universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Major depressive disorder treatment prediction using entropy-based measures

verfasst von | submitted by

Alexander Stähle

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 645
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Data Science
Betreut von | Supervisor:	RNDr. CSc. Katerina Schindlerova Privatdoz.
Mitbetreut von | Co-Supervisor:	Dipl.-Ing. Dr.techn. Lukas Miklautz BSc

Acknowledgements

I would like to express my heartfelt gratitude to my supervisors, Dr. Katerina Schindlerova (Kateřina Hlaváčková-Schindler) and Dr. Lukas Miklautz, for their unwavering support throughout the development of this thesis. Dr. Schindlerova's encouragement and constructive criticism challenged me to think more critically and approach problems from different perspectives. I am also immensely grateful to Dr. Lukas Miklautz, whose mentorship not only contributed to the academic success of this thesis but also to my personal and professional growth.

I would also like to thank my family and friends for their constant support and encouragement throughout this journey. A special note of appreciation goes to Maeva, whose constant support and faith in my abilities have helped me navigate the challenges of this journey. I also dedicate this work to her late mother, whose loving influence on her daughter is reflected in the way Maeva cares for others.

This thesis was completed as part of the international project "Learning Synchronization Patterns in Multivariate Neural Signals for Prediction of Response to Antidepressants," in collaboration with the Data Mining and Machine Learning Research Group of the University of Vienna and the Czech Academy of Sciences.

Abstract

The traditional 30-day evaluation period for depression treatment is often lengthy, potentially worsening symptoms and reducing remission chances due to the exhaustive process. Predicting treatment response earlier using electroencephalography (EEG) data could significantly shorten treatment times, improve antidepressant selection, and enhance remission outcomes. This thesis explores the utilization of machine learning models to predict depression treatment outcomes by analyzing EEG data collected after seven days of treatment.

We address this as a classification problem, employing both linear and non-linear methods, with an emphasis on entropy-based techniques to extract features from EEG signals and assess brain dynamics. We begin with time-frequency decomposition using wavelet transform analysis and calculate an extensive set of features. To our knowledge, this thesis is the first to use a comprehensive set of entropy-based measures for predicting depression treatment outcomes. We also introduce a novel measure, symbolic influence entropy (SIE), derived from transfer entropy, designed to quantify a signal's past influence on its present value by excluding external information shared across another signal.

Classification is performed in two settings: using each measure individually and combining an ensemble of measures. We evaluate the performance of three feature selection techniques: minimum Redundancy Maximum Relevance (mRMR), F-Test, and L1-regularization-based selection, and analyze results from five machine learning models. Our analysis identifies key measures, frequency ranges, and EEG channels that may serve as markers for predicting treatment response in depression.

The dataset includes 176 patients, divided using nested cross validation (CV) to obtain an average test score across all folds. Among the models, the Support Vector Machine (SVM) demonstrated the highest precision, achieving an F1 score of 0.646 using symbolic transfer entropy (STE) as a single feature. Notably, using multiple measures did not improve precision compared to single measures. STE's strong performance aligns with its frequent selection in multi-measure settings and significance in statistical analysis. T6 emerges as a promising channel through statistical analyses, with delta frequencies slightly better at discriminating between responders and non-responders. Additionally, the choice of entropy variant has only a minimal impact on classification scores, which consistently remain within 0.04 of the best-observed score across different tests. In contrast, other linear and non-linear measures can produce more variable results, with scores ranging from 0.07 to 0.15 points below the best-observed score, making entropy variants more reliable in maintaining high classification precision. Despite some overfitting, our findings suggest EEG data contain brain dynamics that can be extracted using entropy-based methods, potentially aiding in identifying critical remission markers in depression.

Kurzfassung

Die herkömmliche 30-tägige Evaluationsphase zur Behandlung von Depressionen ist oft langwierig, was die Symptome verschlimmern und die Chancen auf eine Remission aufgrund des erschöpfenden Prozesses verringern kann. Eine frühere Vorhersage des Behandlungserfolgs auf Basis von electroencephalography (EEG)-Daten könnte die Behandlungszeit erheblich verkürzen, die Auswahl von Antidepressiva verbessern und somit die Remissionsergebnisse optimieren. Diese Arbeit untersucht den Einsatz von Machine Learning-Modellen zur Vorhersage der Behandlungsergebnisse von Depressionen durch Analyse von EEG-Daten, die nach nur sieben Behandlungstagen erhoben wurden.

Wir betrachten dies als ein Klassifikationsproblem und verwenden sowohl lineare als auch nichtlineare Methoden, wobei wir den Schwerpunkt auf Entropie-basierte Techniken legen, um Feature aus EEG-Signalen zu extrahieren und die Komplexität der Dynamik des Gehirns zu evaluieren. Zunächst führen wir eine Zeit-Frequenz-Dekomposition mittels Wavelet-Transformation durch und berechnen eine umfangreiches Set an Featuren. Soweit uns bekannt ist, ist diese Arbeit die erste, die eine umfassende Reihe von entropiebasierten Metriken in die Vorhersage von Behandlungsergebnissen bei Depressionen einbezieht. Wir führen außerdem ein neue Metrik ein, das wir symbolic influence entropy (SIE) nennen, welche aus dem Konzept der Transferentropie abgeleitet ist und dazu dient, den Einfluss der Vergangenheit eines Signals auf seinen aktuellen Wert zu quantifizieren, indem externe Informationen, die mit einem anderen Signal geteilt werden, ausgeschlossen werden.

Die Klassifikation erfolgt in zwei Szenarien: Einmal unter der Verwendung jeder Metrik einzeln und einmal durch Kombination eines Ensembles verschiedener Metriken. Wir bewerten die Leistung von drei Feature-Selection-Techniken: minimum Redundancy Maximum Relevance (mRMR), F-Test und L1-regularisierungsbasierte Selektion und analysieren die Ergebnisse von fünf Machine Learning-Modellen. Unser Fokus liegt darauf, wichtige Metriken, Frequenzbereiche und EEG-Kanäle zu identifizieren, die als Marker zur Vorhersage der Behandlungsreaktion bei Depressionen dienen könnten.

Der Datensatz umfasst 176 Patienten und wird mittels nested cross validation (CV) aufgeteilt, um einen durchschnittlichen Testscore zu erhalten. Unter den Modellen zeigte die Support Vector Machine (SVM) die höchste Präzision und erzielte einen F1-Score von 0,646 unter Verwendung von symbolic transfer entropy (STE) als alleinige Metrik. Bemerkenswert ist, dass die Verwendung mehrerer Metriken die Präzision im Vergleich zur Verwendung einzelner Metriken nicht verbesserte. Die starke Leistung von STE stimmt mit seiner häufigen Auswahl im Multi-Measure-Ansatz und seiner Signifikanz in der statistischen Analyse überein. In den statistischen Analysen zeigt sich T6 als vielversprechender EEG-Kanal, wobei Delta-Frequenzen eine leicht bessere Unterscheidung zwischen Respondern und Non-Respondern ermöglichen. Zudem hat die Wahl der Entropie-Variante nur einen minimalen Einfluss auf die Klassifikationswerte,

die in verschiedenen Tests konstant innerhalb von 0,04 des besten beobachteten Wertes bleiben. Im Gegensatz dazu führen andere lineare und nichtlineare Metriken zu variableren Ergebnissen, bei denen die Werte um 0,07 bis 0,15 Punkte unter dem besten beobachteten Wert liegen können, wodurch Entropie-Varianten zuverlässiger in der Aufrechterhaltung einer hohen Klassifikationspräzision sind. Obwohl alle Klassifikationspipelines ein gewisses Maß an Overfitting aufweisen, legen unsere Ergebnisse nahe, dass EEG-Daten dynamische Eigenschaften des Gehirns enthalten, die mittels Entropie-basierter Methoden extrahiert werden können, was möglicherweise bei der Identifizierung kritischer Marker für die Remission bei Depressionen hilfreich ist.

Contents

List of Tables

List of Figures

1. Introduction

In this chapter, we present the problem we aim to address and discuss its significance. Section 1.1 overviews the current depression treatment process, highlighting areas where improvements are needed. In Section 1.2, we outline our approach to addressing these challenges and detail our key contributions. Finally, Section 1.3 provides a roadmap for the following chapters.

1.1. Problem Statement

Major depressive disorder (MDD) is a prevalent brain disease worldwide. The World Health Organization reports that an estimated 5% of all adults suffer from depressive disorder [Ins21]. MDD not only impairs health in general but also increases the risk of suicide, the 18th leading cause of death according to the WHO in 2019 [22]. Although several treatment options are available for managing depression, including psychotherapy, a variety of antidepressant medications and transcranial magnetic stimulation techniques, initial remission rates remain below 50% [Gol+02]. Therefore patients revert to a trial-and-error approach to find the optimal treatment method. Unfortunately, this approach yields limited success, with remission rates plateauing at a threshold of 50% even for consecutive antidepressant treatments [Pet+05]. Moreover, this trial-and-error approach is closely associated with the emergence of psychiatric and medical comorbidities, an aggravation of the severity of symptoms in MDD, a reduced likelihood of favorable responses to new treatments and even an increased risk of suicide for the patient [MO04]. Accordingly, the possibility to predict the treatment outcome, thereby mitigating or altogether circumventing lengthy and ineffective therapy procedures, would be desirable.

The database collected to tackle this problem was recorded using a non-invasive method called electroencephalography (EEG). EEG data contain information about the activity of the brain resulting from a large number of neurons being activated in the cerebral cortex [SC13]. It is obtained non-invasively by placing electrodes, that measure the electric fields, along the scalp of the patient. Although there is a more accurate approach in which sensors are inserted surgically into the patient's head, this non-invasive method allows for a risk-free and cost-effective recording [SC13]. Given that EEG produces a substantial volume of data, with data points recorded over time at each individual electrode, managing this extensive dataset is challenging when employing machine learning models. The extraction of meaningful features from EEG data, which contain valuable information for machine learning algorithms to predict treatment outcomes, is highly significant. Ineffective feature extraction may compromise overall performance and effectiveness of machine learning models, potentially limiting their ability to provide meaningful insights and predictions

in the context of MDD treatment.[LC16].

1.2. Aims and Contributions

The aim of this master thesis, conducted within the context of the "Learning Synchronization Patterns in Multivariate Neural Signals for Prediction of Response to Antidepressants" research project, is to determine the consistency of treatment response prediction using a large EEG dataset. The overarching research project involves a diverse set of methods for analysing the acquired EEG data. This work in particular plays a crucial role in the project, focusing on the detailed examination and analysis of a specific method within this complex research. Emphasizing non-linear, entropy-based approaches, this work assesses the ability of these methods to extract predictive features for treatment prediction. We approach the treatment prediction challenge by framing it as a classification task. For this purpose, we employ EEG data collected on the seventh day of treatment and assess the treatment response on the twenty-eighth day, which serves as our target variable.Predictions are generated by extracting features from the EEG data and applying these features to commonly used machine learning classification models. Ideally, the classifier should be interpretable, enabling us to identify the features that contribute most significantly to its predictions. Moreover, it should exhibit strong generalization capabilities, ensuring that the developed framework can be applied to the broadest possible spectrum of patients.

Efforts have been made to forecast treatment outcomes by employing EEG data with linear features [Bai+19][ET16]. However, to date, the research landscape lacks studies extensively exploring the subject of MDD treatment outcomes using non-linear methods, which are renowned for their effectiveness in assessing depression in patients [Cuk+20][Zit+22]. In light of this gap, this work's main contribution lies in applying non-linear methods, specifically concentrating on methods based on the concept of entropy, a measure of uncertainty in a system or signal, to predict treatment outcomes among individuals affected by MDD. Additionally, a type of entropy inspired by the computation of transfer entropy is examined. Transfer entropy quantifies the amount of uncertainty reduced in predicting future values of one variable by knowing the past values of another, given the past of the first variable [Sch00]. The here-examined adaptation of transfer entropy, named Influence Entropy, quantifies the amount of uncertainty reduced about future values of one variable provided by its past, independent of the influence of another variable's past.

1.3. Outline

This thesis is structured as follows: Chapter 2 provides the necessary theoretical background. Chapter 3 reviews related works. Chapter 4 introduces the dataset. Chapter 5 explains the proposed methodology, particularly focusing on feature extraction methods. Chapter 6 presents the experimental results. Chapter 7 discusses the results and offers

insights into potential future research directions. Finally, Chapter 8 provides a summary of the thesis.

2. Foundations

To interpret the results obtained in the master thesis some preliminary background knowledge is required. This chapter will explain the necessary background for the topics of EEG, depression and entropy measures.

2.1. Structure of the brain

Understanding the structure of the brain is crucial for interpreting the results of EEG studies, as the location and function of different brain regions can provide valuable insights into the neural mechanisms underlying various cognitive and emotional processes.
Anatomically, the brain is divided into three main parts: the cerebrum, cerebellum and brainstem [Bri24a].
The cerebrum is the largest part of the brain and is divided into two cerebral hemispheres. The left hemisphere controls language and speech, whereas the right hemisphere interprets visual and spatial information. The cerebrum consists of an inner core of nerve fibres that form a connection with the outer layer, known as the cerebral cortex. It is associated with higher brain functions such as thought and action [MA23][Bri24b].
The cerebellum is located under the cerebrum at the back of the brain and plays a key role in motor control and coordination. It is also involved in some cognitive functions, such as attention and language [MA23][Bri23].
The brainstem, which connects the cerebrum and cerebellum to the spinal cord, performs many basic functions necessary for survival, including heart rate, breathing and sleeping [MA23][Joy23].
Further, the cerebral cortex can be divided into four lobes, which are also roughly defined by major surface folds [Akr24]. The frontal lobe is the largest part of the brain and is responsible for higher cognitive functions, such as decision-making, mood, social and moral reasoning, personality and attention. It also contains the motor cortex which is responsible for planning and coordinating voluntary movements. The parietal lobe is involved in processing sensory information, including vision, hearing, motor, sensory and memory functions. Therefore, it also plays a role in spatial orientation and direction. The occipital lobe is the primary visual processing centre in the brain. It helps determine distance, depth, colour and other aspects of vision. The temporal lobe plays a key role in processing auditory information, memory formation, language comprehension and some aspects of emotion and speech production [MA23][Akr24]. Where depression can be detected in the brain is unclear as multiple papers have reported inconsistent findings [Pan+12].

