble methods leverage two or more base models to swift the skew of bias towards the majority class. An ensemble can integrate different classifiers and employ different aggregation strategies for its final classification decision, such as majority vote, weighted votes or stacking.

9 Patient-Specific EEG Classification in Epilepsy

The development of robust and accurate EEG-based seizure detection and prediction systems is a challenging task due to the high variability in EEG signals across different patients. While a universal classifier trained on data from multiple subjects would be the desirable goal for generalization, evidence from studies strongly suggests that patient-specific models offer superior performance and can better suit for real-world clinical and wearable applications.

Hussein et al.[56] conducted an extensive quantitative analysis of intracranial EEG (iEEG) data and concluded that statistical characteristics of both interictal and preictal EEG vary significantly between patients. For instance, researchers observed that the overall range and interquartile range of iEEG sensor readings differed not only between interictal and preictal states but also across patients. They state that *"there is no typical trend in either interictal or preictal data across different epileptic patients"* and that *"the iEEG data of each patient has its own characteristics and its statistical features can solely be meaningful for building a seizure prediction system for this particular patient"*. This kind of inter-patient variability limits the effectiveness of a global model and underlines the necessity of patient-specific approaches.

Similarly, Qiu et al.[57] highlighted that *"the features of the EEG signals vary among different patients"* and that *"even for the same patient, the EEG signals could change substantially when the patient is at different modes, status (awake or asleep), or ages"*. Authors introduced a lightweight deep learning model, called LightSeizureNet, which was evaluated in both patient-independent and patient-specific settings. The patient-specific model achieved higher accuracy than the patient-independent version, achieving 99.77%. In addition, patient-specific models seem to be more computationally efficient for real-time application, such as wearable systems. Furthermore, authors reported that their patient-specific model required only 3.7 million multiply-accumulate operations (MACs), compared to 6.2 million MACs for the patient-independent model, making it a more suitable solution for resource constrained application domains such as a wearable/implantable device.

Further supporting evidence can be found in a recent study by Vijay et al.[58], in which authors perform a systematic comparisson of five supervised machine learning models for both patient-specific and non-patient-specific seizure detection using the CHB-MIT scalp EEG database. Their results demonstrate a clear performance advantage when patient-specific models are used, because of their ability of understanding and learning from personalized seizure patterns.

10 Evaluation

In the pursuit of efficient and robust seizure detection systems, recent advancements emphasize patient-specific EEG channel selection to mitigate noise and redundancy inherent in multi-channel recordings, particularly for wearable applications. For instance, a 2025 study by Ferrara et al[59]. introduces a PCA-based Channel Weight Coefficient (CWC) method that dynamically ranks and selects just two uncorrelated bipolar channels per patient from the standard 21-electrode 10-20 system, leveraging temporal shifts in neural activity between inter-ictal and ictal phases.

mdpi.com

Validated on the CHB-MIT dataset—a benchmark for pediatric epilepsy similar to the one used in this thesis—the approach achieves a balanced accuracy of 0.83 and a false-positive rate of 0.10 per hour, outperforming a global four-channel temporal montage by 30% in false-positive reduction while maintaining comparable sensitivity (0.67). This per-patient optimization not only enhances interpretability through topographic visualizations of channel importance but also reduces model size to 51 KB, facilitating real-time inference on resource-constrained devices. Inspired by this, our methodology incorporates a hybrid bad-value thresholding and PCA sweep to personalize channel dropping, aiming to preserve seizure-relevant RQA features amid data imbalances.

References

[1] Frolov, N., Maksimenko, V. and Hramov, A. (2020), Revealing a multiplex brain network through the analysis of recurrences *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30(12), 121108.

[2] Recurrence quantification analysis of periodic dynamics in the default mode network in first-episode drug-naïve schizophrenia. *Psychiatry Research: Neuroimaging*, 329, 111583.

[3] Recurrence quantification analysis of rs-fMRI data: A method to detect subtle changes in the TgF344-AD rat model. *Computer Methods and Programs in Biomedicine*, 257, 108378.

[4] Lameu, E. L., Yanchuk, S., Macau, E. E. N., Borges, F. S., Iarosz, K. C., Caldas, I. L., Protachevicz, P. R., Borges, R. R., Viana, R. L., Szezech, J. D., Batista, A. M. and Kurths, J. (2018). Recurrence quantification analysis for the identification of burst phase synchronisation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, *28*(8), 085701. https://doi.org/10.1063/1.5024324

[5] Lombardi, A., Guccione, P. and Mascolo, L. (2014). Analysis of fMRI data using the complex systems approach. In *20th IMEKO TC4 International Symposium and 18th International Workshop on ADC Modelling and Testing* (pp. 293-298). Benevento, Italy.

