Degree Project in Computer Science and Engineering

First cycle, 15 credits

Epileptic Seizure Analysis

Asymmetris in Causality using Convergent Cross Mapping

TENZING SANGPO CHOEDON
THERESE LINDH

Epileptic Seizure Analysis: Asymmetries in Causality using Convergent Cross Mapping

TENZIN SANGPO CHOEDON
THERESE LINDH

Master in Computer Science
Date: May 28, 2025
Supervisor: Arvind Kumar
Examiner: Pawel Herman
School of Electrical Engineering and Computer Science
Swedish title: Epileptisk Anfalls Analys - Asymmetrier i Kausalitet med hjälp av Convergent Cross Mapping metoden

Abstract

Exploring brain connectivity is essential to gain a deeper understanding of epileptic seizure dynamics. Particularly, effective connectivity is used for detecting causal relationships between neural regions, with Granger Causality (GC) being a widely recognized measurement. However, identifying the specific EEG channels that are causally involved in epilepsy remains a challenge. Convergent Cross Mapping (CCM) has been developed to address these limitations of GC. Despite the potential of CCM, there have been minimal real applications in epilepsy. Additionally, previous research suggests that there is a stronger directional causal influence on certain brain regions during seizures. Therefore, this study explores causal asymmetries across channels in epilepsy by applying CCM to EEG data. The methodology of this study involves preprocessing the data from the CHB-MIT dataset, passing it into the CCM algorithm, tuning the CCM parameters, and evaluating the resulting causality across non-seizure, preictal and ictal states through asymmetry measures. The results indicate that causality patterns are generally more asymmetric during pre-seizure and seizure activity compared to non-seizure activity. Furthermore, results from individual channels suggest that channels 20 and 21, as well as channels 6 and 12 most consistently exhibit the highest asymmetry in causality for pre-seizure and seizure activity, respectively, which is partially consistent with previous findings. It is concluded that CCM can be applied, with the improvements of scalable methods, to identify potential EEG channels that are important for underlying directional connectivity involved in seizure dynamics.

Sammanfattning

Att utforska hjärnans anslutningar är essentielt för att förstå dynamiken kring epileptiska anfall. I synnerhet effektiv används för att upptäcka orsakssamband mellan olika områden i hjärnan, med Granger Casuality (GC) som en av de mer erkända metoderna. Dock så fortsätter det vara svårt att identifiera de olika EEG kanalerna som orsakar epilepsi. Convergent Cross Mapping (CCM) har utvecklats för att lösa denna brist med GC. Trots potentialen av CCM så har de faktiska applikationerna av CCM inom området epilepsi varit minimala. Dessutom har tidigare forskning antytt en starkare direktionell kausal påverkan på specifika områden i hjärnan under epileptiska anfall. Därmed kommer den här studien att utforska orsakssamband, mer specifikt orsaksasymmetrier, mellan olika EEG kanaler genom att applicera CCM till EEG data. Den här studiens metodik använder datafrån CHB-MIT datasetet, vilken matades genom CCM algoritmen, vilken finjusterades och slutligen evaluerades orsakssamband över föriktal-, iktal- och normalfaserna genom asymmetri. Resultatet indikerade att orsaksmönstren är generellt mer asymmetriska under den föriktala och iktala fasen jämfört med normaltillståndet. Utöver detta indikerade resultaten från de individuella EEG kanalerna att kanal 20 och 21, samt kanal 6 och 12, visade mest konsistent höga nivåer av sambandsasymmmetri i föriktala- och iktalafasen, respektive, vilket är konsistent med tidigare fynd. Slutsatsen är att CCM kan appliceras, givet att skalbara metoder förbättras, i syftet att identifiera potentiella EEG kanaler som är viktiga för underliggande direktionella orsaker involverade i dynamiken kring epileptiska anfall.

Acknowledgements

We acknowledge researcher Satarupa Chakrabarti from the Division of Computational Science and Technology at KTH for providing us with guidance and support throughout the project. Satarupa is conducting a study on causal asymmetry in parkinson's disease, which has been helpful in informing our study. Special thanks to our supervisor Arvind Kumar from the Division of Computational Science and Technology at KTH for offering the opportunity to meet with Satarupa and continuously guiding us.

List of Abbreviations

Abbreviation	Full Terminology
EEG	Electroencephalography
SSR	State-Space Reconstruction
BC	Brain Connectivity
AC	Anatomical Connectivity
FC	Functional Connectivity
EC	Effective Connectivity
CCM	Convergent Cross Mapping

Contents

Chapter 1

Introduction

Epilepsy is a neurological condition that affects around 50 million people globally. This condition is characterized by frequent and repeated seizures, resulting from abnormal electrical activity in the brain [1]. Electroencephalography (EEG) is a method that captures this activity through measuring the changes in electrical potential in the brain. These EEG signals are typically analyzed to detect seizures and are recorded across different locations that are represented by EEG channels [2]. Despite substantial research, there remains aspects to uncover about the neurophysiological mechanisms that cause seizures, particularly the disruptions in neuronal connectivity that induce excessive electrical discharges [3]. Hence, exploring brain connectivity is essential to gain a deeper understanding of epilepsy.

Recently, analyzing brain connectivity has become increasingly common in EEG studies [4]. This can reveal processes and interactions between different brain regions, which is categorized into anatomical, functional and effective connectivity. Among these categories, effective connectivity (EC) is specifically used for detecting causal relationships between neural regions [3]. Although recent studies have concluded regions such as the hippocampus as a source of causal influences in certain cases of epilepsy [3], there remain aspects to be discovered regarding the specific EEG channels that are involved, due to varying criteria and results across studies [29].

A widely recognized method for measuring EC in epilepsy is Granger causality (GC) [3], which is regarded as the primary advance in exploring causality [5]. However, GC is not suitable for complex dynamical systems, such as ecosystems and biophysical systems, that have interacting parts whose states

vary over time [6]. Thus, several methods that are based on GC have been devised to extend the capabilities [4]. In contrast, a new approach called Convergent Cross Mapping (CCM) has been developed to address the limitations of GC. This method has successfully detected casual dependencies in ecosystems [5], but there have been minimal real applications in biophysical systems especially in epilepsy. Furthermore, a GC study found that the directional causal influence on different brain regions becomes stronger during seizures [3], indicating a change from balanced patterns to more asymmetric causality patterns. Therefore, this study explores the causal asymmetries across channels in epilepsy through applying CCM on EEG data.

1.1 Problem definition

Traditional GC has limitations that do not accommodate the complexities in epilepsy EEG data. This method requires linear interaction between variables, which can be problematic due to the non-linear dependencies in EEG signals [3]. Additionally, it does not cover causation in dynamical systems (e.g. biophysical systems) with nonseparable variables and weak to moderate coupling [5]. Despite the advances in GC-based methods, conclusive identification of the specific EEG channels that are causally involved remains a challenge. The CCM method could provide a promising alternative that could address all these limitations.

1.2 Aim and Research Question

The purpose of this study is, therefore, to explore the directional asymmetries in causality across EEG channels using CCM, which could contribute to a better understanding of seizure dynamics. Thus, causal asymmetries could serve as features to distinguish between states in seizure classification problems. This could also potentially depict the set of EEG channels that are important for seizure propagation, which can be used for channel selection in seizure detection algorithms. Therefore the question being investigated is: How can Convergent Cross Mapping (CCM) be applied to explore causal asymmetries across channels in epilepsy using EEG data, and what information can it reveal about seizure dynamics? In order to answer this question, the following sub-questions are studied:

- How can CCM parameter values be optimized in EEG data through convergence analysis, autocorrelation and simplex projection?

- How causal asymmetries differ across non-seizure, pre-ictal and ictal states in EEG data?
- Which specific channels exhibit the highest causal asymmetry during non-seizure, pre-ictal and ictal activity?

1.3 Scope

This paper has included definitions for different brain states related to seizures, however the focus lies in comparing causality results between ictal (seizure), pre-ictal (pre-seizure) and non-seizure state. The causality scores are computed over a limited range of CCM parameter values that are passed into the pre-existing CCM algorithm. Moreover, the scores are not computed across different datasets or patient groups. It is limited to a single dataset (CHB-MIT) that contains real recordings of 24 cases with intractable epileptic seizures, which includes both focal and generalized seizures. Furthermore, this study does not separate and compare the causality results according to seizure type. There are other causality measures that are mentioned such as GC, but this study does not conduct an in-depth comparative analysis of these measures.

1.4 Approach

In this paper, the overall approach consists of three major steps. Firstly, the data is loaded from the CHB-MIT dataset and then preprocessed to eliminate artifacts and noise that can affect the causality scores. The preprocessed data is passed into the CCM algorithm along with varying values on CCM parameters. Finally, the parameter values are optimized and the resulting causality scores are evaluated through asymmetry measures.

Chapter 2

Background

2.1 Epilepsy

Epilepsy today has a wide variety of causes. While some patients are born with it, other cases are caused by disease and/or injury [7]. Epilepsy as a diagnosis is only set after the patient has had two different unprovoked seizure events, or the patient has had one unprovoked seizure with a high risk of remission [8]. While a vast majority of cases respond well to treatment and are seizure free within the first five years of treatment, a large portion of the affected patients never reach this stage [7]. What causes epilepsy also differs between patients, some patients may develop epilepsy due to previous brain trauma, others might have genetic causes.

Further complicating this is the fact that not all seizures are the same[9]; while some are localized to a smaller area of the brain, called a local or focal seizure, others fire in the whole brain at once, called a generalized-onset, or generalized, seizure. A seizure can also start as a focal seizure and later transition into a generalized state[9]. Adding onto this, the seizure itself is defined as three different stages, the preictal, ictal and postictal state, with a fourth state interictal being defined as the time between seizures[2]. The part of the brain where the seizure starts is called the Seizure Onset Zone [10].

2.2 EEG

A common way to help with the diagnosis of epilepsy is through the use of electroencephalography (EEG) [2][8]. This works through placing electrodes on the scalp of the patient; where each electrode records electrical energy in

the brain and outputs it in the form of a timeseries. A typical pattern these electrodes are placed in is called a 10-20 system (see figure 2.1) in which each electrode gets a coordinate and an assigned position on the scalp [2][11]. The data can then be processed using typical signal-processing methods.