2.2. EEG

The brain transmits information through nerve cells or neurons. They are the fundamental units of the brain and nervous system and are responsible for receiving sensory input and transmitting information throughout the body. Each neuron comprises a cell body, dendrites and an axon. The dendrites receive signals from other neurons, while the axon transmits signals away from the cell body to other neurons. Synapses are the functional connections between neurons, where an electrical signal is converted into a chemical signal to cross the synapse and reach the next neuron [Kan+00].
EEG is a non-invasive method used to record brain electrical activity. This electrical activity represents the summation of the synchronous activity of thousands or millions of neurons transmitting electrical impulses [SC13].

2.2.1. Acquiring of EEG signals

The EEG signal is captured by placing multiple electrodes on the scalp. These electrodes capture the potential differences between each two electrodes or in comparison to a reference electrode [Ole06].
The number of electrodes and their location on the scalp can vary, but they often follow the internationally recognized 10-20 system. This system is based on the relationship between the location of an electrode and the underlying area of the cerebral cortex [SC13]. With this system, 21 electrodes are placed and labelled using a naming scheme that consists of the rough location of the lobe of the cerebral cortex, the placement is related to F (frontal), T (temporal), C (central), P (parietal), O (occipital) and an additional character indicating either the left (odd numbers) or right (even numbers) hemisphere or the middle line (z) [KB12]. The C does not stand for a specific lobe. It approximately covers the motor cortex and parts of the somatosensory cortex and thus spans the central region of the frontal and parietal lobe [NS05]. The placement and labels can be seen in Figure 2.1.
EEG provides high temporal resolution, capturing real-time brain activity changes on the order of milliseconds ($\sim$5ms). However, it has a relatively low spatial resolution (2-5cm), making it challenging to precisely locate the source of the electrical activity within the brain. This trade-off implies that although EEG is excellent for studying the brain activity timing, it may not be the best tool for pinpointing the exact location of that activity [Kan+00]. Nevertheless, this method enables sufficiently precise localization and notably, its high temporal resolution facilitates almost instantaneous capture of brain activity changes.

2.2.2. Frequency bands

EEG signals are generally categorized into frequency bands, each corresponding to different brain states and functions. These frequency bands are typically divided into delta (0.5-4 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta (13-30 Hz) and gamma (>32 Hz) [NS05]. These can be characterized as follows:

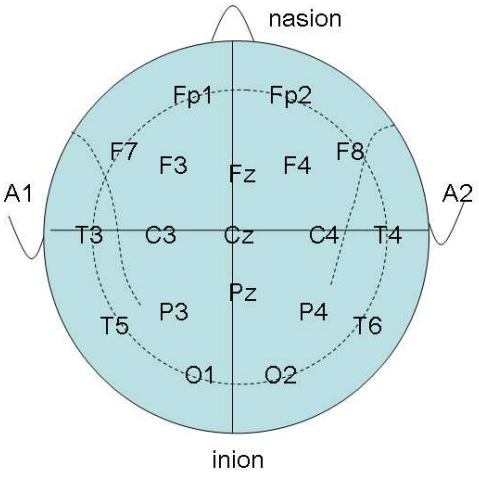


Figure 2.1.: Standardized electrode placement on the scalp according to the 10-20-System [23].

- Delta Waves (<4 Hz): These waves are the slowest and are often associated with deep sleep and certain pathological conditions like cerebral lesions [Jad21].

- Theta Waves (4-7 Hz): These waves are typically associated with drowsiness or the early stages of sleep [Jad21].

- Alpha Waves (8-12 Hz): Alpha waves are a feature of normal wakefulness and have considerable variations. They are often associated with relaxed wakefulness and are most prominent at the parietal and occipital sites [NS05] [Jad21].

- Beta Waves (13-30 Hz): Beta waves are associated with active thinking, active attention, focus on the outside world, or solving concrete problems [NS05] [Jad21].

- Gamma Waves (>30 Hz): Gamma waves are associated with various cognitive processes, such as attention, memory and sensory processing. They have been extensively studied for their role in synchronizing neuronal activity across different parts of the brain, supporting cognitive functions like perception, attentional selection and memory consolidation. [JKL07]

In the context of EEG signal analysis, these frequency bands provide a means of categorizing and interpreting the electrical activity of the brain in more detail. Furthermore, abnormalities in certain frequency bands may indicate neurological or psychiatric disorders [Jad21].

2.2.3. Wavelets

Wavelets are mathematical functions used to decompose signals into components at different scales, which analyze various frequency components, and at different translations, which shift the wavelet function along the time axis [Kas13]. This flexibility makes wavelets a powerful tool for analyzing signals that exhibit varying characteristics across different frequencies. Consequently, wavelets offer advantages in EEG signal analysis compared to traditional filtering techniques that separate the data into frequency bands. The wavelet approach allows for a more nuanced analysis of the EEG signal at multiple scales [Baj21].

There are two types of wavelet transformations, the continuous wavelet transform (CWT) and the discrete wavelet transform (DWT). The CWT analyzes a signal at every possible scale and translation, producing a continuous two-dimensional representation of the signal in the time-frequency space. DWT, on the other hand, efficiently decomposes signals into a hierarchical set of frequency bands, capturing both coarse and fine details. This hierarchical decomposition is achieved through successive high-pass and low-pass filtering [Baj21].

In this work, we will use the DWT. Figure 2.2 shows the repeated decomposition into approximate and detailed coefficients using DWT.

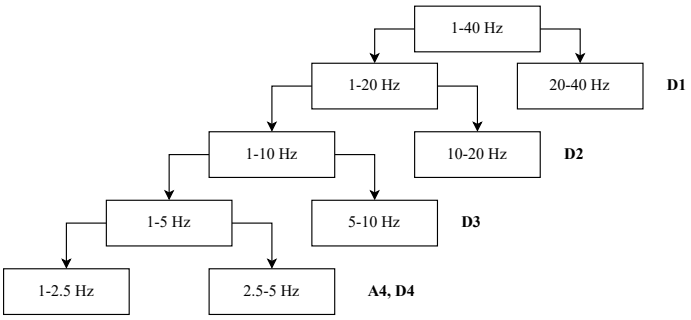


Figure 2.2.: DWT-decomposition of frequency ranges into approximate and detailed wavelet coefficients.

Unlike traditional filters, which use a fixed window size and thus limit their ability to adapt their resolution based on frequency content, wavelets dynamically adjust their window size. This dynamic adjustment allows wavelets to provide precise localization of significant events in the EEG data. It captures short-term high-frequency events with high temporal resolution, as well as low-frequency events with detailed frequency information. Traditional filters with a fixed window size cannot achieve this because they always involve a trade-off between temporal and frequency resolutions [Baj21].

Given that the dataset for this thesis was band-pass filtered between 1 and 40 Hz, four levels of wavelet decomposition are sufficient to extract coefficients corresponding to the previously mentioned frequency bands. Table 2.1 corresponds the wavelet coefficients to the frequency bands they cover.

Wavelet Coefficients	EEG Frequency Bands	Frequency Range (Hz)
D1	Upper Beta + Lower Gamma	20-40
D2	Upper Alpha + Lower Beta	10-20
D3	Upper Theta + Lower Alpha	5-10
D4	Upper Delta + Lower Theta	2.5-5
A4	Delta	1-2.5

Table 2.1.: Wavelet decomposition of EEG signals into frequency bands.

2.3. Entropy

Moving away from EEG, this section briefly introduces the concept of entropy, giving its mathematical formula and provides its interpretation.

While there is no consensus about the definition of a complex system, some characteristics widely associated with complex systems have been identified in the literature. A complex system is composed of many interacting components, that exhibit new properties that cannot be predicted simply from the properties of individual parts [Uth20]. These systems are characterized by nonlinear dynamics, emergence (new patterns), adaptation, self-organization and uncertainty [CCS19]. The complexity of such systems often stems from their diverse, flexible and interdependent elements, which contribute to a level of unpredictability or uncertainty. This is a key aspect of defining and measuring their complexity [Reb+21]. The brain exhibits these signs of a complex system, which highlights the importance of measuring its complexity [Tel+11]. A mathematical framework that deals with uncertainty and unpredictability is information theory.

Information theory is the scientific study of the quantification, storage and communication of digital information [Ash12]. The field was fundamentally established in the 1920s by the works of Nyquist [Nyq24], Hartley [Har28] and Shannon in the 1940s [Sha48]. A key concept in information theory is entropy, which is often considered as a measure of the unpredictability of a sequence of symbols or signals. In simpler terms, it quantifies the amount of information or level of uncertainty in systems [Ash12].

This goes hand-in-hand with the core idea of information theory, which states that the information conveyed through a message depends on the degree to which this information is surprising. Entropy then measures the expected amount of information gained by identifying the outcome of a random trial. The more surprising an event, the more information is gained [Mac03].

Mathematically, entropy is defined as follows: Given a discrete random variable X that takes on values in the alphabet χ and is distributed according to a probability mass function $p : \chi \to [0, 1]$, such that $p(x) := P[X = x]$, entropy is defined as follows:

$$H(X) = -\sum_{x \in \chi} p(x) \log_b p(x)$$

, where b is the base of the logarithm, which is commonly set to 2 [Sha48]. This definition states that the higher the probability of an outcome x_i, the less "surprising" it is and the less it contributes to the overall entropy. This mathematical formulation is crucial as it provides a quantitative tool for measuring information content and supports various applications in information theory.

As information theory has evolved, different variations of entropy have been developed, each adapting the original concept for specific applications or nuances. Section 5 goes deeper into a selected set of these variations, showcasing how this fundamental concept has been adapted and extended to address specific challenges and questions in both theoretical and applied contexts.

When applying complexity analysis methodologies, including entropy, to biological systems, especially the brain, it is important to distinguish between complexity and randomness [Gra12]. While complex systems produce outputs that may seem unpredictable and irregular, they contain inherent meaningfulness that is not found in random systems [YT13]. This key difference has been highlighted in the literature, however, the exact boundary between complexity and randomness remains somewhat blurry. This ambiguity poses challenges when directly determining whether increases or decreases in complexity measure values are related to higher or lower complexity of the system [De +16].

3. Related Work

In this section, a comprehensive review of research conducted on EEG depression data is presented. This study focuses on the utilization of entropy-based methods for feature extraction for classification using machine learning models.
Although the analysis of EEG data to gather insights into MDD is not a new approach, as it has been conducted in the late 1990s, for example, in [Got98] and [AH01], the extensive capabilities of machine learning techniques to extract information from EEG data have heavily improved, leading to deeper and more complex analysis.

3.1. Linear and Non-Linear Approaches

In the early stages of EEG analysis, only linear methods were used to extract features from the brain signals. Most studies that employ linear methods use variants of power spectrum analysis ([Kno+01], [All+04], [Ste+10]). This involves converting the signal from the time domain to the frequency domain and decomposing it into constituent frequency components, allowing for an analysis of the relative power distribution rather than being influenced by the absolute values of the powers [Bac+18]. Methods utilizing power spectrum analysis include the spectral asymmetry index (SASI) [Hin+09], alpha power variability (APV) [Bac+18] and relative gamma power (RGP) [MLP16]. Iešmantas et al. used coherence, phase lag and asymmetry [IA20]. In addition, dimensionality reduction techniques such as independent component analysis (ICA) or principal component analysis (PCA) are also utilized to extract features [GAD08].
Müller et al. conversely argued in their work that linear methods are not able to capture the complex dynamical variations and fractal nature of EEG signals as effectively as non-linear methods [MJA17]. This is supported by an analysis by Stam et al., who argue that except for epileptic seizure brain dynamics, the brain states and dynamics are too complex to be solved by a simple analysis and necessitate non-linear methods [Sta05]. Akar et al. concluded that non-linear analysis is discriminative when explaining the brain dynamics of MDD patients. They even looked at different recording scenarios and identified significant differences between healthy subjects and subjects suffering from MMD in terms of brain dynamics revealed by non-linear methods [Aka+15]. Although the above-mentioned studies solely use non-linear methods, some use both linear and non-linear methods. However, the findings of these studies are inconclusive. At present, it is not clear whether non-linear methods, as in [Bac+15], consistently perform better than linear methods [Bac+18]. Still, many studies have reported a better performance using nonlinear features, as shown in [HMR13], [Ach+15], [Sto+15]. Differences in the

results can be explained by the small sample size in most studies in this field. As a result, the scientific literature arguing for better modelling of brain dynamics using non-linear methods, as suggested in [Bac+15], [Aka+15], [Cuk+20], [HMR13], [Sta05], has led to increasing popularity.
Contrary to linear methods which, when chosen in literature, are mostly based on power spectrum analysis, the types of non-linear methods vary greatly. There is no consensus on which non-linear feature performs best. However, there is research agreement that no single non-linear feature is sufficient to capture all the dynamics of the brain and may even be misleading in this regard [BR15]. Because no clear method exists to determine which features work best, multiple non-linear approaches exist: Čukić et al. attempted to carefully select their features based on different frequency ranges. They chose Higuchi's fractal dimension (HFD) which shows better performance in the higher frequency band along with sample entropy (SampEn) which performs better at lower frequencies [Cuk+20]. Hosseinifard et al., on the other hand, use more non-linear features than just two. They use Detrended Fluctuation Analysis, HFD, correlation dimension (CD) and the Lyapunov exponent (LLE) as their non-linear features, however, their selection is not explained [HMR13]. Zhu et al. used an even larger quantity of non-linear features. These include similar features to those of the two studies mentioned before and use CD, LLE, Lempel-Ziv complexity (LZC), Kolmogorov entropy (Kol), C0-complexity, approximate entropy (ApEn), permutation entropy (PermEn), Singular-value Deposition entropy, Min-entropy, Hartley entropy and Spectral entropy [Zhu+19].
Akar et al. used complementary non-linear features in the sense that different non-linear methods calculate different types of analysis roughly based on the concept of structuring complexity methods into dimensional complexity, regularity and predictability. Therefore, they selected ShannonEn, LZC, Kol, Katz fractal dimension (KFD) and HFD [Aka+15].