[6] Pitsik, E. (2025). Recurrence quantification analysis and theta-band functional networks detect age-related changes in brain sensorimotor system: VR-based approach. *The European Physical Journal Special Topics*. https://doi.org/10.1140/epjs/s11734-025-01509-y

[7] Guglielmo, G., Wiltshire, T. J. and Louwerse, M. (2022). Training machine learning models to detect group differences in neurophysiological data using recurrence quantification analysis based features. *ICAART 2022 - 14th International Conference on Agents and Artificial Intelligence*, 430-434.

[8] Lopes, M. A., Zhang, J., Krzeminski, D., Hamandi, K., Chen, Q., Livi, L. and Masuda, N. (2020). Recurrence quantification analysis of dynamic brain networks. *European Journal of Neuroscience*, 00, 1-20. https://doi.org/10.1111/ejn.14960

[9] Pentari, A., Tzagkarakis, G., Tsakalides, P., Simos, P., Bertstas, G., Kavroulakis, E., Marias, K., Simos, N. J. and Papadaki, E. (2022). Changes in resting-state functional connectivity in neuropsychiatric lupus: A dynamic approach based on recurrence quantification analysis. *Biomedical Signal Processing and Control, 72*, 103285. https://doi.org/10.1016/j.bspc.2021.103285

[10] Pentari, A., Simos, N., Tzagarakis, G., Kagialis, A., Bertsias, G., Kavroulakis, E., Gratsia, E., Sidiropoulos, P., Boumpas, D. T. and Papadaki, E. (2023). Altered hippocampal connectivity dynamics predicts memory performance in neuropsychiatric lupus: A resting-state fMRI study using cross-recurrence quantification analysis. *Lupus Science & Medicine*, *10*(1), e000920. https://doi.org/10.1136/lupus-2023-000920

[11] Gruszczyńska, I., Mosdorf, R., Sobaniec, P., Żochowska-Sobaniec, M. and Borowska, M. (2019). Epilepsy identification based on EEG signal using RQA method. *Advances in Medical Sciences*, 64(1), 58-64. https://doi.org/10.1016/j.advms.2018.08.003

[12] Mo, J., Zhang, J., Hu, W., Wang, X., Zhao, B., Zhang, K. and Zhang, C. (2022). Neural underpinnings of default mode network on empathy revealed by intracranial stereoelectroencephalography. *Psychiatry and Clinical Neurosciences, 76*(12), 659-666. https://doi.org/10.1111/pcn.13470

[13] Vanithamani Palanisamy, A. Ranichitra, and Radhamani Ellapparaj, A differential biomarker based on recurrence quantification analysis of EEG signal and genetic algorithm for epilepsy diagnosis. *Journal of Artificial Intelligence and System Modelling*, 2(2):74–85, 2024. 10.22034/jaism.2024.450400.1046

[14] Ngamga, E. J., Bialonski, S., Marwan, N., Kurths, J., Geier, C. and Lehnertz, K. (2016). Evaluation of selected recurrence measures in discriminating pre-ictal and inter-ictal periods from epileptic EEG data. *Physics Letters A*, *380*(16), 1419–1425. https://doi.org/10.1016/j.physleta.2016.02.024

[15] Fan, M. and Chou, C.-A. (2019). Detecting abnormal pattern of epileptic seizures via temporal synchronization of EEG signals. *IEEE Transactions on Biomedical Engineering*, 66(3), 601-608. https://doi.org/10.1109/TBME.2018.2850959

[16] P. Núñez, J. Poza, C. Gómez, V. Barroso-García, A. Maturana-Candelas, M. A. Tola-Arribas, M. Cano, and R. Hornero, Characterization of the dynamic behavior of neural activity in Alzheimer's disease: exploring the non-stationarity and recurrence structure of EEG resting-state activity, *Journal of Neural Engineering*, vol. 17, no. 1, p. 016071, 2020, doi: https://doi.org/10.1088/1741-2552/ab71e9.

[17] Yang, C., Luan, G., Liu, Z. and Wang, Q. (2019). Dynamical analysis of epileptic characteristics based on recurrence quantification of SEEG recordings. *Physica A: Statistical Mechanics and its Applications*, *523*, 507–515. https://doi.org/10.1016/j.physa.2019.02.017

[18] Rangaprakash, D. (2014). Connectivity analysis of multichannel EEG signals using recurrence based phase synchronization technique. *Computers in Biology and Medicine, 46*, 11-21. https://doi.org/10.1016/j.compbiomed.2013.10.025

[19] Trulla, L. L., Giuliani, A., Zbilut, J. P., and Webber, C. L., "Recurrence Quantification Analysis of the Logistic Equation with Transients," *Physics Letters A*, vol. 223, no. 4-5, pp. 255–260, 1996. `doi: 10.1016/S0375-9601(96)00741-4`

[20] Webber, C. L. and Zbilut, J. P., "Recurrence Quantification Analysis of Nonlinear Dynamical Systems," in *Tutorials in Contemporary Nonlinear Methods for the Behavioral Sciences*, pp. 26–94, 2005.