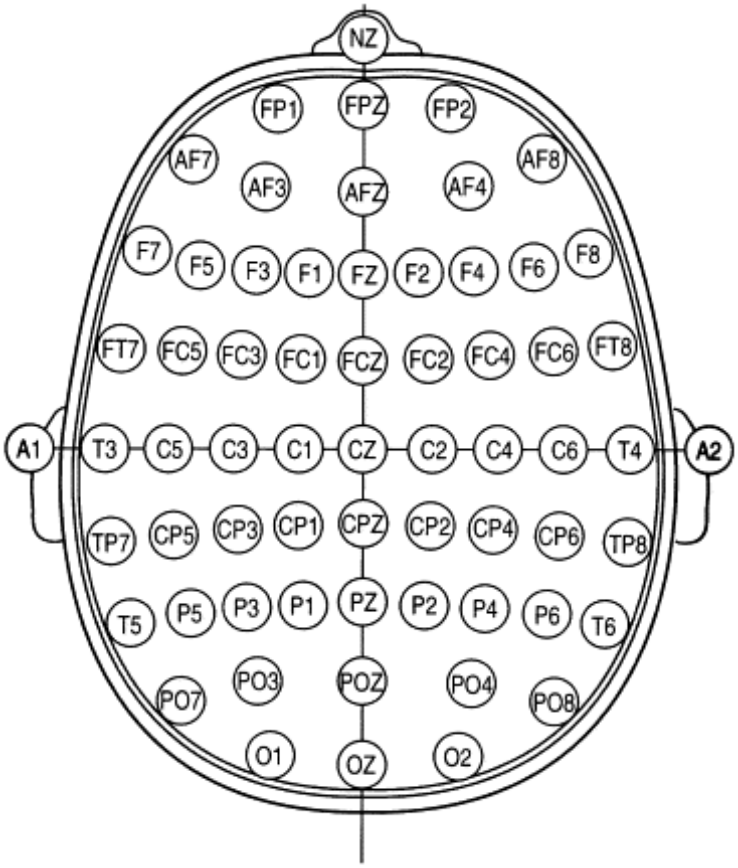

Figure 2.1: Map of the electrodes in a typical 10-20 system [12].

2.3 Brain Connectivity

Brain connectivity, or General connectivity (GC), is an umbrella term for three different forms of connectivity[13]:

- Anatomical connectivity (AC), which creates the white matter in the brain, small fiber tracks which neurons use to communicate between each other.

- Functional connectivity (FC), which is the impact past neuronal activation has on future activations, and reflects statistical dependencies between different regions in the brain.
- Effective connectivity (EC), reflects the casual interactions between different brain regions. It combines FC and AC and effectively describes the effect one region of the brain has on the others, which means that methods like Granger Causality can be used to effectively study brain connectivity[13].

2.4 State-Space Reconstruction

State-Space Reconstruction (SSR) is several different techniques, with Takens' theorem being one of them, used to reconstruct an attractor for a dynamically created system using time-series data[14]. An attractor is defined as a point in space which the system is naturally attracted towards, with curved attractors being called manifolds and depicted with the notation M. Takens' theorem says that manifolds can be reconstructed based on observation of a single variable [6][15].

For example: say that a system is based on the relationship between the three variables X, Y and Z and creates the manifold M. Takens' theorem says that M can be reconstructed as an m dimensional manifold M_x, also known as a shadow manifold, by taking time shifted copies of X so that the state of the system at any time t is given by the equation[15]:

$$M_x = \{X(t), X(t-\tau), \ldots, X(t-(m-1)\tau)\} \tag{2.1}$$

This will, given that a sufficient amount of dimensions are used, preserve essential mathematical properties of M in M_x, such that the system state will map one-to-one between the manifolds[16].

2.5 Convergent Cross Mapping

The method was formulated by Sugihara et al 2012 as an approach that utilizes concepts related to attractors, cross mapping and convergence, in order to inspect causal effects between variables in complex dynamical systems through

studying times series data [6]. The fundamental concept of CCM is that for a dynamical system S: Y = f(X,Y) if X is causally influencing Y then information regarding X is historically embedded in Y, which can be used for predicting values of X [5].

In dynamical systems theory, X and Y has a causal link if they are members of the same dynamic system, meaning if they share a common attractor manifold M_s [5]. However, the true manifold of the system M_s is not known, therefore Takens' embedding theorem is used to compute shadow manifolds of variables X and Y. According to Takens' theorem if X_s, then we can construct a shadow manifold of the true manifold through vectors of lagged (historical) values of X. This results in a shadow manifold M_X that is a projection of M_s, with 1:1 correspondence of the points [6].

Cross mapping is applied to examine if there is a correspondence between the points in the X manifold M_X and in the Y manifold M_Y [6]. If time indices of neighboring points on M_Y can identify points on M_X , then Y can be used to predict X values and vice versa [5]. In order to measure the causality between the variables, the correlation between the predicted values and original values of the time series variable, can be utilized[6].

Convergence is also a key property of the CCM framework [5, 6]. Convergence in CCM means that the estimation of the cross-mapped values improves with the time series length L. The reason is that more data contributes to more data points filling in gaps in the attractor, resulting in closer neighbors and lower estimation error as L increases [5].

The CCM framework mentions different key phenomenon and terminology in causality:

- Coupling: refers to when one variable causally influences the other [6].
- Bidirectional coupling: means the variables are mutually coupled, resulting in cross mapping in both directions [5].
- Unidirectional coupling: equates to when one variable causally influences the other, while the opposite direction has no causal influence [5].
- Strong forcing: pertains to when one variable exerts great deal of influence on the other, while the other variable exerts no influence, which could falsely imply bidirectional coupling (called synchrony) [5].

- Transitivity: is an inherent property in causation where X $\Rightarrow$ Y and Y $\Rightarrow$ Z implies X $\Rightarrow$ Z. CCM enables detection of unidirectional coupling through helping to distinguish variables that are directly coupled from those that only share a driver [5].

2.6 Quantitative Measures and Methods

2.6.1 Asymmetry Index

There is no standardised method for computing the asymmetry index, as there exists varying definitions and formulas across context and discipline. In matrix analysis, a symmetric matrix satisfies A = A^T, where A is the matrix and A^T is its transpose [17]. In this case the difference A - AT is the zero matrix. However, A $\neq A^T$ indicates non-symmetry where the difference A - A^T can be used to quantify the extent of the asymmetry in the matrix. The frobenius norm is applied on this difference to produce a scalar value that summarises the extent of the asymmetry, referred to as Asymmetry Index.

$$AI = ||A - A^T||_F \tag{2.2}$$

where AI denotes the asymmetry index

A is the matrix

A^T is the transpose of A

$||\ldots||_F$ denotes the Frobenius norm, equal to the square root of the sum of the absolute squares of its elements [18]

2.6.2 Autocorrelation

Autocorrelation quantifies the correlation between a time series X_t and the lagged version $Xt - k$ of the same time series where k represents the time lag. The autocorrelation coefficient (k) is calculated by comparing values of X at time t with values of X at time t-k [19].

$$p(k) = \frac{corr(X_t, X_{t-k}}{o(X_t) - o(X_{t-k})} \tag{2.3}$$

where, X_t is a time series variable at time t

k is the time lag

Corr is the correlation value

o is the standard deviation

2.6.3 Simplex Projection

Simplex projections are based on Takens' theorem and the idea that even chaotic time series have some sort of pattern. It predicts future values by looking at past behaviour and uses that to reconstruct a manifold assuming that a sufficiently large dataset is used. This method divides the time series into two halves: X which serves as the library set, and Y which serves as the prediction set [16].

The forecast for a prediction:

$$Y(t_k) = \{Y(t_k), Y(t_k - 1), \ldots, Y(t_k) - E + 1\} \tag{2.4}$$

Where E<len(Prediction set)+1 is given by projecting it's neighbor into that space:

$$X_1, X_2, \ldots, X_{E+1} \tag{2.5}$$

Where:

$$||X_1 - Y(t_k)|| = min(||X - Y(t_k)||) \tag{2.6}$$

For every value of X where X $\neq$ Y applies

This means that X_1 represent the nearest neighbors, X_2 the second closest, X_3 the third closest etc.

The set of E+1 neighboring points will then form a minimal polygon, called a simplex, which encloses the estimated point and the prediction can then be determined by the average value of the projections:

$$X_1, X_2, \ldots, X_{E+1}(t_{e+1} + 1) \tag{2.7}$$

By doing this with different values of E an optimal dimension for the simplex can be found that allows for the most accurate prediction.

2.7 Related Work

Cao [4] reviews and discusses several "EEG-based functional and effective connectivity studies" which were undertaken in the years before the articles publication (2021). His team compared 18 18 methods to quantify the brain connectivity using EEG-data, among them GC. He orders these into two categories, nonparametric and parametric, comparing them within these categories, as well as different ways to visualize the connectivity. This has allowed researchers to analyze and study brain connectivity, which, in combination with machine learning, has allowed researchers to map different brain patterns to emotive states as well as identifying neurological disorders. Cao ends the report by stressing the importance of further research and points out several possible directions of future research[4].

Deyle [9] along with Sugihara introduces two general theorems based on Takens' theorem that expands on the original theorem so that multiple time series can be used to create a manifold, while also proving that Takens' theorem is a special case of these. These theorems could then make State Space Reconstruction applicable to more fields of study, with a primary focus on dynamic systems with weak to moderate correlation, such as ecological, financial and biological[9].

Sugihara [5] proposes an alternative to GC, called CCM, that would enable scientists to use predictability instead of causality in complex systems to detect causation. The major advantage that CCM would have over GC would be accuracy in dynamic systems where the coupling between variables were weak to moderate. As GC checks if the removal of X would majorly negatively affect Y, the weak to moderate causations would go unnoticed, while CCM tries to reconstruct X from Y and Y from X, and would have an easier time picking up on the smaller correlations. This makes CCM more suitable in systems in which correlation between time-series variables may not always be obvious or true, and allows for a deeper understanding of dynamic systems like ecosystems [5].

Schiecke [20] looks at the potentials and limitations of using CCM to investigate nonlinear causality in a dynamic system. They do this by testing both simulated data, with various degrees of simulated noise added, and real epilepsy data taken from the University Hospital Vienna. The results of this are divided into two parts, with the simulated data showing that CCM becomes more ac-

curate with low levels of noise and a lower level of embedding dimensions, but at higher noise levels the importance of more embedded dimensions becomes apparent. The real life data shows definite signs of correlation between different electrodes, with a higher library length giving a higher correlation coefficient, while at the same time showing a sudden rise in value during ictal and postictal periods. Schiecke ends the paper with noticing that the nonlinear approach of CCM is an advantage, but the fact that it can only check two variables, $X \Rightarrow Y$ and $Y \Rightarrow X$ is a definite disadvantage[20].

Chapter 3

Method

The proposed method in this research, for analysing the causal dependencies in epilepsy EEG data consists of the following steps: preparing the data, implementing the CCM framework and evaluating the causality scores. The workflow of the proposed method is illustrated in figure 3.1.