3.2. Entropy-based methods

A relevant set of features, given the entropy-based foundation of this thesis, are computed using methods that are based on the concept of entropy. Entropy-based methods are frequently referred to as entropy measures in the literature, which stem from the nature of entropy itself. Entropy is a metric to quantify data uncertainty or information content, however, entropy measures can also be utilized as features in machine learning models. In this thesis, entropy measures are thus regarded as the input features of a machine learning model, if not stated otherwise.

Although there are still studies that do not use entropy methods, the trend in the literature appears to be towards using entropy variations ([Cuk+20], [Lau+22], [Sup+20],[Zhu+19]). In fact, some studies only used variants of entropy methods to engineer their features. For example, Liang et al. investigated twelve entropy measures for monitoring the depth of anaesthesia [Lia+15]. They found that all entropy variants could detect changes in the EEG data during different anaesthesia states but the selection of the best-performing feature was situation-dependent. This reinforces the importance

of selecting features with care. Faust et al. most notably solely utilize entropy-based features and achieved a classification accuracy of 99.5% in identifying MDD patients [Fau+14].
Other entropy-based approaches mentioned in the literature in connection with depression EEG data include symbolic transfer entropy (STE), which is commonly used as a measure to evaluate the flow of information in dynamic and multidimensional systems and is therefore useful in analyzing psychiatric disorders such as MDD [SL08] [Mar+22].
Azami et al, conducted a comprehensive study on entropy analysis of biomedical signals [Aza+22]. They first detail the theory behind the selected methods to form a heterogeneous set of entropy measures. Additional contributions include a comparison in terms of computational efficiency as well as limitations and advantages considering noise, data length and their ability to discriminate uncertainty in a system. Among the investigated methods, fuzzy entropy (FuzzyEn) was the overall best-performing method among the methods that denote the rate of information production (conditional entropy (CE)). In addition, FuzzyEn appears to be strongly correlated with many other methods. On the other hand, among the methods that quantify the total amount of information (Shannon entropy (ShannonEn)), permutation entropy (PermEn) and distribution entropy (DistEn) are not strongly correlated with other methods. PermEn, for example, has shown promising results in the detection of epileptic episodes [NG12], while FuzzyEn was used by Mohamaddi et al. to distinguish the severity of MDD [MM21].

3.3. Treatment response prediction

Depression research can be divided into several areas. The majority of the research focuses on differentiating between depressed and healthy subjects, which includes determining whether or not a subject is depressed using features engineered from EEG data. As previously stated, we will focus on determining whether a medication in the early phase of treatment will effectively reduce depression symptoms and apply the appropriate treatment to each patient individually.
Although this field of study is not as well known as depression detection, there are a few promising studies. Watts et al., in particular, recently published a review of treatment prediction efforts. The studies covered are divided into two types of treatments: neurostimulation therapy and antidepressants. The authors report on the accuracy of each paper and find that predicting neurostimulation responses is generally more precise. All of these studies demonstrated high accuracies when predicting neurostimulation response, ranging from 89% to 91% [Wat+22].
Some relevant studies are mentioned in the following. Bailey et al. predicted treatment response to repetitive transcranial magnetic stimulation (rTMS), a non-invasive, safe and painless treatment that works by sending brief repetitive magnetic pulses that can modulate neuron activity in the target area of the brain [Lef+14]. The authors achieved an accuracy of 91% on working memory EEG data obtained from 39 participants using linear features [Bai+18]. They also predicted rTMS response based on resting EEG data

as a baseline and EEG data from 32 participants after one week of treatment, achieving an accuracy of 86% [Bai+19]. Erguzel et al. employed EEG cordance, a quantitative measure to compare absolute and relative power in different regions of the brain, to obtain an accuracy of 89% in predicting rTMS response [Erg+15].
Contrary to the approaches in the previous two papers, Hasanzadeh et al. implemented their work by utilizing nonlinear features. Their prediction of rTMS response incorporated not only nonlinear features but also analyses of power spectrum, bispectral and cordance features. They selected LZC, CD and KFD as non-linear features. In comparison to other depression classification studies, this is a relatively small subset of non-linear features. In addition, no entropy-based methods were employed. Using a combination of all features, the authors obtained an accuracy of 87%. They discovered that power spectrum features outperformed cordance features in distinguishing between responder and non-responder groups, which contradicts Erguzel et al.'s findings [HMR19]. This could be attributed to the fact that Hasanzadeh et al. had a relatively small sample size of 46 participants, whereas Erguzel et al. had a dataset of 147 participants. Both participation counts are relatively low but difficult to beat in practice.
Arns et al. conducted two separate studies on linear and non-linear methods for predicting treatment outcomes in patients undergoing rTMS and psychotherapy. They found that both linear and non-linear methods effectively differentiated between responders and non-responders. Linear measures included pre-frontal cordance, fronto-central theta and anterior individual alpha peak frequency [Arn+12]. Non-linear measures included LZC, LLE and False Nearest Neighbors. Among these features, only LZC was able to notably distinguish between responders and non-responders, primarily by detecting an opposing shift in values between minute 1 and minute 2 of EEG recordings [Arn+14].
The current research on predicting the treatment response of MDD exhibits a notable scarcity of entropy-based methodologies. Shalbaf et al. utilize an entropy-based measure but lack the variety of other variants of entropy. They use PermEn as their only measure in combination with empirical mode decomposition instead of classical separation of frequency bands and obtain an area under the curve (AUC) of 0.8 [Sha+18].
In the scope of the parent project of this thesis, two previous works have been published, which employ different approaches on the same dataset that is used in this thesis and thus provide valuable benchmarks for comparison. Hlaváčková-Schindler et al. [Hla+23] used Granger-causal graphs to extract features and trained two separate classifiers to differentiate between male and female patients, reporting test F1 scores of 0.6 for female patients and 0.2 for male patients. Additionally, Kraljevska [Kra24] utilized motif-discovery algorithms and achieved a best F1 score of 0.782 on a separate test set.

3.4. Overview

This thesis investigates the ability of entropy-based features to distinguish responders from non-responders to not only rTMS but also different types of MDD treatments, such as different neurostimulation therapy approaches and antidepressants. Specifically, this study addresses the gap in the use of entropy-based methods in treatment prediction,

which is more commonly applied in depression detection (see Table 3.1). The dataset of EEG screenings used in this thesis is quite large, with 176 participants, making the results more generalizable than in previous studies. By utilizing a comprehensive set of extracted features and a large dataset within the domain, this thesis aims to guide future studies in identifying potential predictors of treatment outcomes.

Table 3.1.: Number of studies that incorporated entropy-based measures in their feature set by problem setting.

Setting	**Count**
Depression detection	6
Depression treatment prediction	1

Table 3.2.: Summary of Research Studies in the field of MDD. Non disclosed values are represented by "/".

Authors	**Target**	**Measures**	**Model(s)**	**Accuracy**	**Dataset Size**
Hosseinifard et al. [HMR13]	detection	power, DFA, HFD, CD, LLE	Logistic regression	90%	90 patients (45 depressed, 45 healthy)
Akar et al. [Aka+15]	detection	KFD, HFD, ShannonEn, LZC, KC	/	/	30 patients (15 depressed, 15 healthy)
Bachmann et al. [Bac+15]	detection	LZC	/	/	34 patients (17 depressed, 17 healthy)
Bachmann et al. [Bac+18]	detection	LZC, DFA, SASI, power, HFD	Logistic regression	92%	26 patients (13 depressed, 13 healthy)
Cukic et al. [Cuk+20]	detection	HFD, SampEn	MLP, LR, SVM, DT, RF (best), NB	90.24 - 97.56%	/

Continued on next page

Table 3.2 continued from previous page

Authors	Target	Measures	Model(s)	Accuracy	Dataset Size
Zhu et al. [Zhu+19]	detection	ApEn, LZC, Kol, PermEn, CD, LLE, C0, SVDen, ShannonEn, Min-E., Hartley E., SpectralEn	SVM, RF, XGB, MLP	83.42%	/
Faust et al. [Fau+14]	detection	ApEn, SampEn, Renyi E., bispectral phase E.	PNN, SVM, DT, KNN, NB, GMM, Fuzzy Sugeno Classifier	99.5%	/
Margarette Sanchez et al. [Mar+22]	distinct between depression and bipolar disorder	STE	SVM, RF, KNN	84.9%	142 patients (71 MDD and 71 Bipolar)
Mohammadi et al. [MHM19]	depression level distinction	FuzzyEn, KFD, fuzzy fractal dimension	SVM, fuzzy function neural network	90% (binary)	60 patients
Bailey et al. [Bai+18]	treatment response prediction	power, connectivity	SVM	91%	39 patients with treatment-resistant depression
Bailey et al. [Bai+19]	treatment response prediction	power, connectivity, cordance	SVM	86.6%	32 patients with treatment-resistant depression
Erguzel et al. [Erg+15]	treatment response prediction	cordance	NN	89.12%	147 patients with treatment-resistant depression

Continued on next page

Table 3.2 continued from previous page

Authors	Target	Measures	Model(s)	Accuracy	Dataset Size
Hasanzadeh et al. [HMR19]	treatment response prediction	power, bispectrum analysis, cordance, KFD, LZC, CD	KNN	87%	46 patients with MDD
Arns et al. [Arn+12]	treatment response prediction	power, cordance, P300, Individual alpha peak frequency	/	/	90 patients with MDD
Arns et al. [Arn+14]	treatment response prediction	LZC, LLE, False Nearest Neighbors	/	/	90 patients with MDD
Shalbaf et al. [Sha+18]	treatment response prediction	PE	/	/	87 patients (62 depressed, 25 healthy controls)
Mumtaz et al. [Mum+17]	treatment response prediction	Coherence, P300, Fourier transform	LR	87.5%	64 patients (34 depressed, 30 healthy controls)
Hlaváčková-Schindler and Pacher et al. [Hla+23]	treatment response prediction	Granger-causal graphs	DT	F1: 0.61 (females), 0.2 (males)	176 patients (84 responders, 92 non-responders)
Kraljevska [Kra24]	treatment response prediction	motif-discovery algorithms	SVM	F1: 0.782	176 patients (84 responders, 92 non-responders)

4. Dataset

This chapter describes the dataset used in the experiments presented in this thesis. The chapter will cover the data acquisition process, the demographics of the participants as well as a definition of response to the treatment. The Synchronisation project members at the Czech Academy of Sciences recorded the EEG data and performed the preprocessing and labelling steps outlined in this section.

4.1. Study participants

The study comprised a total of 228 subjects who suffered from depression without any treatment being done beforehand. However, some of the recordings of participants were not usable because of technical issues during recording or an increased number of artefacts in the EEG data. Therefore, data of a final total of 176 subjects was analyzed. Statistics on the subjects can be found in table 4.1. The data revealed a significant bias towards female patients, which will be addressed in Chapter 5. In addition, male patients generally have a lower response rate to treatments. The participants mostly underwent one type of treatment, with the treatment choice being different for each participant, with some patients undergoing a combination of treatments. An overview of treatments used can be found in table 4.2.

Table 4.1.: Summary of patient data and treatment outcomes, R- responders, NR- non-responders.

	R	**NR**	**Total**
Age [years]	45.99 (±11.21)	45.99 (±12.82)	45.99 (±11.70)
Gender (female/male)	66/18	62/30	128/48
MADRS Pre-treatment	27.67 (±3.43)	27.80 (±4.15)	27.70 (±3.61)
MADRS Post-treatment	12.04 (±6.02)	25.25 (±5.60)	15.45 (±8.27)
Responders at final	84	92	176

4.2. Definition of response

To assess the impact of treatments on the participants' mental health, a psychological evaluation was performed both before treatment began and after 28 days of treatment. The Montgomery-Åsberg Depression Rating Scale (MADRS) was used in this evaluation

Table 4.2.: Distribution of treatments among responders and non-responders. (At the time of the thesis not all treatment plans were available).

Treatment	**R**	**NR**	**Total**
Selective Serotonin Reuptake Inhibitors	22	5	27
Serotonin and Norepinephrine Reuptake Inhibitors	2	1	3
Norepinephrine and Dopamine Reuptake Inhibitors	9	5	14
Repetitive Transcranial Magnetic Stimulation	10	6	16
Transcranial Direct Current Stimulation	15	5	20
Noradrenergic and Specific Serotonergic Antidepressants	13	6	19

to assess depression severity on a scale of 0-60. The scale is often used in the literature (see [Wat+22]) and Kjaergaard et al. deem MADRS, among others, "acceptable as screening instruments for MDD" [Kjæ+14].

Following the analysis, participants were divided into two groups based on their reaction to the treatment: those who responded positively and those who showed no significant improvement. This categorization is a foundation for using the data in machine learning models, particularly supervised learning, by assigning clear class labels to the subjects. Treatment success can be determined by two criteria: response, defined as a reduction of the MADRS score to less than half of its initial value after 28 days and remission, defined as a MADRS score falling below a certain threshold, typically between 8 and 10, after the same period. It is worth noting that in this thesis, response is used rather than remission. Consequently, the study included 176 participants, with 84 responding positively to the treatment and 92 non-responders. Remission rates can be considered to be even lower. If the participants would revert to the before-mentioned trial-and-error approach a result in line with [Pet+05], in which they state that remission rates plateau at approximately 50%, could be the case.

4.3. Data Acquisition

Participants underwent multiple recording sessions in the laboratory, each taking place on a different day. This thesis only examines the recordings from the second session. The initial session was scheduled for day zero, immediately before the start of the new treatment plan. The follow-up session took place on day 7 of treatment. During these sessions, recordings were recorded with 19 electrodes set up across the participants' scalps, following the international 10-20 system guidelines. Figure 2.1 shows the specific positions and labels of the electrodes. Each recording session lasted 10 minutes, with participants relaxed and their eyes closed. Through the preprocessing steps outlined below, the recordings varied in length. 4.5 shows the final lengths of the recordings. Any signs of sleepiness were quickly addressed to ensure alertness throughout the recording.