[21] Thomasson, N., Hoeppner, T. J., Webber, C. L., and Zbilut, J. P., "Application of Recurrence Quantification Analysis to EEG Signals," *International Journal of Computer Research*, 2002.

[22] Marwan, N., Schinkel, S., and Kurths, J., "Recurrence Plots 25 Years Later – Gaining Confidence in Dynamical Transitions," *EPL (Europhysics Letters)*, vol. 101, no. 2, p. 20007, 2013. `doi: 10.1209/0295-5075/101/20007`

[23] Heunis, T., Aldrich, C., Peters, J. M., Jeste, S. S., Sahin, M., Scheffer, C., and de Vries, P. J., "Recurrence quantification analysis of resting state EEG signals in autism spectrum disorder –

a systematic methodological exploration of technical and demographic confounders in the search for biomarkers," *BMC Medicine*, vol. 16, no. 1, p. 101, 2018.

[24] Eckmann, J.-P., Kamphorst, S. O. and Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhysics Letters*, *4*(9), 973–977.

[25] Timothy, L. T., Krishna, B. M. and Nair, U. (2017). Classification of mild cognitive impairment EEG using combined recurrence and cross recurrence quantification analysis. *International Journal of Psychophysiology*, *120*, 86–95. https://doi.org/10.1016/j.ijpsycho.2017.07.006

[26] Kashyap, A. and Keilholz, S. (2019). Dynamic properties of simulated brain network models and empirical resting-state data. *Network Neuroscience, 3*(2), 405-426. https://doi.org/10.1162/netn_a_00070

[27] Carrubba, S., Frilot, C. and Marino, A. A. (2019). Optimization of recurrence quantification analysis for detecting the presence of multiple sclerosis. *Journal of Medical and Biological Engineering*, *39*(6), 806–815. https://doi.org/10.1007/s40846-019-00462-1

[28] Shalbaf, R., Behnam, H., Sleigh, J. W., Steyn-Ross, D. A., & Steyn-Ross, M. L. (2014). Frontal–temporal synchronization of EEG signals quantified by order patterns cross recurrence analysis during propofol anesthesia. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 22(6), 1164–1174. https://doi.org/10.1109/TNSRE.2014.2350537

[29] Jiang, X., Bian, G.-B., and Tian, Z. (2019). Removal of Artifacts from EEG Signals: A Review. Sensors, 19(5), 987. https://doi.org/10.3390/s19050987

[30] Dhanaseivam, P. S., and Chellam, C. N. (2023). A review on preprocessing of EEG signal. In *2023 9th International Conference on Biosignals, Images and Instrumentation (ICBSII 23)*. Madurai, India.

[31] Jiang, X., Bian, G.-B., and Tian, Z. (2019). Removal of artifacts from EEG signals: A review. *Sensors, 19*(5), 987. https://doi.org/10.3390/s19050987

[32] Sen, D., Mishra, B. B., and Pattnaik, P. K. (2023). A review of the filtering techniques used in EEG signal processing. In *Proceedings of the 7th International Conference on Trends in Electronics and Informatics (ICOEI 2023)* (pp. 270–278). Bhubaneshwar, India.

[33] Mihajlović, D. (2019). EEG Spectra vs Recurrence Features in Understanding Cognitive Effort. *In Proceedings of the 2019 International Symposium on Wearable Computers (ISWC '19)*, September 9–13, 2019, London, United Kingdom. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3341163.3347746

[34] Takens, F. (1981). Detecting strange attractors in turbulence. In *Lecture Notes in Mathematics* (Vol. 898, pp. 366-381). Springer.

[35] Kennel, M. B., Brown, R., & Abarbanel, H. D. I. (1993). Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45(6), 3403–3411.

[36] Fraser, A. M., & Swinney, H. L. (1986). Independent coordinates for strange attractors from mutual information. *Physical Review A, 33*(2), 1134–1140. https://doi.org/10.1103/PhysRevA.33.1134

[37] Marwan, N., Carmen Romano, M., Thiel, M., & Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports, 438(5–6)*, 237–329. https://doi.org/10.1016/j.physrep.2006.11.001

[38] Marwan, N. (2003). Encounters with neighbours: Current developments of concepts based on recurrence plots and their applications (*Doctoral dissertation, University of Potsdam, Germany*). Institut für Physik, Universität Potsdam.