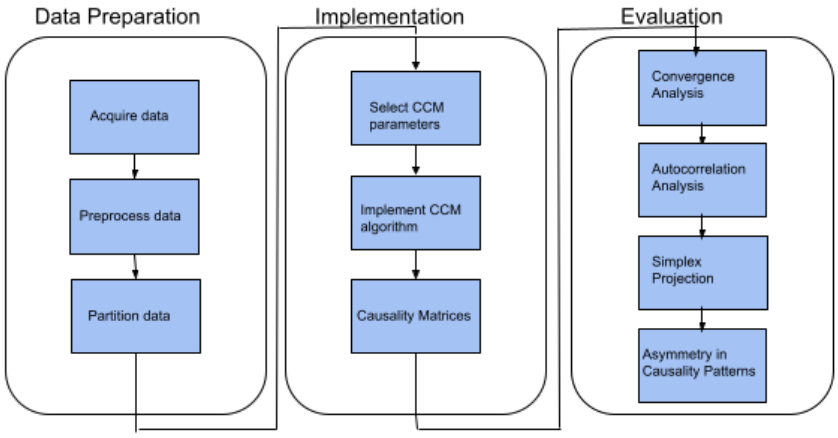

Figure 3.1: Overview of the proposed workflow, which is sectioned into three phases with substeps.

3.1 Data Preparation

This section presents the pipeline to prepare the data for subsequent implementation, which includes data acquisition, preprocessing and partitioning.

3.1.1 Data

In this study, the CHB-MIT dataset related to epilepsy was used. This dataset contains 982 hours of EEG time series recordings from 23 pediatric subjects with intractable seizures. The EEG signals are captured at a sampling rate of 256 Hz across a minimum of 23 channels. The recorded data consists of 24 cases where case chb01 and chb21 was obtained from the same subject 1.5 years apart [21][22]. The data from the 23 cases were collected at the Children's Hospital Boston and one case was collected at Beth Israel Deaconess Medical Center [23]. The data registered a total of 198 seizures which were confirmed as clinical seizures by experts [22]. For each case, the experts annotated the seizure onset and offset times. The data was stored in one hour long segments that were categorized into non-seizure records and seizure records to indicate whether the record contains a seizure [23]. The EEG channels utilized bipolar reference to measure the potential differences for pairs of electrodes [22][24]. This can provide better artifact rejection and more precise spatial localization compared to referential montages [24]. The dataset that was used in this paper is accessed via PhysioNet website [21].

Subjects chb01, chb02, chb03, chb04, chb05, chb06, chb08, chb09, chb10 and chb23 were selected primarily due to consistent data availability of channel number $1, \ldots, 23$. These 23 EEG channels (FP1–F7, F7–T7, T7–P7, P7–O1, FP1–F3, F3–C3, C3–P3, P3–O1, FP2–F4, F4–C4, C4–P4, P4–O2, FP2–F8, F8–T8, T8–P8, P8–O2, FZ–CZ, CZ–PZ, P7–T7, T7–FT9, FT9–FT10, FT10–T8, T8–P8) were selected for the analysis because they are the minimum common set of channels among all subjects. These subjects have a sufficient number of seizures for the analysis, as shown in Table 3.1.

subject_id	seizures
chb01	7
chb02	3
chb03	7
chb04	4
chb05	5
chb06	9
chb08	5
chb09	3
chb10	6
chb23	7

Table 3.1: Number of seizures per subject

3.1.2 Data Preprocessing

Preprocessing is necessary for preparing the EEG data for further analysis. This step involves reducing artifacts and noise in the EEG data to improve the accuracy of epilepsy seizure detection [25]. In EEG signals, noise can arise due to inherent physical and electrical interferences, while artifacts arise due to external sources such as eye and muscle movements, disturbances of the environment and movement of EEG sensors [11]. Preprocessing techniques, specifically filtering, were applied on EEG data before passing it to the CCM algorithm. The EEG records were preprocessed primarily through the MNE python package [26].

Filtering is used to suppress frequency components outside of the cutoff values. The purpose is to remove low frequency drifts (artifacts) and high frequency noise [27]. We applied band-pass filtering on the segmented data using the default filter method in MNE. The default method is a zero-phase FIR (finite filter response) filter, which is stable and sufficient for electrophysiological data analysis [28]. The range of frequencies that were selected for the pass-band was 0.5 - 40 Hz since this frequency range contains useful EEG signals (below 40 Hz) and no motion or electrical artifacts (above 0.5 Hz) [11]. This allows frequencies within the specified range to pass through while lower frequencies and higher frequencies that are outside of these bounds are attenuated [27].

3.1.3 Data Partitioning

In epilepsy seizure analysis, the approach is primarily categorized into: patient-dependent, patient-independent and cross-patient models. This study implemented a patient-specific approach for the analysis which is standard practice due to the inter-individual variability in EEG patterns [29]. This ensures that personalized characteristics such as age, brain anatomy, medication are accounted for, therefore reducing the risk of bias being introduced. The preprocessed data was segmented into ictal and non-seizure samples for each patient. Additionally, the 30 second window preceding each seizure was segmented into pre-ictal samples. This 30-second window is a standardised length in EEG analysis that effectively captures the potential alterations in brain connectivity during the pre-ictal phase [30]. These samples were used to construct patient-specific files: *control data*, corresponding to a selected non-seizure sample; *pre-ictal data*, consisting of a set of pre-seizure samples; and *ictal data*, encompassing a set of seizure samples. The causality scores of these files were then computed during the implementation phase for the analysis.

3.2 Implementation

This portion of the paper outlines the implementation of the CCM framework, which involves selecting the CCM parameter values and implementing the algorithm.

3.2.1 Selection of CCM Parameters

In order to improve the accuracy of the causality scores, we computed them using varying values of the CCM parameters. The parameters are L, E and τ which represent library length, embedding dimensions and the time lag step. The ranges of these parameter values are shown in Table 3.2. The selected ranges were chosen to explore a comprehensive space while maintaining computational feasibility. The range of L values were determined based on a preliminary test of convergence in causality values for a single channel pair of a time series. We computed the causality matrix for each combination of parameter values (L, E, τ). The causality matrix contains causality scores for all channel pairs in a $N \times N$ matrix, where N is the number of channels. After computing the necessary scores, the most appropriate values for the CCM parameters were identified during the evaluation phase.

Table 3.2: Parameter ranges explored during the CCM analysis.

Parameter	**Values**
L	$\{6000,\ 7000,\ 8000,\ 9000,\ 10000\}$
E	$\{2,\ 3,\ 4,\ 5\}$
τ	$\{1,\ 2,\ 3,\ \ldots,\ 10\}$

3.2.2 Description of CCM Algoritm

This study implemented the CCM algorithm developed by Sugihara et al., which is available on GitHub [6]. The algorithm contains the following steps, assuming that the time-series variables X and Y are given, which are defined in Equations 3.1 and 3.2.

$$X = \{X(1), X(2), \ldots, X(L)\} \tag{3.1}$$
$$Y = \{Y(1), Y(2), \ldots, Y(L)\} \tag{3.2}$$

1. Construct the shadow manifold M_X through computing vectors $x(t)$ of lagged coordinates from time series X (see Equation 3.3).

$$M_X = \{x(t) \mid t \in [1 + (E-1)\tau, L]\} \tag{3.3}$$
$$x(t) = \langle X(t), X(t-\tau), X(t-2\tau), \ldots, X(t-(E-1)\tau)\rangle \tag{3.4}$$

2. Find the $E+1$ nearest neighbors of the selected $x(t)$ in M_X, and mark them at time indices $t_1, \ldots, t_{E+1}$, in order to identify the corresponding points in Y. The $E+1$ neighbors is the minimum number of points required for an embedding with E dimensions.

3. Generate the weights w_i using the corresponding neighbor vector $x(t_i)$ (see Equation 3.5), through computing the euclidean distance between the selected $x(t)$ and neighbor $x(t_i)$ and scaling to $d(x(t), x(t_1))$ as multiples to the closest point. The greater distance for the neighbor, the lower the weight by $exp(-k)$.

$$u_i = \exp\left[-\frac{d(x(t), x(t_i))}{d(x(t), x(t_1))}\right], \quad w_i = \frac{u_i}{\sum_{j=1}^{E+1} u_j} \tag{3.5}$$

4. The predictive model $\hat{Y}|M_X$ is defined using the generated weights w_i for the corresponding $Y(t_i)$ (see Formula 3.6).

$$\hat{Y}|M_X = \sum_{i=1}^{E+1} w_i Y(t_i) \tag{3.6}$$

5. Compute the correlation coefficient between Y and $\hat{Y}|M_X$ to see how much information from Y is stored in X. If the correlation is strong, we say Y causally influences X.

6. Generate $\hat{Y}|M_X$ for different values of L, since it should converge to Y as L increases. If X and Y are dynamically coupled, more data points lead to denser $E+1$ points in M_X, corresponding to a denser $E+1$ cluster in M_Y. The resulting correlation between Y and $\hat{Y}|M_X$ should stabilize as L increases.

3.2.3 Implementation of CCM

The ccm algorithm was implemented to quantify the causal influence between EEG channel pairs (see Algorithm 1). The data segment was selected using a fixed length window $start\_index : end\_index$ to ensure consistency across all pairs. This window was selected using a randomized approach, which ensures variation by selecting different starting points for each data file, including a wider range of EEG patterns in the subsequent analysis. The starting point does not include the beginning of the EEG recording to avoid potential artifacts that can occur there. The CCM was applied in both directions to explore the extent of X causally influencing Y and Y causally influencing X. The upper triangle of the causality matrix, denoted as $C_{X\to Y}$, stores the causality value from $X \to Y$ and the lower triangle of the causality matrix, denoted as $C_{Y\to X}$, stores the causality value from $Y \to X$. This ensures that both directions for each pair are stored in the matrix, enabling further analysis of the asymmetries in directionality. The resulting matrix C contains entries $C_{i,j}$ that represent the causality score from channel i to j (see formula 3.7), facilitating visualization and analysis tasks such as plotting heatmaps and computing asymmetry index values.

$$C = \begin{bmatrix} C_{1,1} & C_{1,2} & \cdots & C_{1,N} \\ C_{2,1} & C_{2,2} & \cdots & C_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ C_{N,1} & C_{N,2} & \cdots & C_{N,N} \end{bmatrix} \tag{3.7}$$

3.3 Evaluation

This section covers the pipeline for selecting the appropriate values of CCM parameters using convergence analysis, autocorrelation and simplex projection, as well as evaluating the causality scores through computing the asym-

metry indices of causality matrices across states and asymmetry of individual channels for channel selection.