4.4. Data preprocessing

To ensure the EEG data's consistency and cleanliness across all participants, the following preprocessing steps were implemented by the project team:

- To maintain uniformity, electrodes that exceeded the standard count of 19 were excluded.
- Since sampling frequencies varied between 250 and 1000 Hz, recordings that were originally sampled at 1000 Hz were downsampled to 250 Hz by selecting every fourth sample.
- To reduce the impact of artefacts, each recording's initial 10-minute segment was trimmed, with the initial and final 30-second portions removed.
- To standardise the analysis of the EEG data, an average reference was applied to the EEG data.
- The data were bandpass filtered between 1 and 40 Hz.
- The data were separated into 2-second intervals or epochs and analysed for quality. Epochs polluted by significant artefacts (an epoch is considered unacceptable if it contains a proportion of 'bad' channels greater than 20% of the total, based on power deviations) were excluded. This step was crucial for preserving data integrity and allowing the analysis of continuous data segments without interference from artefacts.

4.5. Exploratory Data Analysis

Tables 4.1 and 4.2 already provide some insights into the data. This section provides deeper insights into the data to better understand its biases.

First we take a look at the distribution of age between males and females in Figure 4.1. Female patients show a Gaussian curve, whereas male patients seem to be more at the extremes of the age range.

Looking at the impact of age on the difference in MADRS scores we can see from Figure 4.2 that age does not show a significant effect on either male or female patients.

Figure 4.3 illustrates the development of MADRS scores, highlighting that responders typically begin with lower initial scores, particularly among males. Both male and female non-responders start with approximately the same MADRS scores, but by the final week, male non-responders exhibit higher scores and a lower response to treatment.

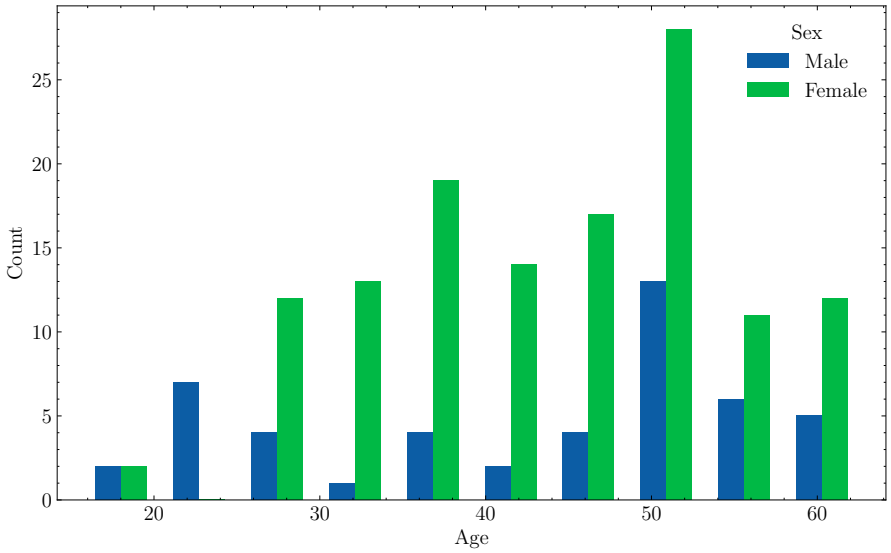


Figure 4.1.: Distribution of age for male and female patients.

Key Findings

- The dataset contains considerably more female than male patients (Table 4.1)
- Age does not have a significant impact on the change in MADRS scores for either gender (Figure 4.2).
- Responders generally have lower initial MADRS scores, especially males. By the final week, male non-responders have higher scores and show less improvement than females (Figure 4.3).

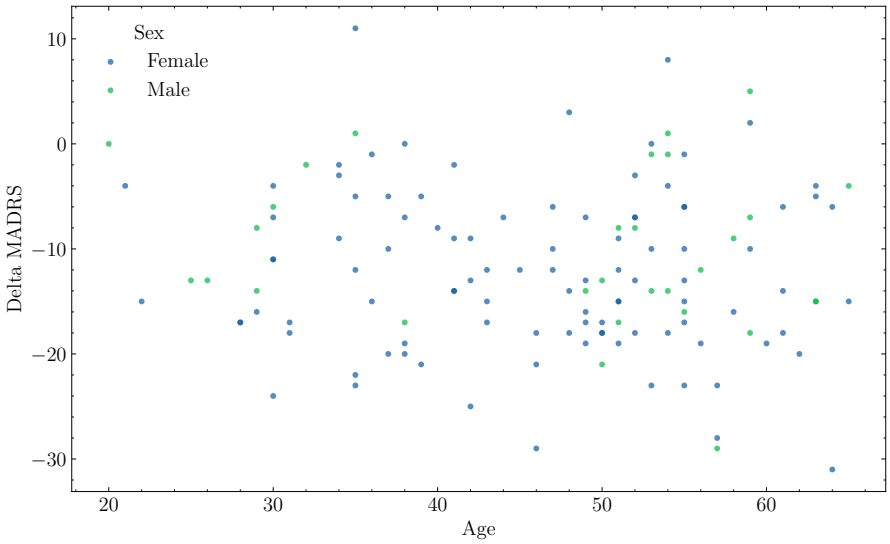


Figure 4.2.: Scatter plot of age and the MADRS-score delta between first and last visit.

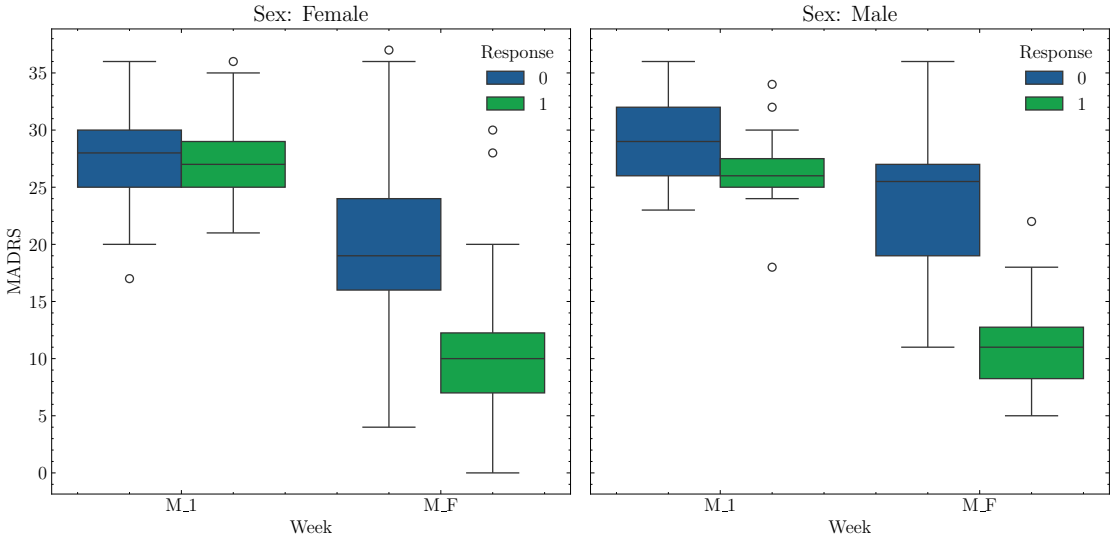


Figure 4.3.: MADRS scores on the first and last visit by sex.

5. Methodology

This chapter describes all of the methods used in this thesis. This overviews the whole workflow, as well as introduces the specific measures used for feature engineering and the statistical methods used to analyze the extracted features.

5.1. Workflow

The workflow adopted in this thesis follows the conventional phases of a data science pipeline, which include without limitation feature extraction, dataset splitting, feature selection, hyperparameter tuning, classification and evaluation. Below, we discuss key considerations and decisions made at each stage. Figure 5.1 illustrates an overview of the workflow.

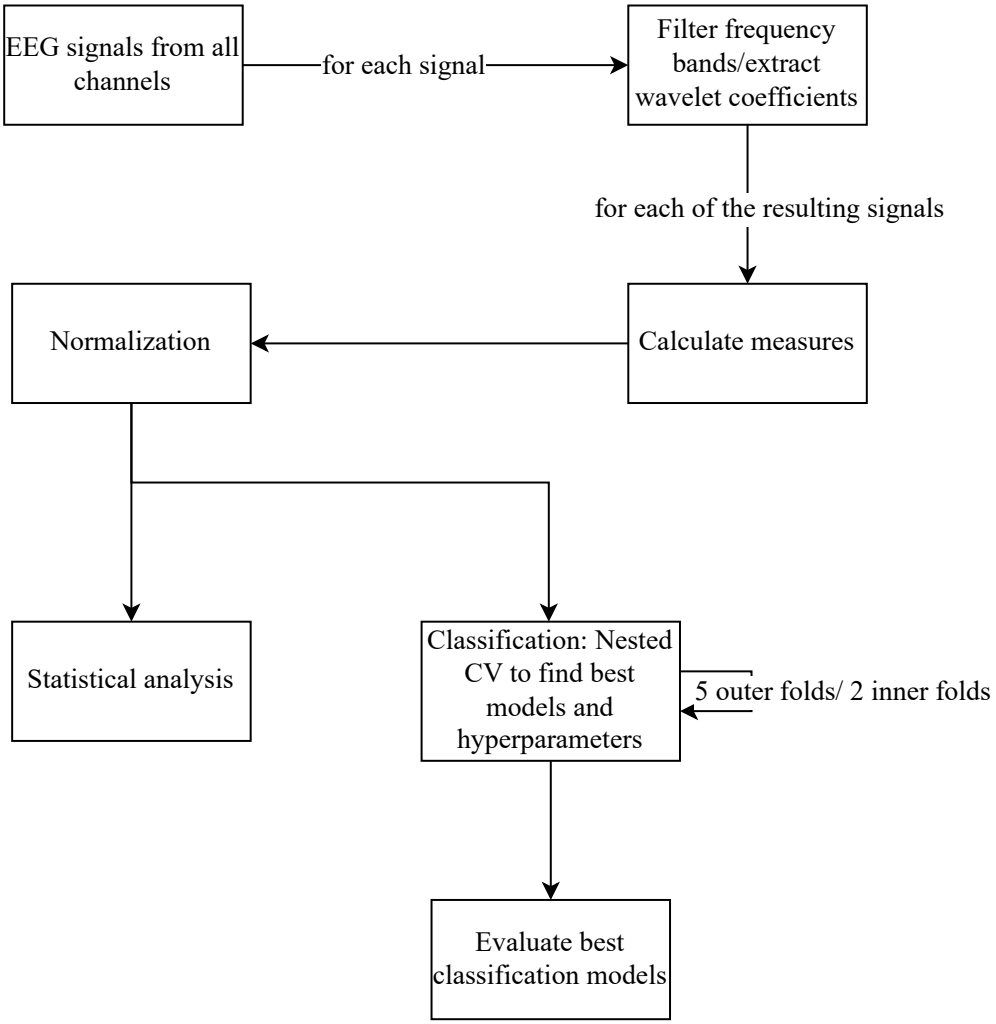


Figure 5.1.: Illustration of the workflow.

Feature extraction Following the preprocessing steps detailed in 4.4, the EEG data is prepared for feature extraction. Utilizing all 19 electrodes to capture brain dynamics results in 19 signals per patient. These signals are subdivided into wavelet coefficients as specified in 2.2.3, generating $19 \times 5 = 95$ signals.

Each signal is then segmented into 2-second intervals, considered to be static based on the discussion in 4.4. Measures, further elaborated in 5.2, are calculated for each segment and subsequently averaged to obtain one value per measure, wavelet coefficient and channel combination. Due to computations involving pairwise channel comparisons, the resulting feature matrix is high-dimensional, consisting of 176 rows for each patient and 6199 features. After removing features with low variance (threshold of 10^{-5}) we are left with a total of 5459 features.

Normalization Normalization is a crucial process in the data preprocessing pipeline, especially when dealing with a variety of entropy-based and other (non-)linear methods as they produce values in different ranges. By scaling the data to a common scale, we make sure that measures with larger scales do not dominate the learning process, which makes it a highly adopted practice [HMR19][Mum+17][Zhu+19][Cuk+20]. In this thesis we employ standardization or z-score normalization which ensures zero-mean and unit standard deviation in each feature. This process mitigates the above mentioned problem and in some cases leads to speedups in the training process for gradient-based classification models. Singh et al. also show performance improvements when scaling features and identify standardization as a generally well performing normalization technique [SS20].

Statistical analysis In this work, we incorporate statistical tests as a fundamental component of our analysis. This dual approach of classification and statistical analysis is motivated by the necessity to not only categorize data into distinct groups, but also to assess the statistical significance of the observed patterns within these categories. Employing statistical tests allows us to validate the reliability and robustness of the classification outcomes, which offers a more comprehensive understanding of the underlying data structures. This approach ensures that our findings are not a product of chance but are supported by quantifiable evidence, which, in turn, strengthens the validity of our conclusions in the broader context of the field.

Different statistical models come with distinct sets of assumptions about the data they analyze, which makes the selection of an appropriate model whose assumptions align with the characteristics of our dataset critical. In the context of our work, the dataset's complexity stems from the simultaneous investigation of multiple dependent variables, specifically measures, coefficients and channels, which establishes it as multivariate in nature. On top of that a hierarchical structure is evident in the nested arrangement of data: measures are nested within coefficients, which in turn are nested within channels.

We considered various statistical tests. In the end, Hotelling's T^2 test emerged as a valuable method for our analysis. Hotelling's T^2 test is an extension of the univariate t-test for multivariate data. It is designed to test the mean differences between two groups across multiple dependent variables [Hot31]. Hotelling's T^2 test therefore effectively

addresses the multivariate nature of the data.

The reliability of Hotelling's T^2 test depends on several assumptions. The data within each group should be approximately normally distributed, for which we employed the Shapiro-Wilk test to assess normality. Additionally, the test assumes that the variance-covariance matrices of the two groups are equal. We verified this through use of Box's M test, which is a commonly employed test for this purpose.

To model the three-way interactions between channels, coefficients and measures, we apply a univariate t-test to examine the effects of the unique channel-coefficient-measure combinations and use Hotelling's T^2 for multivariate cases (i.e. one- and two-way interactions).