[39] Berényi, A., Kurics, T., Bence, M., Németh, M., Laszlov'szky, T., Nádasdi, M., Ráfi, P., Vincze, V. and Högye, S. *Systems and Methods for Seizure Detection and Closed-Loop Neurostimulation*. U.S. Patent Application No. US 2025/0195894 A1, filed December 13, 2024, published June 19, 2025.

[40] Zhongnan, Z. and Weizhen, L. (2018)*An EEG signal classification model based on genetic algorithm and random forest*, Chinese Patent CN108615024A, Applicant: Xiamen University.

[41] Chen, Y., Hao, H. and Li, L. (2019). *Implantable closed-loop deep brain stimulation system* (Patent No. CN106512206B). State Intellectual Property Office of the People's Republic of China.

[42] Becker, K., Ranft, A., Zieglgänsberger, W., Eder, M. and Dodd, H.-U. (2008). *Monitoring neuronal signals* (U.S. Patent Application No. US20080234597A1). U.S. Patent and Trademark Office.

[43] Rawald, T., Sips, M., Marwan, N. (2017): PyRQA - Conducting Recurrence Quantification Analysis on Very Long Time Series Efficiently. - *Computers and Geosciences*, 104, pp. 101-108.

[44] Norbert Marwan. *Recurrence Plot and Cross Recurrence Plot Quantification* URL: http://www.recurrence-plot.tk/rqa.php, 2025 *Accessed*: 2025.

[45] Gao, X., Yan, X., Gao, P., Gao, X., & Zhang, S. (2020). Automatic detection of epileptic seizure based on approximate entropy, recurrence quantification analysis and convolutional neural networks. *Artificial Intelligence in Medicine*, *102*, 101711. https://doi.org/10.1016/j.artmed.2019.101711

[46] Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P., & Elger, C. E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. Physical Review E, 64(6), 061907.

[47] Guttag, J. (2010). CHB-MIT Scalp EEG Database (version 1.0.0). PhysioNet. RRID:SCR_007345. https://doi.org/10.13026/C2K01R

[48] Fredkin, D.R and Rice, J.A *Method of false nearest neighbors: A cautionary note,* Phys. Rev. E., vol. 51(4), pp. 2950–2954, 1995.

[49] C. Rhodes and M. Morari, *False-nearest-neighbors algorithm and noise-corrupted time series,* Phys. Rev. E., vol. 55(5), pp. 6162–6170, 1997.

[50] McSharry, P.E., He, T., Smith, L.A. and Tarassenko, L. *Linear and non-linear methods for automatic seizure detection in scalp electro-encephalogram recordings.* Med. Biol. Eng. Comput. 40, 447–461 (2002). https://doi.org/10.1007/BF02345078

[51] Chawla, N. V., Bowyer, K. W., Hall, L. O. and Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357. https://doi.org/10.1613/jair.953

[52] He, H., Bai, Y., Garcia, E. A. and Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning, *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, Hong Kong, 2008, pp. 1322-1328, doi: 10.1109/IJCNN.2008.4633969

[53] Devinsky, O., Vezzani, A., O'Brien, T., Jette, N., Scheffer, I.E., Curtis, M. and Perucca, P. Epilepsy.*Nature Reviews Disease Primers* 4, 18024 (2018). https://doi.org/10.1038/nrdp.2018.24

[54] He, H. and Garcia, E. A. . Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263-1284. https://doi.org/10.1109/TKDE.2008.239

[55] Grobbelaar, M., Phadikar, S., Ghaderpour, E., Struck, A.F., Sinha, N., Ghosh, R. and Ahmed, M.Z. I. (2022). A survey on denoising techniques of electroencephalogram signals using wavelet transform. *Signals*, **3**(3), 577–586. https://doi.org/10.3390/signals3030035

[56] Hussein, R., Ahmed, M. O., Ward, R., Wang, Z. J., Kuhlmann, L. and Guo, Y. (2019). Human intracranial EEG quantitative analysis and automatic feature learning for epileptic seizure prediction. arXiv preprint arXiv:1904.03603.

[57] Qiu, S., Wang, W. and Jiao, H. (2023). LightSeizureNet: A lightweight deep learning model for real-time epileptic seizure detection. *IEEE Journal of Biomedical and Health Informatics*, 27(4), 1845–1856.

[58] Vijay, C. P., Mahendra, G., Bharath, K. N., Suhaas, K. P., Kumara, B., Divyashree, R. H. and Sudheesh, K. V. (2025). Supervised machine learning evaluation for patient-specific and non-patient-specific epileptic seizure detection with multichannel scalp EEG. *SN Computer Science*, 6(680).

[59] Ferrara, R., Giaquinto, M., Percannella, G., Rundo, L. and Saggese, A. (2025). Personalizing Seizure Detection for Individual Patients by Optimal Selection of EEG Signals. *Sensors*, 25(9), 2715. https://doi.org/10.3390/s25092715