3.3.1 Parameter Testing

The parameter testing was performed using the subject chb01 due to computational constraints. Moreover, tuning the parameters on a single subject is more feasible than scaling the process across multiple subjects for CCM analysis. Further limitations were introduced for parameter testing due to computational constraints. The causality matrices used for the parameter tuning were derived from the control data file because it represents non-seizure activity, which serves as a reference. Additionally, the parameter testing was limited to a subset of channels such as: 2, 4, 6, 7. In a comprehensive review [29], Faul and Marnane reported that these channels demonstrated high seizure probability and seizure detection performance using probability-based filtering.

Convergence analysis is an integral component of the CCM framework [5]. The convergence analysis was conducted on control data. The purpose is to identify the library length (L) that produces the most stable and convergent approximation of causality score [5]. Each parameter set (L, E, τ) has a corresponding causality matrix with scores for all channel pairs. In order to study overall convergence, the mean causality score was computed for each matrix. These mean values were plotted across L for each combination of (E, τ) (see Algorithm 2). The appropriate L^* was selected based on the convergence in the line plot and was used for all subsequent analysis.

Autocorrelation is an essential concept of time series analysis. This concept measures the correlation of the variable at different time lags [19]. In order to find the appropriate τ, the first zero-crossing of the autocorrelation function was observed because it indicates the minimal time lag that ensures maximum linear independence among the coordinates of the embedding vector [31]. This analysis was performed for the control file to identify the optimal time lag value. The causality matrices for each combination of parameter values (L^*, E, τ) were used. To inspect the overall autocorrelation, the mean causality score was computed for each matrix. The mean values were passed into the autocorrelation function from the statsmodels python module [32]. The resulting autocorrelation values were plotted across τ for each E (see Algorithm 3). The appropriate τ^* was selected based on the first zero-crossing in the line plot and was used for all subsequent analysis.

Simplex projection is commonly used to find optimal embedding dimension (E) [16]. This method was performed for each channel, where we computed a simplex object for each E through pyEDM python package [33]. This simplex object used τ^* for the state reconstruction and L^* data points for training the embedding and the remaining L^* data points to test predictions. The resulting simplex object has the observation and prediction values, which were used to compute the pearson correlation values to measure the predictive skill [16]. The mean correlation value was selected for representing that specific embedding dimension E. The optimal E for that channel was selected through identifying the E with the best prediction skill i.e. highest mean correlation value (see Algorithm 4). After identifying the optimal E values for each channel, the overall appropriate E* was selected as the most frequently occurring E among the channels and this was used for all subsequent analysis.

To assess whether the parameters (L^*, E^*, τ^*) are appropriate for the remaining files, we performed convergence checks for each state per selected subject in Table 3.1. Thus the convergence was examined for all patient-specific files of these subjects to determine if it produces stable causality values for a representing channel pair at (L^*, E^*, τ^*).

3.3.2 Asymmetries in Causality Patterns

The causality matrices with parameter values (L^*, E^*, τ^*) were computed for all the patient-specific files, including the control data, ictal data and pre-ictal data of the subset of subjects. Thus the subjects had a single causality matrix for each state: non-seizure, pre-ictal, ictal. This matrix contained causality scores between 23 EEG channels that we selected for the analysis. This contrasts with the parameter testing, which used a limited number of channels. These causality matrices were also plotted as heatmaps to visualize the strength of causal relationships and differences in directionality. These causality patterns typically encompasses inherent asymmetries in causal influences due to the directional nature of causal relationships.

The asymmetries in the resulting causality matrices with (L^*, E^*, τ^*) was investigated through computing the asymmetry index for each causality matrix. To examine how these asymmetries vary across the different states for all subjects, the distribution of the asymmetry index values were visualized across states through a boxplot. In order to explore the sources that are driving these

asymmetry patterns, the asymmetry value of each EEG channel was measured. The asymmetry value of an arbitrary channel X in a causality matrix was computed through calculating the absolute value of the difference between total causal outflow from the row, where X is the driving variable, and total causal inflow from the column, where X is the response variable. The channel exhibiting the highest asymmetry of each causality matrix was identified. To examine whether these channels are consistently dominant in each states for all subjects, horizontal bar charts were created for each state, displaying the top three most frequently occurring channels with the highest asymmetry value.

Chapter 4

Results

The following section demonstrates the results from the parameter testing to find the most optimal CCM parameter values.

4.1 Parameter Tuning

The results of the parameter testing is comprised of Figures 4.1, 4.2, 4.3. The convergence analysis is based on figure 4.1, which shows the mean causality scores with different (E, τ) values for the control data across the L values. The solid line plots represents the mean causality values from $X \rightarrow Y$, while the dotted lines represents the mean causality values from $Y \rightarrow X$. The causality values stabilizes close to $L = 10000$, indicating that $L^* = 10000$ is the appropriate library length.

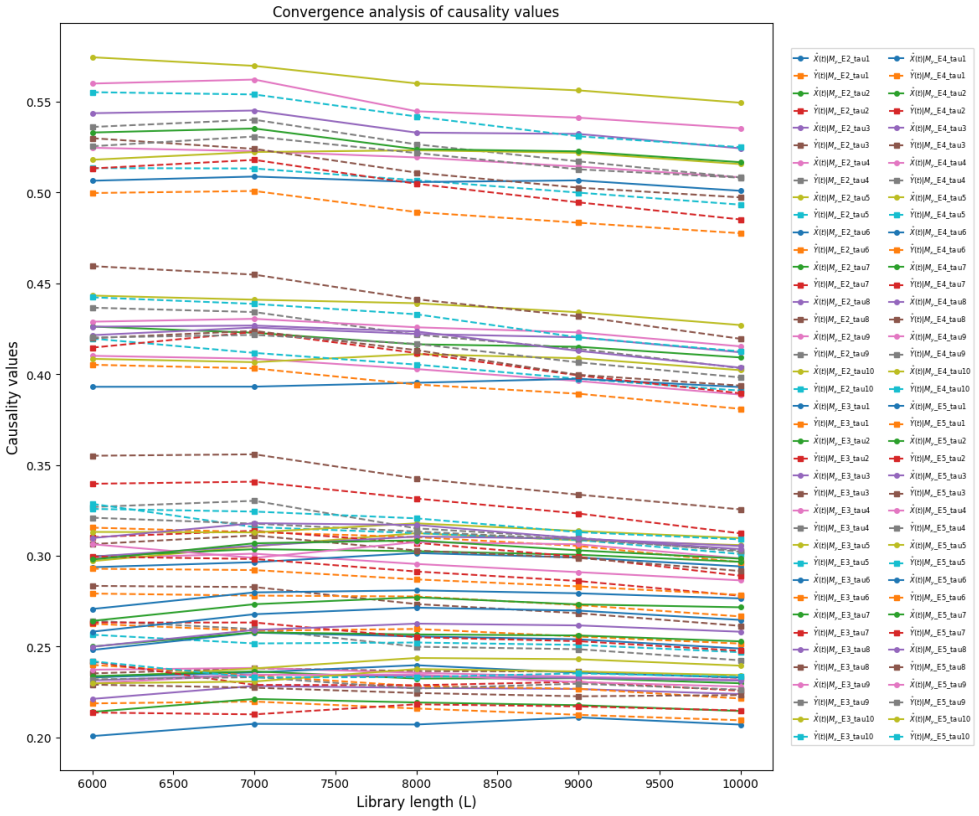

Figure 4.1: Convergence behavior of causality values with various parameter combinations (L, E, τ) for the control data

Figure 4.2 displays the results of the autocorrelation analysis. The lines corresponds to autocorrelation values of the mean causality scores with various parameter combinations (E, τ) and $L^* = 10000$ for the control data. The solid line plots represent the mean causality values from $X \to Y$, while the dotted lines represents the mean causality values from $Y \to X$. The first zero-crossing is observed at $\tau = 3$ and also $\tau = 4$, which is the most frequently occurring value at the first zero-crossing. Thus $\tau^* = 4$ is the appropriate time lag.

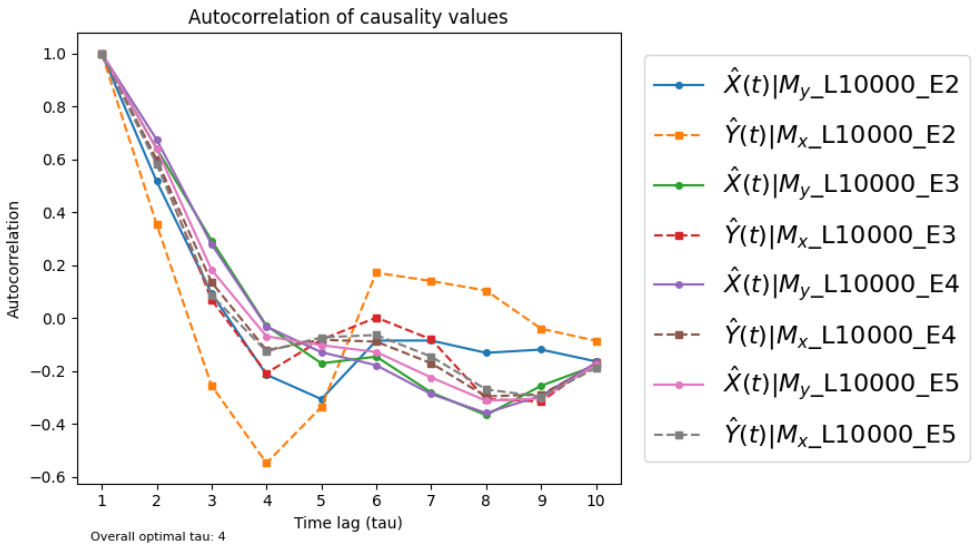

Figure 4.2: Autocorrelation of causality values with various parameter combinations (E, τ) and $L = 10000$ for the control data

In figure 4.3 the results of the simplex projection is shown. The lines represents the prediction skill of the state space reconstruction with various E values, $L^* = 10000$ and $\tau^* = 4$ of control data from channels 2, 4, 6, 7. The highest correlation value for each channel is observed at $E = 2$, $E = 5$ and also $E = 4$, which is the most frequently occurring E value as the highest correlation i.e. best prediction skill. Thus $E^* = 4$ is the appropriate embedding dimension.