Dataset splits Although our dataset is rather large in a medical context, it is considered undersized in the context of training machine learning models [Vab+19]. Given this limitation, a typical split into training, validation and test datasets is not viable, as we would loose too much data that is needed in the model learning process. In terms of model training we have to account for two key tasks: tuning optimal hyperparameters as well as selecting an appropriate model. Since these processes are distinct in their nature, trying to find solutions to both problems simultaneously would introduce a bias into the process [Ras20]. Therefore, a robust method for evaluating and selecting models is critical. Nested cross validation (CV) provides a solution to this challenge by offering a more accurate estimation of a model's generalization error in the case of sparse data. In nested CV, an inner loop is responsible for model selection and hyperparameter tuning and an outer loop evaluates the model's performance. This approach ensures that the evaluation of the models generalization performance is not optimistically biased by the model selection process, because selecting a model with hyperparameters tuned specifically on the same validation set may not lead to reliable generalization. The outer CV loop helps mitigate the risk of overfitting on the validation set, which is a common problem when the same data is used for both model tuning and evaluation. By using separate data subsets for hyperparameter tuning and model performance evaluation, the nested CV method provides a more reliable estimate of how well a model is expected to perform on unseen data, which in turn helps making a more informed decision in the selection of the most appropriate model [Ras20]. Figure 5.2 illustrates the process.

We demonstrate the effectiveness of this workflow by applying it to a seizure detection dataset. The results can be found in Section A.4.

Oversampling The dataset is heavily skewed towards female patients. In the experiments outlined in Chapter 6 we noticed a decrease in performance among male patients. This problem has also been outlined in another thesis work on the same EEG dataset and leads to significant improvements in the evaluation scoring if addressed accordingly [Kra24]. To do so, we employ Synthetic Minority Over-sampling Technique (SMOTE) which is a popular choice to over-sample the minority class. This then gives us 256 rows of data.

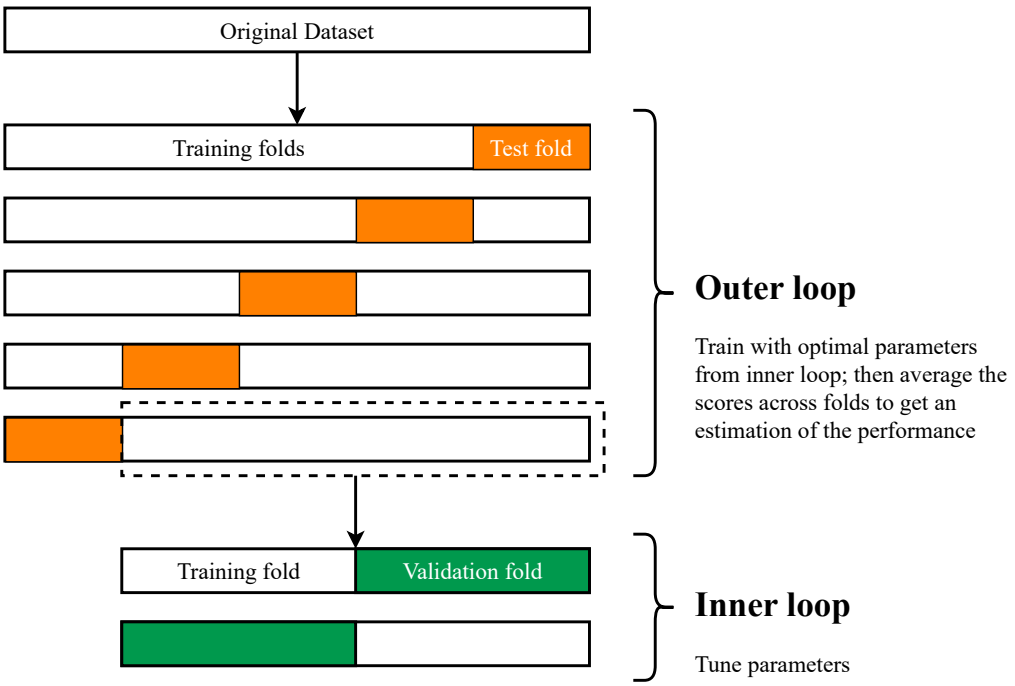


Figure 5.2.: Illustration of the process of nested CV (adapted from [Ras20]).

Feature selection Given the complexity of our feature matrix, which holds an extensive array of features, we implement feature selection during each iteration within the previously mentioned nested CV. Watts et al. highlighted the lack of comparison of feature selection methods and their influence on performance in the existing literature on depression treatment prediction [Wat+22]. In response, we conduct a comparison among three distinct feature selection techniques: minimum Redundancy Maximum Relevance (mRMR), ANOVA F-value and L1-regularization based feature selection.

The mRMR method selects features based on a criterion that jointly maximizes relevance to the target variable while minimizing redundancy among features. This technique aims to identify a set of features that are collectively most predictive of the outcome, ensuring that each selected feature provides unique information not offered by other features in the set [DP03].

The ANOVA F-value technique, on the other hand, assesses the individual predictive power of each feature using the statistical F-test. The proposed method evaluates the null hypothesis that no linear relationship exists between a feature and a target variable. Features with lower p-values, for which we reject the null hypothesis, are considered more significant and are selected for model training.

Finally, the L1-regularization based feature selection method, commonly known as the least absolute shrinkage and selection operator (LASSO), operates by adding a penalty

term to the model that is proportional to the absolute value of the coefficients. This regularization technique drives some feature coefficients to zero, effectively selecting a subset of the most important features while excluding those deemed less relevant. Lasso is particularly useful in high-dimensional data where multicollinearity is present, as it can shrink the coefficients of correlated features selectively [Tib96].

Classification To predict treatment outcomes, we employ a variety of classification methods. Given our limited dataset, we favor simpler algorithms that require less training data, while ensuring that these models can both mitigate overfitting and maintain interpretability. This interpretability is crucial, as it allows us to not only examine which features were selected, but also the importance of each feature. With these considerations in mind, we utilize the following models: Support Vector Machine (SVM), Random Forest (RF), eXtreme Gradient Boosting (XGB), Logistic Regression (LR) and Naive Bayes (NB). The implementations of these models are sourced from the scikit-learn library [Ped+11].

The hyperparameters were fine-tuned through Bayesian optimization, which employs Gaussian processes. This method assumes that the function values, describing the optimal hyperparameters, follow a multivariate Gaussian distribution. Essentially, Bayesian optimization with Gaussian processes predicts the performance of hyperparameters by estimating a probability model of the objective function. This approach iteratively updates the model with the goal of finding the optimal set of hyperparameters by evaluating where the model predicts a high likelihood of improvement [Moc89]. The hyperparameter spaces that were searched for each model are listed below (note that the Linear SVM utilizes the 'Linear Support Vector Classification' implementation of the scikit-learn library which allows to set different penalty terms):

- **SVM**
  - C: 1×10^{-6} to 1×10^{2}
  - Kernel: rbf, poly
- **Linear SVM**
  - C: 1×10^{-6} to 1×10^{2}
  - Penalty: l1, l2
- **Random Forest**
  - Number of Estimators: 10 to 500
  - Max Depth: 1 to 100
  - Min Samples Split: 2 to 50
- **XGBoost (XGB)**
  - Number of Estimators: 10 to 500
  - Max Depth: 1 to 100
  - Gamma: 1×10^{-6} to 1×10^{2}

  - Learning Rate: 0.01 to 0.9

- **Logistic Regression**
  - C: 1×10^{-6} to 1×10^{2}
  - Penalty: l1, l2

- **Naive Bayes**
  - Var Smoothing: 1×10^{-9} to 1×10^{-6}

In addition we also optimize the number of features to select by setting a hyperparameter space of 1 to 50.

Benchmark To establish a benchmark to evaluate the workflow presented in this thesis, we will compare it against a baseline and two sophisticated deep learning models. The benchmark includes a random guessing classifier and two variants of Convolutional Neural Networks (CNNs).

A random guessing classifier serves as a minimal benchmark, producing random outcomes within the set of possible classes. This establishes a baseline, indicating that any performance exceeding this threshold suggests that the model effectively learns patterns from the data rather than making random predictions.

Deep learning models, particularly CNNs, excel at learning intricate patterns in data due to their ability to automatically extract hierarchical features from raw inputs. Unlike traditional machine learning approaches that rely heavily on manually engineered features and often struggle with high-dimensional data, deep learning models can handle vast amounts of data and identify subtle, complex patterns that may be overlooked by conventional methods. This makes them especially powerful for tasks involving large datasets with intricate structures, such as image and signal processing [Alz+21].

In this study, we employ a simple CNN [FMI83][LeC+99] and a more advanced variant, the EEGNet architecture [Law+18]. Originally designed for image recognition, the simple CNN model requires our data to be represented in a compatible format. CNN layers expect input data in the form of three-channel images, typically with dimensions that include height, width and the number of channels. However, our EEG data is structured as time-series data, with dimensions representing samples, time and channels.

To transform the EEG data, we first ensure all sequences are of equal length by padding shorter sequences and trimming longer sequences. Next, we reshape the EEG data to add a dummy channel dimension, converting it from a three-dimensional format (samples, time, channels) to a four-dimensional format (samples, time, channels, 1). We then repeat this dummy dimension to create a three-channel format, resulting in dimensions that include samples, time, channels and three repeated channels. This allows the models to process the EEG data as if it were an image, facilitating the use of powerful image-based CNN architectures for our classification tasks.

The EEGNet architecture is a model architecture that is specialized for EEG data and tuned to effectively capture the unique spatial and temporal patterns present in EEG data. We take the implementation from the 'torchEEG' library [ZZL24].

5.2. Measures

As highlighted in Chapter 3, the current body of literature on treatment prediction lacks an extensive application of entropy-based measures. Consequently, this thesis focuses primarily on exploring entropy-based measures. This section provides a comprehensive overview of the measures employed, including their mathematical underpinnings and the rationale behind their selection.

As mentioned in Chapter 2, each entropy-based measure adjusts the general concept to a more nuanced and particular use case. ShannonEn, for example, is regarded as a widely applicable entropy variant, however, other entropy measures were specifically designed for time-series data, making them more desirable for this work.

Sample Entropy SampEn is an extension of ApEn intended to overcome its shortcomings. ApEn, developed by Pincus, measures the likelihood that similar patterns of observations will not be followed by additional similar observations. It quantifies the unpredictability of fluctuations in a time-series dataset [Pin91]. SampEn, introduced by Richman et al. [RLM04], excludes self-matches in its calculations, making it less biased. Looking at the formula of ApEn, it becomes apparent that the above explanation holds:

- Consider a time series of length N: $\{x_1, x_2, \ldots, x_N\}$.
- Form sequences of length m: Define vectors of length m from the time series:

$$u_i = \{x_i, x_{i+1}, \ldots, x_{i+m-1}\}, \quad i = 1, 2, \ldots, N - m + 1.$$

- Calculate the distance between vectors: For each pair of vectors u_i and u_j, calculate the distance between them using any appropriate distance measure (e.g., maximum norm, Euclidean distance, etc.).
- Calculate the proportion of similar vectors: For each vector u_i, calculate the proportion of vectors u_j (including self-matches) that satisfy:

$$C_i^m(r) = \frac{\text{Number of } u_j \text{ such that distance}(u_i, u_j) \leq r}{N - m + 1},$$

  where r is the tolerance level.
- Average over all vectors to obtain $\phi(m)$:

$$\phi(m) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln C_i^m(r)$$

  Note that the sum is performed first, and the logarithm is taken afterward.
- Repeat the process for sequences of length $m + 1$ to obtain $\phi(m + 1)$.

- Approximate Entropy is given by the difference:

$$\text{ApEn}(m, r, N) = \phi(m) - \phi(m+1)$$

  [Pin91].

SampEn has a slight variation to this process:

- Form sequences of length m: Similar to ApEn, define vectors of length m from the time series:

$$u_i = \{x_i, x_{i+1}, \ldots, x_{i+m-1}\}, \quad i = 1, 2, \ldots, N - m + 1.$$

- Calculate the distance between vectors: For each pair of vectors u_i and u_j, calculate the distance between them using any appropriate distance measure.

- Count the number of matching vectors, excluding self-matches: For each vector u_i, count the number of vectors u_j (**excluding** self-matches) that satisfy:

$$B_i^m(r) = \text{Number of } u_j \text{ such that distance}(u_i, u_j) \leq r, \quad j \neq i.$$

- Repeat the process for sequences of length $m + 1$ to obtain the count $A_i^{m+1}(r)$ for sequences of length $m + 1$:

  $A_i^{m+1}(r) = \text{Number of } u_j \text{ such that distance}(u_i, u_j) \leq r \text{ for vectors of length } m{+}1, \quad j \neq i.$

- Sample Entropy is calculated as the negative natural logarithm of the ratio of these counts:

$$\text{SampEn}(m, r, N) = -\ln\left(\frac{\sum_{i=1}^{N-m} A_i^{m+1}(r)}{\sum_{i=1}^{N-m} B_i^m(r)}\right)$$

  Note that the sum is performed inside the logarithm [RLM04].

Fuzzy Entropy FuzzyEn, introduced by De Luca and Termini [DT93] in 1993, measures the degree of fuzziness or uncertainty in time-series data. Like SampEn, FuzzyEn attempts to quantify the unpredictability of patterns within the series, but it does so by applying fuzzy logic principles, which allow for a more nuanced consideration of similarity between data segments [Aza+22]. The computation of FuzzyEn shows a crucial distinction to SampEn in terms of the propagation of similarity. While SampEn has a hard boundary, FuzzyEn returns a similarity score between 0 and 1.

- Similar to ApEn and SampEn, construct vectors of length m from the time-series data.

- For each pair of sequences, FuzzyEn calculates the degree of similarity using a fuzzy membership function, which is typically a function of the Euclidean distance between the sequences and a predefined tolerance r. This fuzzy membership function returns values between 0 and 1, representing the degree of similarity, where 1 indicates perfect similarity and 0 indicates complete dissimilarity.

- Average the fuzzy similarities for all pairs of sequences of length m to obtain $\mu_m(r)$. Repeat for sequences of length $m+1$ to obtain $\mu_{m+1}(r)$.
- FuzzyEn is given by the difference in the logarithms of the average similarities of sequences of length m and $m+1$, formally defined as:

$$FuzzyEn(m, r, N) = -\ln\left(\frac{\mu_{m+1}(r)}{\mu_m(r)}\right)$$

  [Aza+22].