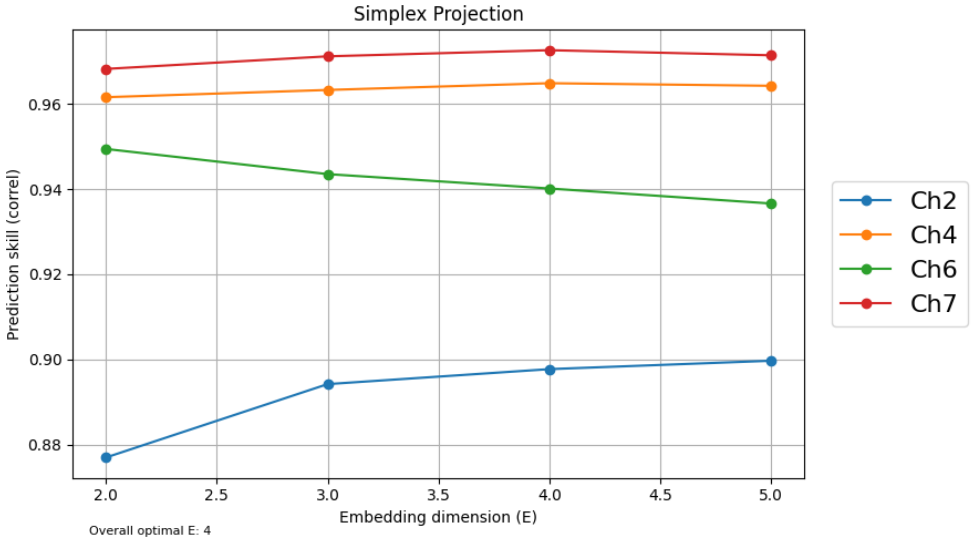

Figure 4.3: Simplex Projection with various E values, L=10 000, τ=4 of control data from limited channels

The convergence of all data of the selected subjects did not consistently stabilize at $L^* = 10000$, $E^* = 4$ and $\tau^* = 4$. Figure 4.4 shows the convergence behaviour for chb01 of a representing channel pair (1,2) as an example of non stabilizing values. However, individual parameter tuning would be time-consuming, thus these parameter values were selected for the entire analysis.

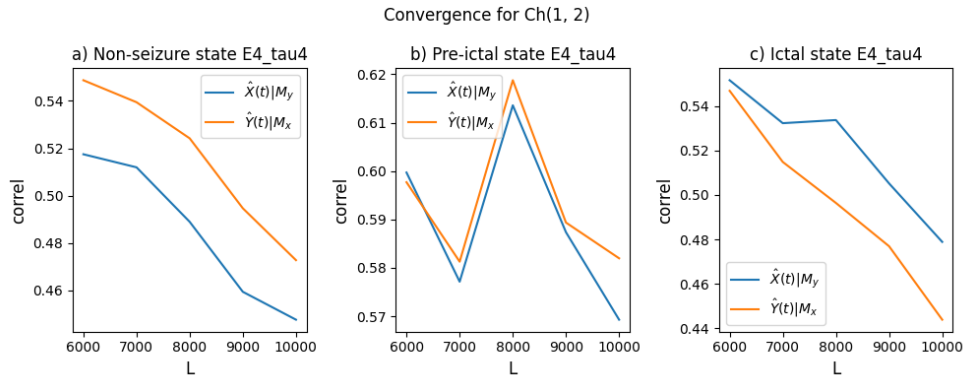

Figure 4.4: Convergence behaviour of different states across L for chb01 with $E = 4$ and $\tau = 4$.

4.2 Asymmetry Index Across States

In this section the asymmetric behavior of the resulting causality patterns for $L^* = 10000$, $E^* = 4$ and $\tau^* = 4$ are presented. In figure 4.5 heatmaps of causality matrices with 23 EEG channels of control, preictal and ictal data of chb03 are shown as an example of asymmetric behaviour across states. The causality patterns of non-seizure data, as seen in 4.5a) are strongly symmetric compared to the other states in 4.5b) and 4.5c).

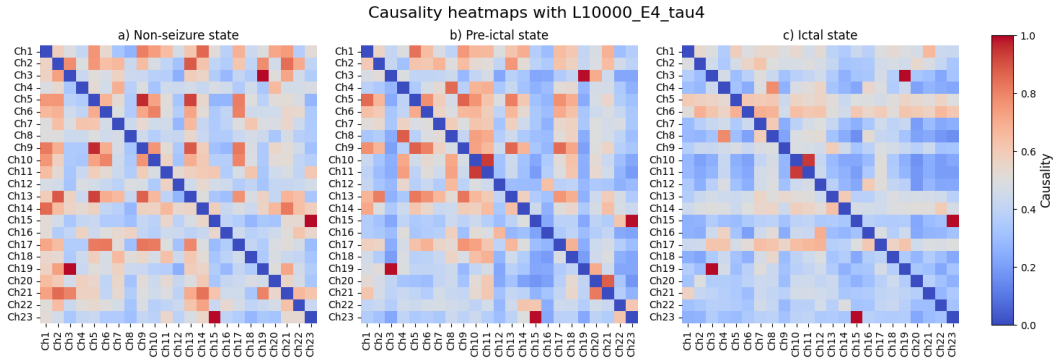

Figure 4.5: Example Heatmaps of Causality Matrix for control, preictal, and ictal state of chb03 with L=10 000, E=4, τ=4.

To quantify the asymmetry in these causality matrices of the selected subjects, the asymmetry index values were calculated and compiled in Table 4.1. This table shows that the asymmetry index of ictal and preictal is higher compared to the control data for chb03, which is also illustrated in figure 4.5.

subject_id	control	preictal	ictal
chb01	1.47	1.21	1.90
chb02	1.61	2.44	1.21
chb03	1.68	2.83	3.66
chb04	1.61	1.70	1.71
chb05	1.55	1.70	1.70
chb06	1.67	1.10	1.03
chb08	1.07	2.01	1.12
chb09	1.17	0.92	1.11
chb10	1.40	1.64	2.07
chb23	1.35	1.89	1.61

Table 4.1: Subject-specific Asymmetry Index for control, preictal, and ictal state with L=10 000, E=4, τ=4.

The distribution of these asymmetry index values in table 4.1 is displayed in figure 4.6 grouped according to states. The asymmetry index values computed from the preictal and ictal data is higher compared to control data for 70% of the selected subjects respectively. Moreover, the causality of control data have less variability in asymmetry index values compared to preictal and ictal states, according to the figure 4.6. Lastly, the asymmetry index values derived from the causality of preictal data is higher than ictal data for 50% of the selected subjects.

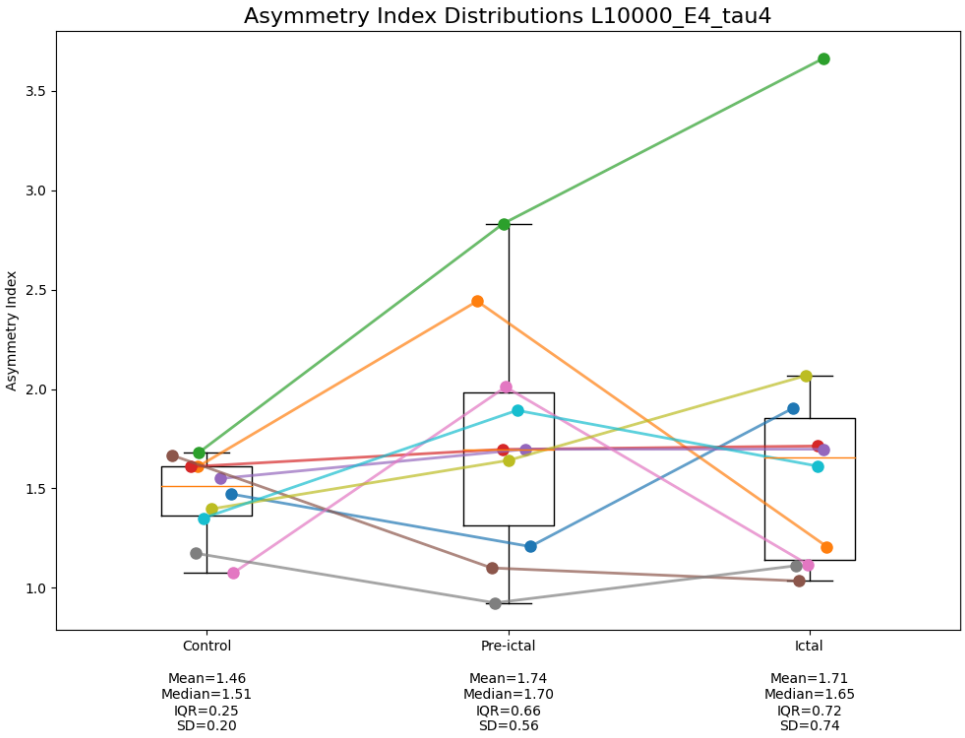

Figure 4.6: Asymmetry Index Distribution across Control, Pre-ictal and Ictal data with L=10 000, E=4, τ=4.

4.3 Highest Asymmetry Channel Across States

To further investigate the primary sources of the observed asymmetry patterns, the asymmetry values of the individual EEG channels were analyzed. The channel with the highest asymmetry of each state per subject is shown in table 4.2. Channel 6 in the the ictal state for chb03 demonstrates the highest asymmetry among the states and subjects. These channels were used for frequency analysis in figure 4.7.

subject_id	control	preictal	ictal
chb01	Ch(11): 1.063	Ch(20): 1.578	Ch(12): 3.172
chb02	Ch(21): 1.407	Ch(22): 2.637	Ch(6): 1.509
chb03	Ch(16): 1.985	Ch(18): 3.306	Ch(6): 5.028
chb04	Ch(8): 1.611	Ch(5): 1.693	Ch(15): 1.94
chb05	Ch(21): 2.259	Ch(21): 1.897	Ch(6): 2.012
chb06	Ch(22): 1.449	Ch(13): 1.233	Ch(1): 1.21
chb08	Ch(2): 1.13	Ch(3): 2.092	Ch(9): 1.147
chb09	Ch(22): 1.135	Ch(20): 0.928	Ch(13): 1.63
chb10	Ch(22): 1.568	Ch(21): 2.1	Ch(14): 2.56
chb23	Ch(17): 1.336	Ch(12): 2.4	Ch(12): 2.322

Table 4.2: Subject-specific Channels with the Highest Asymmetry in control, preictal, and ictal data for subjects.

The channels in table 4.2 were examined to determine whether they consistently show dominant asymmetry across the selected subjects. The frequency of the top asymmetric channels is illustrated as bar charts in figure 4.7. This figure is restricted to the three channels that is most frequently occurring among the set of top-ranking asymmetric channels. Channel 6 is the most frequently occurring for the ictal state across the selected subjects with an occurance rate of 30%. The remaining channels in figure 4.7 has a frequency of 2 or 1 which corresponds to 20% or 10% respectively.

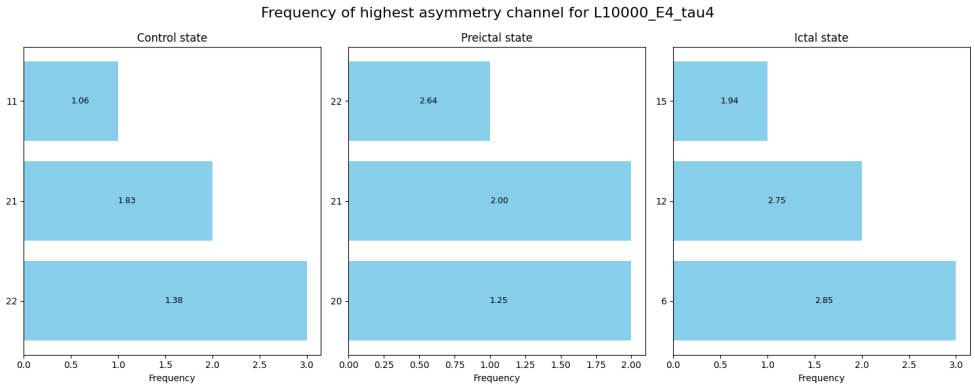

Figure 4.7: Frequency and the Mean of Channel with the Highest Asymmetry of Control, Pre-ictal and Ictal states with L=10 000, E=4, τ=4.

Chapter 5

Discussion

This section discusses the findings of the study in light of the research question: **How can CCM be applied to explore causal asymmetries across channels in epilepsy using EEG data, and what information can it reveal about seizure dynamics?** The results are analyzed in the context of existing work and limitations in our study, followed by considerations of improvements and implications.

5.1 Analysis

An observation is that the $L^* = 10000$, $E^* = 4$ and $\tau^* = 4$ are the optimal parameters of control data from chb01, which can be seen in figures 4.1, 4.2 and 4.3. However, the causality values of the selected subjects did not consistently stabilize at (L^*, E^*, τ^*) (see figure 4.4), which may be a consequence of suboptimal hyperparameter tuning.

The asymmetry indices of causality from the ictal data are higher compared to control data in 70% of the instances shown in figure 4.6. This suggests that, in most cases, directional differences in causal influences increase during seizures. This is consistent with findings from a GC study [3], which revealed increased directional causality in certain brain regions during seizures. Similarly, the asymmetry indices derived from the causality of preictal data are higher than those of the control data in 70% of cases. This may indicate stronger directional biases in causal influences closer to seizure onset. This aligns with the results of [34] that demonstrated an increase in outgoing network connectivity during preictal activity. Moreover, the asymmetry indices computed from preictal data are higher than ictal data for 50% of the selected

subjects in figure 4.6. This may imply that, in some cases, directional biases in causality are more pronounced during the preictal phase than ictal phase. We theorize that this is due to directional connectivity peaking at seizure onset, and then partially stabilizing during the ictal phase. Additionally, the variability in asymmetry indices is also stronger in the pre-ictal and ictal states compared to the non–seizure state, which can stem from hyperparameters only being tuned to control data of chb01.

The channels that are asymmetrically dominant for the non-seizure state are channel 22 (FT10–T8) and channel 21 (FT9–FT10) with frequency rates of 30% and 20%, respectively, according to figure 4.7. The mean asymmetry of channel 22 and 21 increases from 1.38 to 2.64 and 1.83 to 2.00, respectively, in the preictal state, indicating heightened asymmetric causality in the temporal lobe region (FT9, T7, FT10, T8) preceding seizures. This could be related to temporal lobe epilepsy; however, the seizure onset zone for temporal lobe epilepsy has been reported to be located in the hippocampus or entorhinal cortex, which has shown considerable increase in synchronization during the preictal phase [35]. Moreover, channel 21 has 20% occurrence rate among the top asymmetric channels in preictal state, which may be linked to previous results identifying FT9–FT10 according to Duun-Henriksen channel selection method, as it showed maximum variance during seizures [36]. Additionally, channel 20 (T7–FT9) was asymmetrically dominant in the preictal state for 20% of subjects, which has been previously demonstrated to be involved with ictal discharges for 98% of the patients in another study [37]. Channels 20 and 21 exhibit the highest asymmetry in 20% of subjects, indicating that they could potentially be important for underlying connectivity involved in seizure initiation.

In contrast, channel 6 (F3-C3) and channel 12 (P4-O2) are the most consistently dominant asymmetric channels in the ictal state, with frequency rates of 30% and 20%, respectively. Particularly, channel 6 for chb03 during ictal activity achieved the highest asymmetry across the states and subjects, which is shown in table 4.2. This channel corresponds to the left frontal-central region, also known as the motor cortex, which may be related to motor cortex seizures originating from frontal lobe epilepsy [38]. The involved electrode F3 has been identified as one of the seizure onset locations for this seizure type [38]. Furthermore, as previously mentioned, channel 6 has also demonstrated high seizure probability and seizure detection performance using probability-based filtering [29]. Channels 6 and 12 also exhibit higher mean asymmetry

compared to previously dominant asymmetric channels in the preictal state. In particular, channel 12 corresponds to the right parietal-occipital region, which may be associated with occipital lobe epilepsy [39]. These channels may display dominant directional causality patterns linked to seizures arising from frontal lobe epilepsy and occipital lobe epilepsy.

5.2 Limitations and Anomalies

The inconsistent convergence around (L^*, E^*, τ^*) across selected subjects of a representative channel pair can be due to the limitations in the parameter testing process, particularly insufficient convergence evaluation. This approach did not consistently converge, even for the control data (including chb01, see figure 4.4), because the optimal hyperparameters were determined using the mean of a subset of channel pairs from the control data for chb01, which overlooked the individual convergence behaviour of channel pairs. The solution is to ensure that the subsequent convergence checks are performed on each individual channel pair for the selected subjects, or alternatively, the mean of the channel pairs to capture overall convergence trends for each subject. However, due to computational constraints, a single channel pair was selected for the subsequent convergence checks across subjects.

Another limitation is insufficient hyperparameter tuning. Hyperparameter tuning was conducted specifically on the control data from chb01. While this could increase the causality accuracy for the control data, it may result in less accurate causality scores for the preictal and ictal states. In some cases, the causality values reached 1 in magnitude (see figure 4.5), indicating perfect causality. This causality score can be caused by strong unidirectional forcing or artifacts from poor hyperparameter tuning. These anomalies could lead to less stable causality values, impacting the subsequent convergence checks around (L^*, E^*, τ^*), particularly for preictal and ictal state across subjects. The solution would be to implement hyperparameter tuning for each state on chb01, rather than using a uniform set of hyperparameters across all states. However, due to computational constraints, the hyperparameter tuning was conducted for control data from a single subject and a subset of channels.

Furthermore, the limited sample size may introduce bias into the results. Due to computational constraints, this study analyzed 10 subjects out of the 24 subjects from the CHB-MIT dataset. Therefore, these results reflect a small

portion of the dataset and may not represent the broad population of epilepsy patients. This constraint may limit the generalizability of the results, which can undermine the validity of the findings. Thus, including a larger sample size could potentially reduce sampling bias and strengthen the ability to generalize across a wider population.

A concern regarding the generalizability of the findings is the low consistency of asymmetrically dominant channels, with frequency rate 30% and 20%. This reflects high variability across subjects, highlighting individual differences in asymmetry patterns, which is consistent with the inter-subject variability in EEG data [29]. Although the results show high variability, the identified channels during preictal and ictal activity are supported by existing literature, indicating that the findings hold validity for individual subjects. Therefore, this variability suggests the potential for identifying key channels for each subject, providing a subject-specific approach rather than generalizing across subjects.

Another notable concern is that the authors of the CCM framework have acknowledged that anomalies can occur in the causality results. In some cases, both directions of causality reach a magnitude of 1 (see figure 4.5), which can be caused by the synchrony phenomenon derived from strong unidirectional forcing [5]. However, an extensive time-lag analysis could potentially address this anomaly, which is further discussed in the following section.

5.3 Future Work

In this study, the data segments used in the CCM analysis were selected using a randomized approach. A future consideration is to implement the sliding window approach to check convergence across all overlapping segments, rather than selecting a single random segment. This approach is the most commonly utilized method for capturing dynamic neural patterns [40], which ensures robust signal processing. In the case that the causality values fails to consistently converge across all windows for the hyperparameters (L^*, E^*, τ^*), additional tuning may be necessary.

To improve the time-lag analysis, the range of time lags steps can be extended to negative values. This extension can help detect delayed causal effects and causal feedback loops in dynamical systems, which Ye et al. has shown in ecological systems [41]. This approach provides a way to differentiate external forcing from true bi-directional causality [41]. In cases of strong unidirec-

tional forcing, the response variable tends to better predict past values of the driving variable, indicating a negative lag, while the driving variable better predicts the future response, indicating positive lag [41]. This approach can result in a more robust and accurate hyperparameter tuning of the time delay τ^*.

The data consists of a heterogeneous set of seizure types, including both focal and generalized seizures. Focal seizures originate from a specific brain region in one hemisphere while the generalized seizure spreads to both hemispheres simultaneously [9]. The analysis should be conducted separately for the seizure types to distinguish between local and global seizure dynamics. Particularly, local dynamics are important for pinpointing initial sources of abnormal electrical discharges and revealing patterns of seizure propagation, as this study [42] has shown through successfully detecting directional connections of seizure onset zones using focal seizures. Moreover, analysing this seizure type could lead to more consistent and stable causality results with better convergence than generalized seizures that have more complex and widespread electrical activity.

To improve the preprocessing pipeline, Independent Component Analysis (ICA) can be utilized in combination with the current band-pass filter. Band-pass filtering remains important for selecting the range of frequencies that the analysis focuses on. However, the ICA method provides better artifact removal, as it avoids eliminating affected data portions [43]. Moreover, the GC study [3] also implemented the ICA method to efficiently clean the data from bad components before exploring the causality. These preprocessing steps can enhance the reliability of the CCM analysis by preserving meaningful EEG signals while minimizing artifacts.

In this paper, the pre-ictal window length was set to 30 seconds to capture relevant connectivity patterns, as Ye et al. [30] showed significant functional connectivity patterns during this pre-ictal length. However, identifying the optimal preictal period (OPP) is challenging due to the inter-subject variability in EEG patterns [43]. In previous studies, the OPP has ranged from 5 - 173 minutes [43]. Therefore, performing tests on different preictal lengths is advised. The OPP is determined by the preictal length that produces the highest causality scores i.e. strongest causality patterns.