Permutation Entropy While the measures until now focused on the predictability of sequences of similar amplitude values, with PermEn, Bandt et al. [BP02] aimed to assess complexity through the analysis of ordinal patterns or permutations of values within a time series. This method is based on the idea that the complexity of a system can be captured by examining the order relations among consecutive values in the time series. It is a simple and fast approach [Aza+22]. The formula can be broken down as follows:

- Divide the time series into overlapping segments, each containing m data points. For each segment, determine the order of the values to identify the specific rank pattern (permutation) of the data points.
- Identify all possible permutation patterns for sequences of length m. The total number of permutations is $m!$.
- Count the frequency of each permutation pattern in the time series to construct a probability distribution of these patterns.
- Calculate the Shannon entropy of this probability distribution to obtain PermEn:

$$PermEn(m) = -\sum_{i=1}^{m!} p(\pi_i)\log(p(\pi_i))$$

  where $p(\pi_i)$ represents the probability of occurrence of the i-th permutation pattern among all observed permutations [BP02].

Spectral Entropy With SpectralEn, Pan et al. introduced a unique approach within the spectrum of entropy-based measures in that it focuses on the frequency domain rather than the time domain [PCL09]. It evaluates how energy is distributed across various frequency components of the signal. A more evenly distributed power spectrum indicates higher complexity or irregularity, as it suggests the presence of many contributing frequencies. In contrast, a signal dominated by a few frequencies would exhibit lower complexity. The SpectralEn formula is computed as follows:

- Compute the power spectrum of the signal:

$$P(f) = |F(f)|^2$$

  where $F(f)$ is the Fourier transform of the signal at frequency f.

- Normalize the power spectrum to get a probability distribution:

$$P_n(f) = \frac{P(f)}{\sum_{f=1}^{N} P(f)}$$

where $P_n(f)$ is the normalized power at frequency f and N is the total number of frequencies.

- Calculate Spectral Entropy:

$$SpectralEn = -\sum_{f=1}^{N} P_n(f) \log_2(P_n(f))$$

[KK92].

Dispersion Entropy DispEn, proposed by Rostaghi et al. [RA16], is a more recent measure. It is based on the application of ShannonEn to sequences of symbols that resemble dispersion patterns. These patterns are calculated from the amplitude values of a time series by using value ranges and mapping these to classes, therefore not only the order of values is considered, but also the values themselves [RA16]. DispEn is said to be effective for short-time signals, has a fast computation time and is designed to alleviate the shortcoming of PermEn in terms of loosing amplitude information [ASR22][RA16]. The formula can be outlined as follows:

- Map the time series into C classes based on dispersion levels:

$$d(i) = \left\lfloor \frac{X'(i) - X'_{\text{min}}}{D} \right\rfloor + 1$$

where $d(i)$ represents the class of the i-th data point, X'_{min} is the minimum value of X' and D is the dispersion defined as $\frac{X'_{\text{max}} - X'_{\text{min}}}{C}$, with C being the number of classes.

- Form embedding vectors of length m with a delay of τ and map each vector to a unique symbol based on its dispersion pattern.

- Calculate the probability distribution of all symbols:

$$p_j = \frac{\#\text{ of embedding vectors mapped to symbol } j}{\#\text{ of embedding vectors}}$$

- Calculate Dispersion Entropy:

$$DispEn = -\sum_{j=1}^{C^m} p_j \log(p_j)$$

where C^m is the total number of possible symbols (or patterns) for embedding vectors of length m.

Slope Entropy In their 2019 study, Cuesta-Frau [Cue19] introduced SlopEn, a novel approach that is similar to the setup of DispEn and PermEn, however, rather than looking at permutations or dispersions, it evaluates the time series' slope first. It is designed to assess the complexity of time series data by quantifying the variability in the slope patterns between consecutive data points, which allows for more sensitive detection of dynamic changes. This essentially makes it a combination of entropy and the popular linear measure LZC [Cue19].

The idea of the formula is similar to PermEn:

- Calculate the slopes between consecutive data points:

$$S(i) = \frac{X(i+1) - X(i)}{\Delta t}$$

  where $S(i)$ is the slope between the i-th and $(i+1)$-th data points, $X(i)$ is the value of the time series at point i and Δt is the time interval between consecutive points.

- Normalize the slopes to obtain a slope time series:

$$S'(i) = \frac{S(i) - \mu_S}{\sigma_S}$$

  where μ_S is the mean of the slopes S and σ_S is the standard deviation of the slopes S.

- Divide the normalized slope time series into N non-overlapping windows of equal length m.

- For each window, calculate the local slope entropy as the Shannon entropy of the slope signs (positive, negative, or zero) within the window.

- The Slope Entropy is the average of the local slope entropies across all windows:

$$SlopeEn = \frac{1}{N} \sum_{k=1}^{N} H_k$$

  where H_k is the Shannon entropy of the slope signs in the k-th window, calculated as:

$$H_k = - \sum_{j \in \{+,-,0\}} p_j \log(p_j)$$

  and p_j is the probability of occurrence of each slope sign (positive, negative, or zero) in the window.

Symbolic Transfer Entropy Transfer entropy is a concept developed by Schreiber [Sch00] in 2000 to quantify information transfer between two systems or time series. Unlike traditional entropy measures, which assess uncertainty or disorder within a single system, transfer entropy focuses on the directional flow of information from one system to another. This makes it a powerful tool for identifying causal relationships in time series data, because it can capture the extent to which the future state of one variable reduces the uncertainty of another variable's future state, taking into account the history of both systems [Sch00].

For a clearer understanding, Figure 5.4 visually illustrates how transfer entropy is calculated using Venn diagrams. Staniek et al. [SL08] extended this concept and introduced symbolic transfer entropy (STE). The symbolic part of STE adds an additional layer of abstraction to the original transfer entropy framework. In STE, the signals are first converted into symbolic sequences, similarly to PermEn. This symbolic representation enables the computation of transfer entropy on a more abstract level, where the sequences of symbols, rather than raw data points, are analyzed to determine the information flow between systems, thus mitigating noise. [SL08].

- Convert the continuous or discrete time series into symbolic sequences:

$$S = \text{map}(Y)$$

  where Y is the original time series and S is the symbolic sequence.

- For two processes X and Y, form pairs of symbolic sequences S_X and S_Y.

- Define the history length k for Y and l for X to consider past values that may influence the transfer entropy.

- Compute the probabilities of symbol patterns:

$$p(s_{y,n+1}, s_{y,n}^{(k)}, s_{x,n}^{(l)}) = \frac{\#\text{ of times the symbol pattern occurs}}{\#\text{ of symbol sequences}}$$

  where $s_{y,n+1}$ is the next symbol in sequence Y, $s_{y,n}^{(k)}$ is the history of k symbols in Y up to time n and $s_{x,n}^{(l)}$ is the history of l symbols in X up to time n. k and l are often chosen to be the same value.

- Calculate STE:

$$T_{X\to Y}^{S} = \sum p(s_{y,n+1}, s_{y,n}^{(k)}, s_{x,n}^{(l)}) \log \frac{p(s_{y,n+1}|s_{y,n}^{(k)}, s_{x,n}^{(l)})}{p(s_{y,n+1}|s_{y,n}^{(k)})}$$

  [SL08].

Symbolic Influence Entropy In this thesis, we introduce the concept of influence entropy. Although closely related to transfer entropy and derived from its computation, this work is, to the best of our knowledge, the first to explore this variation. Unlike transfer entropy, which measures the extent to which one signal influences another, influence entropy assesses the total impact of a signal's past values, independent of the past values of any other signal. Consequently, we eliminate all shared information between multiple signals and focus exclusively on the information derived from a signal's past values.

The concept of influence entropy is visually explained in 5.5.

By taking the symbolic variation of the influence entropy we get symbolic influence entropy (SIE). The formula of SIE differs from STE in the following way:

$$IE^{S}_{X\to Y} = \sum p(s_{y,n+1}, s^{(k)}_{y,n}, s^{(l)}_{x,n}) \log \frac{p(s_{y,n+1}|s^{(k)}_{y,n}, s^{(l)}_{x,n})}{p(s_{y,n+1}|s^{(l)}_{x,n})}$$

, note $p(s_{y,n+1}|s^{(l)}_{x,n})$ rather than $p(s_{y,n+1}|s^{(k)}_{y,n})$ in the denominator.

Measures used as comparison In assessing the effectiveness of entropy-based metrics, we utilize a diverse array of measures, both non-linear and linear, to serve as benchmarks. Detailed mathematical formulations and further discussions on these measures are available in the cited literature. The following is a brief overview:

- **Band power**: Band power measures the power of the signal within specific frequency bands, namely alpha, beta, gamma, delta and theta bands. Band power is valued for its simplicity and interpretability, however, it is often outperformed by more sophisticated measures [JPB14] [Vid+09] [HMR13].

- **Hjorth parameters**: Hjorth Parameters, introduced by Hjorth [Hjo70] in 1970, include Activity, Mobility and Complexity, offering a straightforward approach to analyze EEG signals. Activity measures the signal's variance, indicating power. Mobility assesses the mean frequency, reflecting signal speed and Complexity assesses the signal's frequency change, capturing its dynamical behavior [Hjo70]. Applications in depression detection, for example, as described by Sakib et al., underline its utility. [SIF23].

- **SASI**: Hinrikus et al. [Hin+09] introduced a more advanced linear measure with SASI. It analyzes the power balance between two EEG frequency bands, above and below the alpha spectrum's peak, within a single channel [Hin+09]. Bachmann et al. use it along with HFD to detect depression [Bac+13].

- **LZC**: According to Lempel and Ziv, LZC measures algorithmic complexity by counting the number of distinct segments in a signal and is simple to compute [LZ76]. It is a popular choice as a non-linear measure as it has been employed successfully numerous times in combination with EEG data [Bac+15] [Li+08]. Méndez specifically analyzed treatment response and found a decrease in LZC values after treatment [Mén+12].

- **HFD**: In 1988 Higuchi introduced HFD, a non-linear measure that calculates the fractal dimension of time series data, offering insights into the irregularity and complexity of EEG patterns [Hig88]. HFD is notable for its efficiency and minimal data requirement, especially in comparison to DFA and CD, making it suitable for both clinical and research settings [Bac+13]. Applications are found in many areas of EEG analysis, including depression detection [Bac+13] [HMR13].

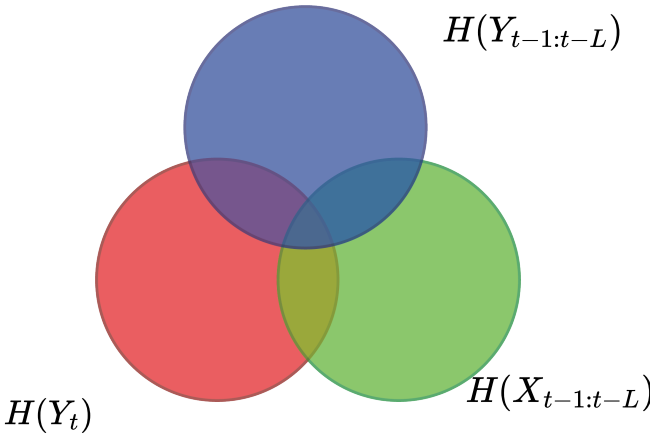


Figure 5.3.: Assuming two signals X and Y transfer entropy of Y given X is calculated using $H(Y_t)$, the entropy of Y, $H(Y_{t-1:t-L})$, the entropy of the past values of Y and $H(X_{t-1:t-L})$, the entropy of the past values of X (adapted from [Khu23]).

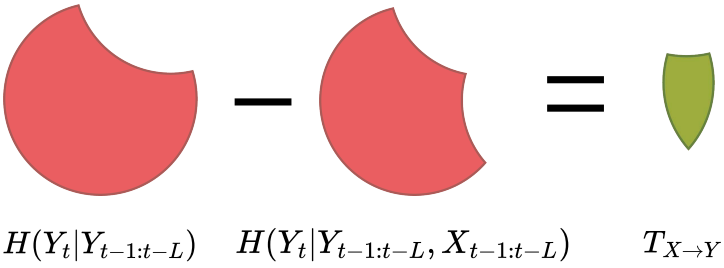


Figure 5.4.: Transfer entropy is calculated by subtracting the entropy of Y_t given its past $Y_{t-1:t-L}$ and the past of the variable X_t from the entropy of Y_t given its past $Y_{t-1:t-L}$ without looking at X (adapted from [Khu23]).

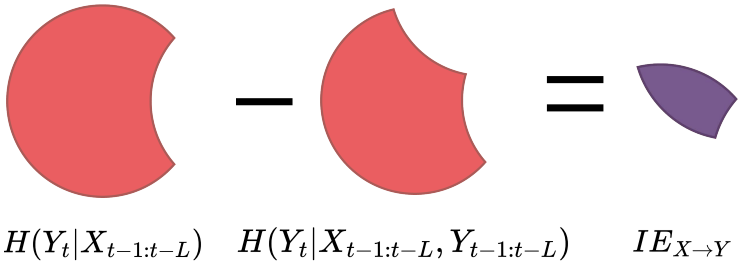


Figure 5.5.: Influence entropy is calculated by again subtracting the entropy of Y_t given its past $Y_{t-1:t-L}$ and the past of the variable X_t but this time from the entropy of Y_t given the past of X without considering the past of Y.

6. Results

In the following chapter, the results of the experiments conducted in this thesis are presented. The results can be divided into a preliminary analysis of the feature values for the two classes, a statistical analysis and classification. Chapter 7 connects the three analysis in an overarching discussion of the results.

6.1. Preliminary feature separation analysis

To obtain a general overview of the effectiveness of the employed measures we examine the mean absolute difference of feature values between the two patient groups.

The heatmap in Figure 6.1 shows the mean absolute difference for all single-channel or unilateral measures split by wavelet coefficients. The T6 channel generally exhibits high difference values across multiple measures, with F4 and O2 showing similar trends. Regarding coefficients, A4 displays extreme values, with low differences for most combinations but some high differences for specific channel and measure combinations. D4 has the lowest difference values across almost all channels and measures, except for the T6 channel. The Fp1, F7, P4, Pz, T3 and T4 channels do not perform as well, showing lower absolute difference values across almost all coefficient and measure combinations. Notably, DFA shows higher difference values in D1 and D3, but is zero across all channels in A4 and D4. Some specific combinations of channels, coefficients and measures show spikes in their absolute mean difference values i.e. for FuzzyEn for coefficient A4 and channel T3.