The classical CCM method requires manual assessments of convergence [5],

which is problematic in terms of scalability and lack of uniform criteria. Therefore, future implementations require a scalable CCM method. Recently, a few approaches have been proposed, such as parallelized CCM using Apache Spark [44]. This improved method relies on numerical convergence assessments and parallelized computations for different library lengths. Thus, a scalable method is recommended for more robust and faster CCM processing across multiple subjects.

5.4 Implications

This study identified the key channels in causal asymmetry for pre-seizure and seizure activity, which vary between subjects, emphasizing the need for a subject-specific method. This could propose a novel approach where the selection criterion of channels for epilepsy is based on the asymmetry in causality using CCM. The asymmetry index is typically higher during pre-seizure and seizure activity compared to the non-seizure state, particularly for key channels, implying that the asymmetry can possibly be extracted as features to distinguish between states. These contributions can potentially provide complementary insights into seizure dynamics and improve the accuracy and performance of seizure detection algorithms.

The broader implication of this research is its potential application to other neurological conditions. The proposed methodology may provide a new perspective on analyzing causality patterns in EEG data for disorders such as alzheimer's, schizophrenia and parkinson's disease - if not already being explored in those contexts. Studies on these conditions can potentially benefit from complementary insights into the underlying directional connectivity patterns in the brain, establishing a new approach for analyzing brain signals and interactions. A deeper understanding of these brain dynamics can lead to significantly improved and effective treatment and management methods for patients.

Chapter 6

Conclusions

This study investigates the question: **How can CCM be applied to explore causal asymmetries across channels in epilepsy using EEG data, and what information can it reveal about seizure dynamics?** Although, the traditional CCM has inherent limitations, particularly regarding the convergence checks and hyperparameter tuning, which may potentially affect the accuracy of the causality scores, the findings of this study support the potential of its application. Specifically, the results of the asymmetry index values corroborate the notion that causality patterns are generally more asymmetric during pre-seizure and seizure activity compared to non-seizure activity. Furthermore, results from individual channels suggest that channels 20 and 21, as well as channels 6 and 12 most consistently exhibit the highest asymmetry in causality for pre-seizure and seizure activity, respectively, which is partially consistent with previous findings. This leads to the conclusion that CCM can be applied, with the improvements of scalable methods, to identify potential EEG channels that are important for underlying directional connectivity involved in seizure dynamics.

Bibliography

[1] World Health Organization. *Epilepsy*. en. URL: https://www.who.int/news-room/fact-sheets/detail/epilepsy (visited on 04/01/2025).

[2] C. P. Panayiotopoulos. "Optimal Use of the EEG in the Diagnosis and Management of Epilepsies". en. In: *The Epilepsies: Seizures, Syndromes and Management*. Bladon Medical Publishing, 2005. URL: https://www.ncbi.nlm.nih.gov/books/NBK2601/ (visited on 04/01/2025).

[3] Robert Coben and Iman Mohammad-Rezazadeh. "Neural Connectivity in Epilepsy as Measured by Granger Causality". In: *Frontiers in Human Neuroscience* 9 (July 2015), p. 194. ISSN: 1662-5161. DOI: 10.3389/fnhum.2015.00194. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4500918/ (visited on 04/02/2025).

[4] Jun Cao et al. "Brain functional and effective connectivity based on electroencephalography recordings: A review". In: *Human Brain Mapping* 43.2 (Oct. 2021), pp. 860–879. ISSN: 1065-9471. DOI: 10.1002/hbm.25683. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8720201/ (visited on 04/01/2025).

[5] George Sugihara et al. "Detecting Causality in Complex Ecosystems". In: *Science* 338.6106 (Oct. 2012). Publisher: American Association for the Advancement of Science, pp. 496–500. DOI: 10.1126/science.1227079. URL: https://www.science.org/doi/10.1126/science.1227079 (visited on 04/02/2025).

[6] Prince Joseph Erneszer Javier. *Time Series Analysis Handbook*. URL: https://phdinds-aim.github.io/time_series_handbook/06_ConvergentCrossMappingandSugiharaCausality/ccm_sugihara.html (visited on 05/10/2025).

[7] Ettore Beghi, Giorgia Giussani, and Josemir W. Sander. "The natural history and prognosis of epilepsy". en. In: *Epileptic Disorders* 17.3 (2015). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1684/epd.2015.0751, pp. 243–253. ISSN: 1950-6945. DOI: 10.1684/epd.2015.0751. URL: https://onlinelibrary.wiley.com/doi/abs/10.1684/epd.2015.0751 (visited on 04/30/2025).

[8] Socialstyrelsen. *Nationella riktlinjer för vård vid epilepsi: stöd för styrning och ledning*. sv. Stockholm: Socialstyrelsen, 2019. ISBN: 978-91-7555-486-0. (Visited on 05/02/2025).

[9] The Johns Hopkins University. *Types of Seizures*. en. Aug. 2021. URL: https://www.hopkinsmedicine.org/health/conditions-and-diseases/epilepsy/types-of-seizures (visited on 05/10/2025).

[10] Sai Sanjay Balaji and Keshab K. Parhi. "Seizure Onset Zone Identification From iEEG: A Review". In: *IEEE Access* 10 (2022), pp. 62535–62547. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3182716. URL: https://ieeexplore.ieee.org/abstract/document/9795000 (visited on 05/08/2025).

[11] Ramendra Nath Bairagi et al. "Epileptic seizure identification in EEG signals using DWT, ANN and sequential window algorithm". In: *Soft Computing Letters* 3 (Dec. 2021), p. 100026. ISSN: 2666-2221. DOI: 10.1016/j.socl.2021.100026. URL: https://www.sciencedirect.com/science/article/pii/S2666222121000150 (visited on 05/05/2025).

[12] Marc R. Nuwer et al. "IFCN standards for digital recording of clinical EEG". In: *Electroencephalography and Clinical Neurophysiology* 106.3 (Mar. 1998), pp. 259–261. ISSN: 0013-4694. DOI: 10.1016/S0013-4694(97)00106-5. URL: https://www.sciencedirect.com/science/article/pii/S0013469497001065 (visited on 04/13/2025).

[13] E. W. Lang et al. "Brain Connectivity Analysis: A Short Survey". In: *Computational Intelligence and Neuroscience* 2012 (2012), p. 412512. ISSN: 1687-5265. DOI: 10.1155/2012/412512. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3477528/ (visited on 05/03/2025).

[14] J. P. Huke. *Embedding Nonlinear Dynamical Systems: A Guide to Takens' Theorem*. en. MIMS Preprint. ISSN: 1749-9097 Issue: 2006.26 Number: 2006.26 Place: Manchester, UK Publisher: Manchester Institute for Mathematical Sciences, University of Manchester. Mar. 2006. URL: https://eprints.maths.manchester.ac.uk/175/ (visited on 04/13/2025).

[15] Anna Krakovská et al. "State space reconstruction techniques and the accuracy of prediction". In: *Communications in Nonlinear Science and Numerical Simulation* 111 (Aug. 2022), p. 106422. ISSN: 1007-5704. DOI: 10.1016/j.cnsns.2022.106422. URL: https://www.sciencedirect.com/science/article/pii/S1007570422001010 (visited on 04/27/2025).

[16] Corpuz James Olegario and Francis. *Time Series Analysis Handbook*. URL: https://phdinds-aim.github.io/time_series_handbook/05_SimplexandSmapProjections/05_Empirical%20Dynamic%20Modelling%20%28Simplex%20and%20SMap_Projections%29.html (visited on 05/10/2025).

[17] GeeksforGeeks. *Symmetric & Skew Symmetric Matrix - Definition, Properties & FAQs*. en-US. Section: Technical Scripter. URL: https://www.geeksforgeeks.org/what-is-symmetric-matrix-and-skew-symmetric-matrix/ (visited on 05/02/2025).

[18] Eric W. Weisstein. *Frobenius Norm*. en. Text. Publisher: Wolfram Research, Inc. URL: https://mathworld.wolfram.com/FrobeniusNorm.html (visited on 05/02/2025).

[19] GeeksforGeeks. *AutoCorrelation*. en-US. Section: Data Science. URL: https://www.geeksforgeeks.org/autocorrelation/ (visited on 05/02/2025).

[20] Karin Schiecke et al. "Convergent Cross Mapping: Basic concept, influence of estimation parameters and practical application". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. ISSN: 1558-4615. Aug. 2015, pp. 7418–7421. DOI: 10.1109/EMBC.2015.7320106. URL: https://ieeexplore.ieee.org/document/7320106 (visited on 04/27/2025).

[21] Ali Shoeb. *CHB-MIT Scalp EEG Database*. 2010. DOI: 10.13026/C2K01R. URL: https://physionet.org/content/chbmit/ (visited on 04/03/2025).

[22] Jonathan Dan. *Description of the Seizure detection challenge (2025)*. en. URL: https://epilepsybenchmarks.com/challenge-description/ (visited on 04/03/2025).

[23] Ali Shoeb and John Guttag. "Application of Machine Learning To Epileptic Seizure Detection". en. In: *Proceedings of the 27 th International Confer- ence on Machine Learning* (2010). URL: https://physionet.org/content/chbmit/1.0.0/shoeb-icml-2010.pdf (visited on 04/05/2025).

[24] Hitten P. Zaveri, Robert B. Duckrow, and Susan S. Spencer. "On the use of bipolar montages for time-series analysis of intracranial electroencephalograms". In: *Clinical Neurophysiology* 117.9 (Sept. 2006), pp. 2102–2108. ISSN: 1388-2457. DOI: 10.1016/j.clinph.2006.05.032. URL: https://www.sciencedirect.com/science/article/pii/S1388245706002616 (visited on 05/09/2025).

[25] Xizhen Zhang et al. *Frontiers | A review of epilepsy detection and prediction methods based on EEG signal processing and deep learning*. URL: https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2024.1468967/full (visited on 04/20/2025).

[26] MNE. *Tutorials — MNE 1.9.0 documentation*. URL: https://mne.tools/stable/auto_tutorials/index.html (visited on 04/19/2025).

[27] MNE. *Filtering and resampling data — MNE 1.9.0 documentation*. URL: https://mne.tools/stable/auto_tutorials/preprocessing/30_filtering_resampling.html#tut-filter-resample (visited on 04/25/2025).

[28] GeeksforGeeks. *Normalization vs Standardization*. en-US. Section: Difference Between. URL: https://www.geeksforgeeks.org/normalization-vs-standardization/ (visited on 04/25/2025).

[29] Di Wu, Jie Yang, and Mohamad Sawan. "Bridging the gap between patient-specific and patient-independent seizure prediction via knowledge distillation". eng. In: *Journal of Neural Engineering* 19.3 (June 2022). ISSN: 1741-2552. DOI: 10.1088/1741-2552/ac73b3. (Visited on 04/10/2025).

[30] Hongyi Ye et al. “Pre-ictal fluctuation of EEG functional connectivity discriminates seizure phenotypes in mesial temporal lobe epilepsy”. In: *Clinical Neurophysiology* 151 (July 2023), pp. 107–115. ISSN: 1388-2457. DOI: 10.1016/j.clinph.2023.05.004. URL: https://www.sciencedirect.com/science/article/pii/S1388245723006223 (visited on 05/05/2025).

[31] Lorenze Peppoloni et al. *Examples of the choice of the embedding delay The autocorrelation...* en. URL: https://www.researchgate.net/figure/Examples-of-the-choice-of-the-embedding-delay-t-The-autocorrelation-function-A-B-and_fig3_313664454 (visited on 05/06/2025).

[32] Josef Perktold et al. *statsmodels/statsmodels: Release 0.14.2*. Apr. 2024. DOI: 10.5281/ZENODO.593847. URL: https://zenodo.org/doi/10.5281/zenodo.593847 (visited on 04/15/2025).

[33] George Sugihara. *Basic EDM Examples - pyEDM, rEDM*. URL: https://sugiharalab.github.io/EDM_Documentation/basic_examples/ (visited on 05/10/2025).

[34] Stefan Sumsky and L. John Greenfield. “Network analysis of preictal iEEG reveals changes in network structure preceding seizure onset”. en. In: *Scientific Reports* 12.1 (July 2022). Publisher: Nature Publishing Group, p. 12526. ISSN: 2045-2322. DOI: 10.1038/s41598-022-16877-x. URL: https://www.nature.com/articles/s41598-022-16877-x (visited on 05/03/2025).

[35] Satoshi Fujita et al. “Preictal Activity of Subicular, CA1, and Dentate Gyrus Principal Neurons in the Dorsal Hippocampus before Spontaneous Seizures in a Rat Model of Temporal Lobe Epilepsy”. en. In: *Journal of Neuroscience* 34.50 (Dec. 2014). Publisher: Society for Neuroscience Section: Articles, pp. 16671–16687. ISSN: 0270-6474, 1529-2401. DOI: 10.1523/JNEUROSCI.0584-14.2014. URL: https://www.jneurosci.org/content/34/50/16671 (visited on 04/11/2025).

[36] Shilpa Sj et al. “Classification of Ictal and Preictal Seizure using EEG Signals based on Convolutional Neural Network”. en. In: *2023 IEEE International Conference on Imaging Systems and Techniques (IST)*. Copenhagen, Denmark: IEEE, Oct. 2023, pp. 1–6. ISBN: 979-8-3503-3083-0. DOI: 10.1109/IST59124.2023.10438054. URL: https:

//ieeexplore.ieee.org/document/10438054/ (visited on 04/21/2025).

[37] Maria de los Angeles Castillo Rodriguez, Armin Brandt, and Andreas Schulze-Bonhage. "Differentiation of subclinical and clinical electrographic events in long-term electroencephalographic recordings". en. In: *Epilepsia* 64.S4 (2023). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/epi.1740 S47–S58. ISSN: 1528-1167. DOI: 10.1111/epi.17401. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.17401 (visited on 04/19/2025).

[38] Beatriz García-López et al. "Morphological Description of Frontal EEG Interictal and Ictal Discharges in an Adult Cohort of 175 Patients". In: *Journal of Clinical Medicine* 10.6 (Mar. 2021), p. 1219. ISSN: 2077-0383. DOI: 10.3390/jcm10061219. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7999909/ (visited on 04/27/2025).

[39] P. D. Williamson et al. "Occipital lobe epilepsy: clinical characteristics, seizure spread patterns, and results of surgery". eng. In: *Annals of Neurology* 31.1 (Jan. 1992), pp. 3–13. ISSN: 0364-5134. DOI: 10.1002/ana.410310103. (Visited on 04/28/2025).

[40] Wei Liao et al. "DynamicBC: A MATLAB Toolbox for Dynamic Brain Connectome Analysis". In: *Brain Connectivity* 4.10 (Dec. 2014), pp. 780–790. ISSN: 2158-0014. DOI: 10.1089/brain.2014.0253. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4268585/ (visited on 04/26/2025).

[41] Hao Ye et al. "Distinguishing time-delayed causal interactions using convergent cross mapping". In: *Scientific Reports* 5 (Oct. 2015), p. 14750. ISSN: 2045-2322. DOI: 10.1038/srep14750. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4592974/ (visited on 04/13/2025).

[42] Yaotian Wang et al. *High-Dimensional Directional Brain Network Analysis for Focal Epileptic Seizures*. arXiv:2208.07991 [stat]. Aug. 2022. DOI: 10.48550/arXiv.2208.07991. URL: http://arxiv.org/abs/2208.07991 (visited on 05/06/2025).

[43] Arnaud Delorme. *d. Indep. Comp. Analysis*. en-US. URL: https://eeglab.org/tutorials/06_RejectArtifacts/RunICA.html (visited on 04/24/2025).

[44] Bo Pu, Lujie Duan, and Nathaniel Osgood. *Parallelizing Convergent Cross Mapping Using Apache Spark*. arXiv:1905.00565 [cs]. May 2019. DOI: 10.48550/arXiv.1905.00565. URL: http://arxiv.org/abs/1905.00565 (visited on 04/26/2025).

Appendix A

Pseudocodes

These pseudocodes reflect the python scipts used in this study. These codes are developed by Tenzin Sangpo Choedon and available at this GitHub Repository.

A.1 CCM algorithm

Algorithm 1 Compute causality scores using CCM

function COMPUTE_CCM(X_{all}, L, E, τ)
    $N \leftarrow$ number of channels in X_{all}
    Initialize $C \leftarrow$ zero matrix of size $N \times N$
    for $i \leftarrow 0$ to $N - 2$ **do**
        $X \leftarrow X_{all}[i, start\_index : end\_index]$
        for $j \leftarrow i + 1$ to $N - 1$ **do**
            $Y \leftarrow X_{all}[j, start\_index : end\_index]$
            $ccm\_correl_{X \to Y} \leftarrow$ CCM(X, Y, τ, E, L).CAUSALITY$(())[0]$
            $ccm\_correl_{Y \to X} \leftarrow$ CCM(Y, X, τ, E, L).CAUSALITY$(())[0]$
            $C[i, j] \leftarrow ccm\_correl_{X \to Y}$
            $C[j, i] \leftarrow ccm\_correl_{Y \to X}$
        end for
    end for
    return C
end function

A.2 Convergence Analysis

Algorithm 2 Plot causality across L values for convergence analysis

function PLOT_OVERALL_CONVERGENCE($File, L\_range, E\_range, \tau\_range$)
  for each E in E_range **do**
    for each τ in τ_range **do**
      Initialize $corr\_mean\_vals_{X \to Y} \leftarrow []$
      Initialize $corr\_mean\_vals_{Y \to X} \leftarrow []$
      for each L in L_range **do**
        $C \leftarrow$ Load matrix with (L, E, τ) from $File$
        $corr\_mean_{X \to Y} \leftarrow$ mean(triu(C))
        $corr\_mean_{Y \to X} \leftarrow$ mean(tril (C))
        $corr\_mean\_vals_{X \to Y}.append(corr\_mean_{X \to Y})$
        $corr\_mean\_vals_{Y \to X}.append(corr\_mean_{Y \to X})$
      end for
      $plot(L\_range, corr\_mean\_vals_{X \to Y})$
      $plot(L\_range, corr\_mean\_vals_{Y \to X})$
    end for
  end for

A.3 Autocorrelation Analysis

Algorithm 3 Compute autocorrelation of causality scores to estimate optimal τ

function COMPUTE_OVERALL_ACF($File, L^*, E\_range, \tau\_range$)
    Initialize $opt\_taus \leftarrow []$
    for each E in E_range **do**
        Initialize $corr\_mean\_vals_{X \to Y} \leftarrow []$
        Initialize $corr\_mean\_vals_{Y \to X} \leftarrow []$
        for each τ in τ_range **do**
            $C \leftarrow$ Load matrix with (L^*, E, τ) from $File$
            $corr\_mean_{X \to Y} \leftarrow$ mean(triu(C))
            $corr\_mean_{Y \to X} \leftarrow$ mean(tril(C))
            $corr\_mean\_vals_{X \to Y}.append(corr\_mean_{X \to Y})$
            $corr\_mean\_vals_{Y \to X}.append(corr\_mean_{Y \to X})$
        end for
        $acf_{X \to Y} \leftarrow$ acf($corr\_mean\_vals_{X \to Y}$)
        $acf_{Y \to X} \leftarrow$ acf($corr\_mean\_vals_{Y \to X}$)
        $plot(\tau\_range, acf_{X \to Y})$
        $plot(\tau\_range, acf_{Y \to X})$
        $opt\_tau_{X \to Y} \leftarrow$ first zero-crossing of $acf_{Y \to X}$
        $opt\_tau_{Y \to X} \leftarrow$ first zero-crossing of $acf_{X \to Y}$
        $opt\_taus.append(opt\_tau_{X \to Y})$
        $opt\_taus.append(opt\_tau_{Y \to X})$
    end for
    $overall\_tau \leftarrow$ most frequent value in opt_taus
    return $overall\_tau$
end function=0

A.4 Simplex Projection

Algorithm 4 Compute simplex projection to estimate optimal E across limited channels

function COMPUTE_SIMPLEX_PROJECTION($X_{all}, L^*, E\_range, \tau^*, Channels$)
    Initialize $opt\_Es \leftarrow []$
    for each ch in $Channels$ **do**
        Initialize $corr\_mean\_vals \leftarrow []$
        $X_{ch} \leftarrow$ extract data points of ch from X_{all}
        for each E in E_range **do**
            $res \leftarrow$ apply simplex projection with (L^*, E, τ^*) on X_{ch}
            $corr\_arr \leftarrow$ pearson(res[Observation],res[Predictions])
            $corr\_mean \leftarrow$ mean($corr\_arr$)
            $corr\_mean\_vals.append(corr\_mean)$
        end for
        $plot(E\_range, corr\_mean\_vals)$
        $opt\_E \leftarrow E\_range[\arg\max(corr\_mean\_vals)]$
        $opt\_Es.append(opt\_E)$
    end for
    $overall\_E \leftarrow$ most frequent value in opt_Es
    return $overall\_E$
end function