As we also employ measures that calculate values between two channels we analyze their mean absolute difference values separately. As before, we split each signal into five wavelet coefficients. We then calculate two multi-signal measures, namely symbolic transfer entropy (STE) and, as introduced in Chapter 5, symbolic influence entropy (SIE) between all possible channel-pairs for each wavelet coefficient separately. Note that transfer entropy is a non-linear measure that quantifies how much knowing the past states of one time series (in this case an EEG channel) can reduce the uncertainty about the future state of another series. Influence entropy acts in a similar manner and quantifies the amount of information that is purely given by the past of one signal, irrespective of information from another signal. By comparing STE and SIE values between responders and non-responders, we assess differences in how information flows between specific brain areas in the two groups. The mean absolute difference between channel pairs is then used to identify which connections show the most significant variation between responders and non-responders.

To visually represent these findings, we plot the layout of all 19 EEG electrodes across five separate plots, one for each wavelet coefficient. For each measure, we test three parameter combinations by varying the dimension and delay parameters, which therefore results in six plots, each with five subplots. The connections between channels were represented by lines, with line thickness and color encoding the magnitude of the difference in entropy between responders and non-responders, averaged across all patients. A threshold of 0.2 was set to filter out smaller, less significant differences, thereby allowing us to focus on the most prominent connections. This threshold, set at approximately half of the maximum observed values (around 0.38), ensures that the plots remain uncluttered and highlight the most meaningful patterns. As mentioned, we explore various parameter settings in the STE and SIE calculation. This systematic exploration enables us to evaluate the consistency of our findings across various settings of the mentioned measures, allowing us to validate whether our results are robust and unaffected by adjustments in the calculation parameters. We also identify the top 20 channel pairs with the highest mean absolute entropy value differences across all parameter combinations and wavelet coefficients. These channel pairs are annotated in the plots, marking them as critical connections that consistently differentiate the two groups for a specific wavelet coefficient and measure configuration. The plots are shown in Figure 6.2.

The plots reveal a clear trend where certain channels act as "nests", exhibiting several connections, while others exhibit minimal connectivity. These highly connected channels likely represent key brain regions involved in neurophysiological processes that differentiate responders from non-responders. The dense connectivity around certain nodes suggests that these brain regions may serve as hubs for information processing or communication that are differently modulated by treatment response.

Generally, no single measure, wavelet or channel exhibit large values across all combinations. There is a large variation in channels that can be seen as "nests" across different wavelets and parameter settings. Still we can observe some trends in channels that occur more frequently than others. Tables A.1 and A.2 offer a detailed analysis of the connectivity differences between depression treatment responders and non-responders. These tables present the results of two key metrics, weighted connection count and top percentile mean, calculated across different wavelet coefficients for each channel.

The weighted connection count sums the fourth power of the mean differences in entropy values, emphasizing significant connectivity variations. This metric should balance counts and weigths of the connections.

The top percentile mean focuses on the top 10% of the largest mean differences, highlighting channels with the most substantial variations in connectivity. This measure identifies channels where the most pronounced entropy differences occur, regardless of the number of connections.

Channels with a high weighted connection count might be important for broad, widespread brain connectivity, whereas those with a high top percentile mean could be vital for specific, powerful connections that may have a significant impact on how treatment affects the brain.

Combining the plots and the tables, we make the following observations:

- Lower values for the dimension and delay parameters in multi-channel measures result in slightly higher difference values than larger ones. This suggests that lower dimensions capture shorter, more direct patterns in brain activity, which may be more relevant for distinguishing responders from non-responders. Similarly, shorter delays might highlight more immediate interactions between brain regions.

- No single channel consistently shows high absolute difference values between responders and non-responders across all wavelets, indicating that treatment response is linked to a distributed neural network rather than localized to specific regions.

- The D2 wavelet coefficient shows larger weighted connection counts, indicating broader network involvement in the upper alpha and lower beta ranges, which reflects widespread changes in cognitive control. In contrast, D4, A4, and D1 exhibit larger top percentile mean values, highlighting fewer but more significant connections in the delta and upper beta/lower gamma ranges, pointing to deeper, targeted changes in neural synchronization and mood regulation crucial for distinguishing responders.

- Channel T6 exhibits consistently large difference values between the two groups across different wavelet coefficients (A4, D1, D2, D3), particularly in all configurations of STE. T6 both shows high number of connections and large difference values in its connections. The recurrence of T6 in multiple measures and coefficients indicates that this channel may play a crucial role in the information flow that distinguishes responders from non-responders.

- While channels like O2 or F3 visually appear as nests with many connections in the plots indicating higher difference values, C3 and C4 have fewer connections but with larger difference values, indicating that these channels may exhibit changes in specific connections between channels rather than broader changes.

- Although not consistent across multiple wavelet coefficients or parameter configurations, STE-Del-5-Dim-7 shows different "nests" for different wavelet coeffcients, namely F8 for A4, T5 and T6 for D3, O2 for D2 and Pz, T6, F3 and F4 for D1.

Generally, there is a trend towards higher difference values in measures with lower parameter values for embedding dimension and delay. In these measures with lower parameter values, the larger difference values are observed in A4 and D4. Conversely, for higher parameter values, the trend is opposite with coefficients D1 and D2 having larger difference values. There is a greater difference in values for multiple feature values that include the T6 channel, as it shows many connections across several coefficients and parameter settings. Similar connectivity "hubs" can be found for the C4, F8, T3, O2 and Pz channels. Although these are often specific to particular coefficient or measure configurations, as illustrated in Figure 6.1, their difference values tend to exhibit larger spikes than those associated with the T6 channel.

Key Findings

- T6 shows high absolute value differences in both single- and multi-channel measures and across multiple coefficients. This indicates a relation between the temporal lobe in the right hemisphere treatment response.
- Although Channels C3 and C4 have fewer connections, they show larger differences in specific connections, suggesting localized changes. In contrast, channels like O2 or F3, with more connections and significant differences, indicate broader network-level changes.
- The A4 coefficient does not show high separability for single-channel measures but exhibits larger values for multi-channel settings. This suggests that treatment affects basic brain networks involved in rest and recovery, which are important for stabilizing brain function in depression.
- D1 and D2 show higher differences than D3 and D4 for single-channel measures, while for multi-channel measures, we observe larger weighted connection count values for D2, emphasizing widespread brain connectivity, while A4, D4, and D1 show larger top percentile mean values, indicating specific, powerful connections.

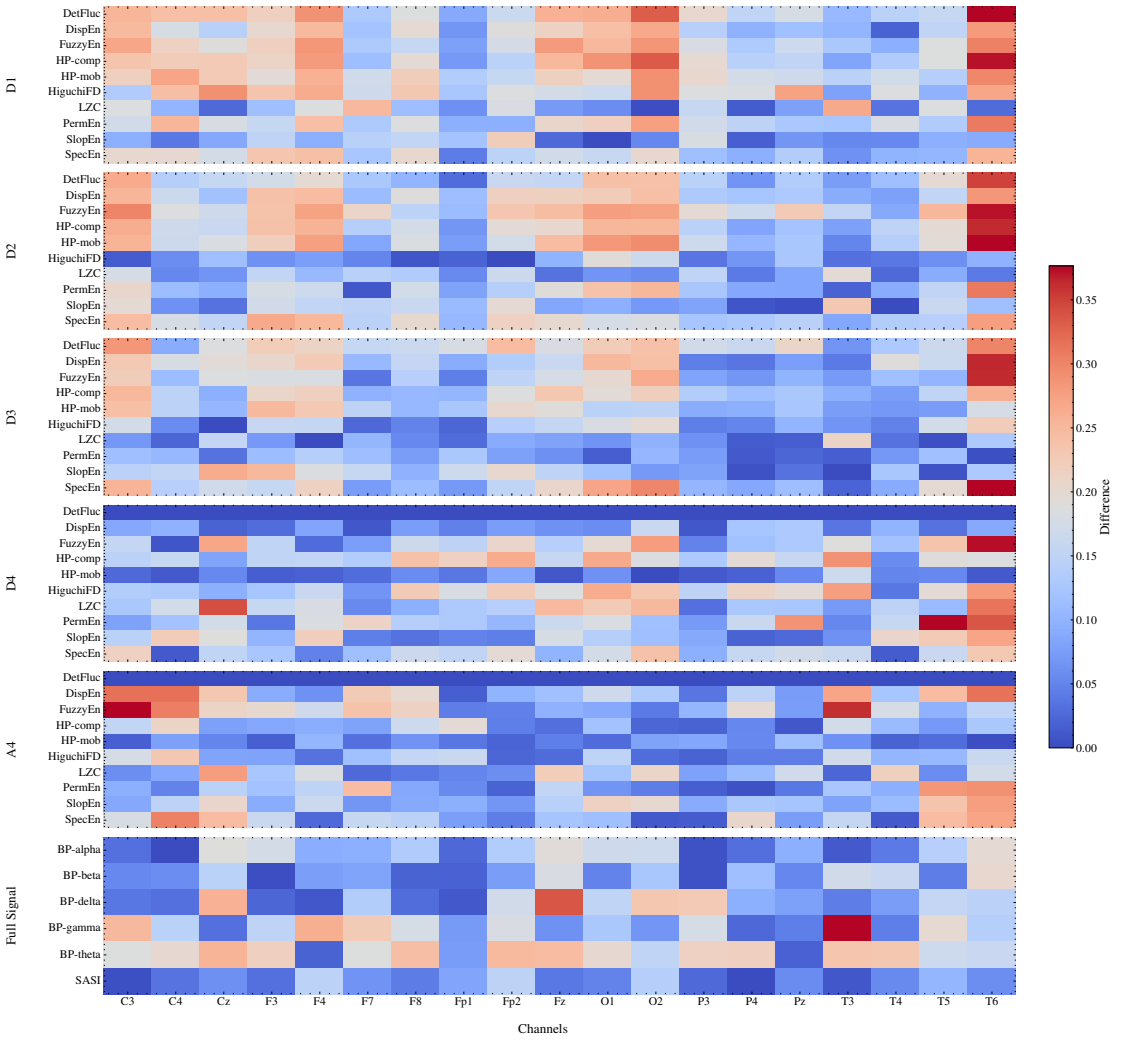
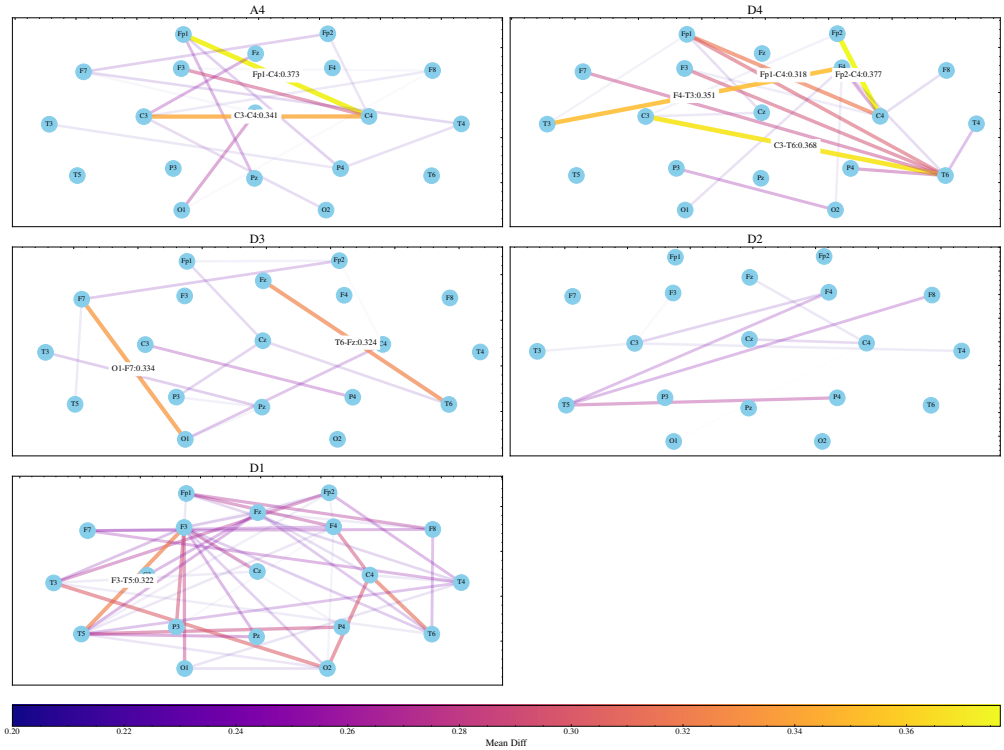
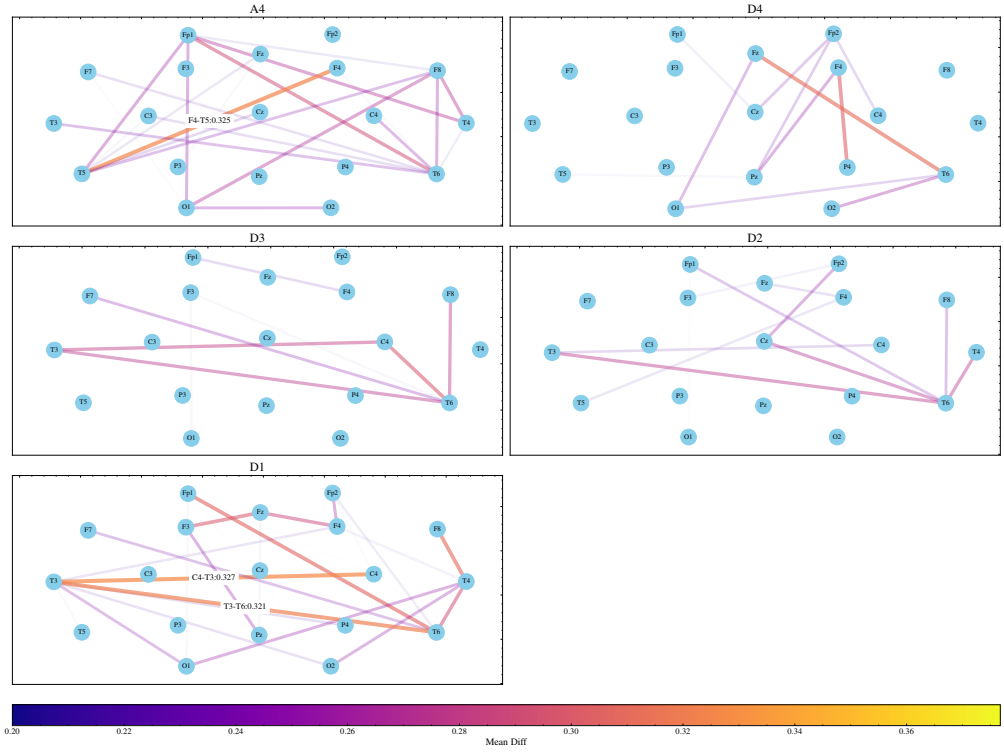


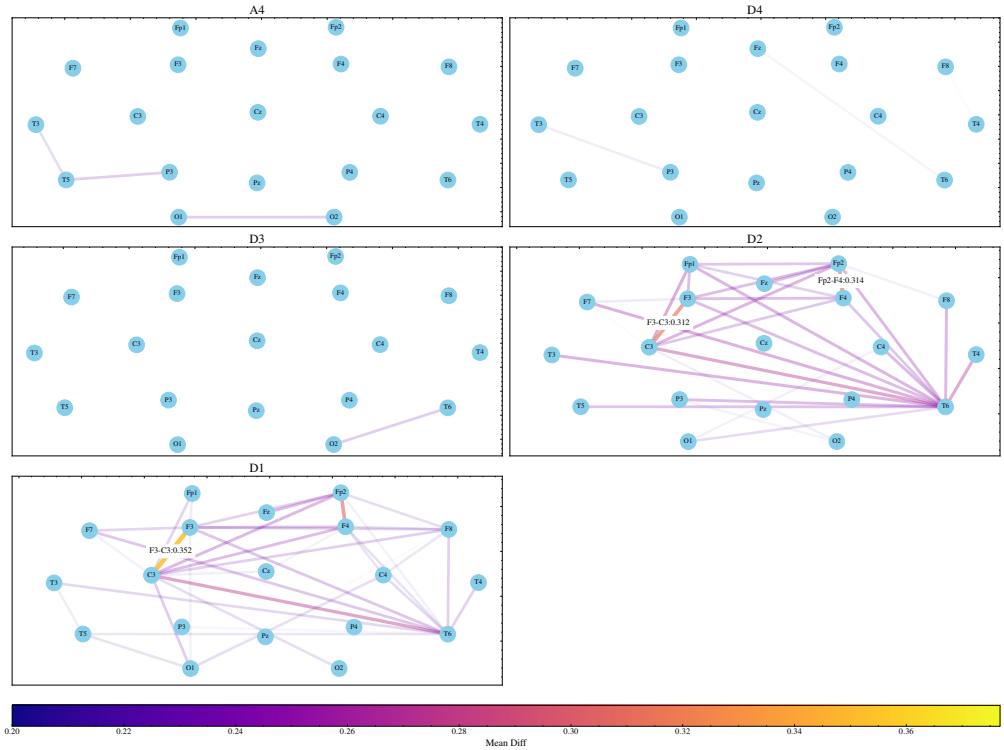
Figure 6.1.: Heatmaps presenting the mean difference scores between the two patient groups across each electrode within each wavelet coefficient (when applicable). X-axis represents the channels, Y-axis represents the measures divided into coefficients.



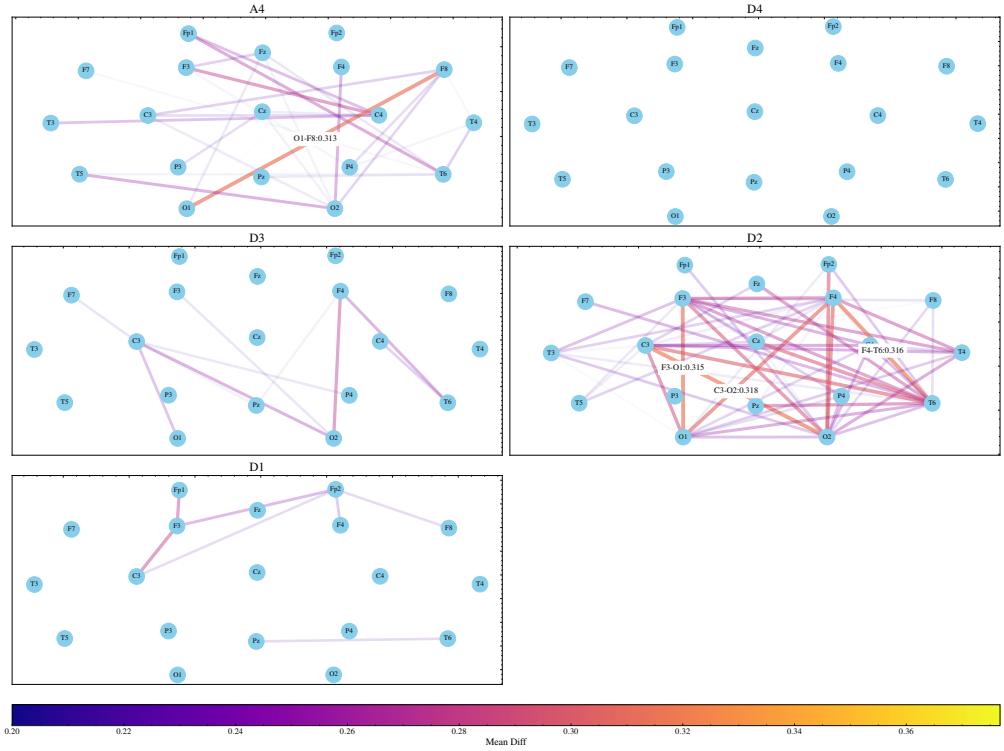
(a) STE with delay 2 and dimension 3.



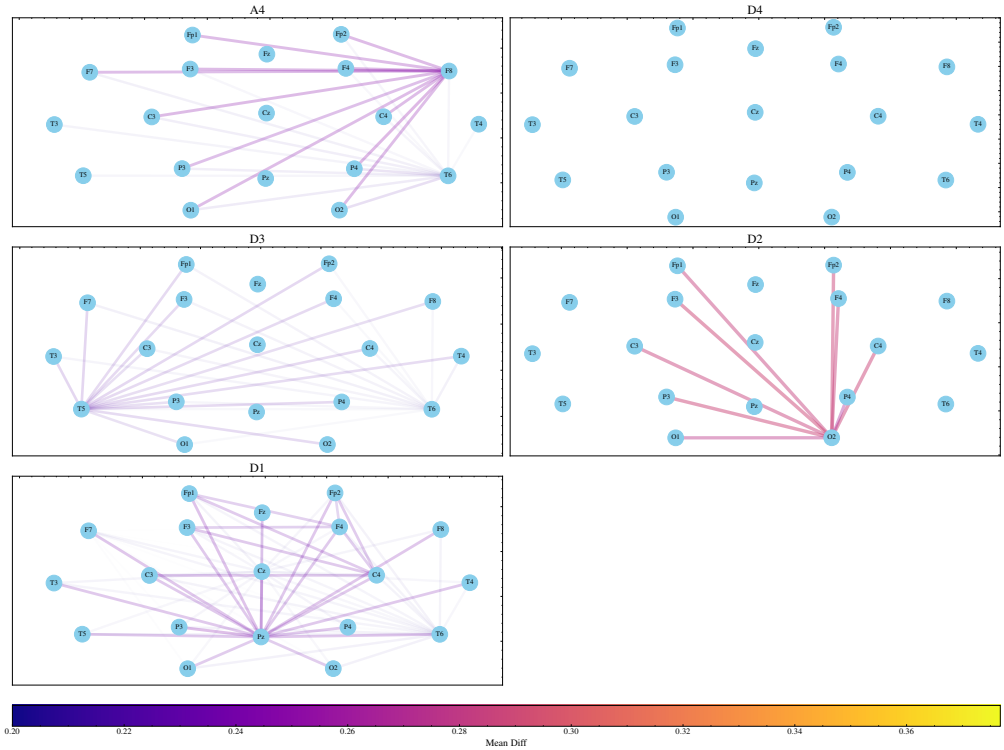
(b) SIE with delay 2 and dimension 3.



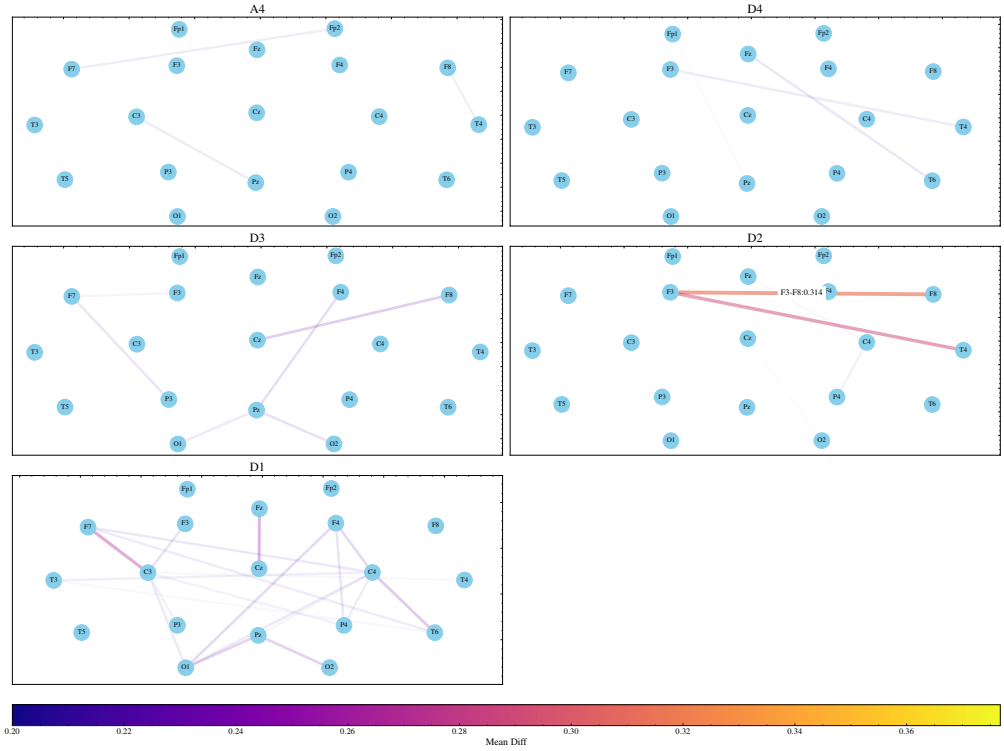
(c) STE with delay 3 and dimension 5.



(d) SIE with delay 3 and dimension 5.



(e) STE with delay 5 and dimension 7.



(f) SIE with delay 5 and dimension 7.

Figure 6.2.: Comparison of mean differences in multi-channel measures (STE & SIE) with varying embedding dimensions and delays. A threshold of half of the maximum difference value was applied to avoid clutter. The top 20 connections with the highest mean differences are labeled with exact values.

6.2. Statistical analysis

The following section presents the results of our statistical analysis, highlighting key findings from the comparison of feature value distributions between responders and non-responders. The analysis is structured to provide insights at different hierarchical levels, starting with single interactions (one-way interactions) and progressively incorporating more variables (two-way and three-way interactions). This approach enables us to track how the separation between responders and non-responders evolves across different channels, including how the significance of these differences changes when additional hierarchical levels, such as coefficients and measures, are considered. The tables are organized to reflect this structure, allowing us to identify the impact of individual channels, coefficients and measures, as well as their combinations, on the separability between groups. Ultimately, we aim to determine which patterns remain consistent across multiple levels and which combinations of variables perform the best when considering all levels of the hierarchy.

6.2.1. Hotelling T^2 & t-Test

As described in Section 5.1, we conduct a univariate t-test for the unique channel-coefficient-measure combinations and Hotelling's T^2 test to assess the differences among channel, coefficient and measure combinations for the multivariate cases whenever one or more variables can take on multiple values (i.e., channel and coefficient are fixed but measure is variable). We report the top 10 most significant results for each hierarchical level at a significance level of $\alpha = 0.05$ in Table 6.1, with the full set of significant results displayed in Table A.3.

Table 6.1.: Results of the t-test and Hotelling's T^2 test for all three-way interactions between channels, coefficients, and measures. The most significant combinations, as well as variables and patterns that appear frequently across different tests and hierarchical levels, are marked in bold.

Combination	**T^2-statistic**	**p-value**
C3-**T6**	52.45	0.0162
C4-**F8**	50.91	0.0212
T4-Pz	49.45	0.0272
C4-O2	48.68	0.0311
C4-T4	48.25	0.0335
Fp1-Cz	46.41	0.0454
O2-**F8**	46.05	0.0482
Fp2-F4	45.90	0.0493

(a) Channel

Combination	**T^2-statistic**	**p-value**
DispEn	313.98	0.0280
LZC	306.38	0.0359

(b) Measure

Combination	**T^2-statistic**	**p-value**
Fp1-C4_**A4**	24.42	0.0004
T5_A4	29.05	0.0019
P4-Pz_D2	20.13	0.0021
O1-**F8_A4**	18.61	0.0038
C3-**T6**_D4	17.85	0.0050
Fp1-Cz_D2	17.02	0.0069
O2-**F8_A4**	15.35	0.0127
F4-O1_D4	15.23	0.0133
C4-T6_**A4**	15.18	0.0136
C3-O1_D1	14.83	0.0154

(c) Channel Coefficient

Table 6.1.: Results of the t-test and Hotelling's T^2 test for all three-way interactions between channels, coefficients, and measures. The most significant combinations, as well as variables and patterns that appear frequently across different tests and hierarchical levels, are marked in bold.

Combination	**T^2-statistic**	**p-value**
Fp1-Cz_STE-Del-2-Dim-3	26.26	0.0002
C3-**T6_STE**-Del-2-Dim-3	20.89	0.0016
C4-**F8_STE**-Del-5-Dim-7	18.33	0.0042
P3-**F8_STE**-Del-5-Dim-7	18.33	0.0042
F7-**F8_STE**-Del-5-Dim-7	18.33	0.0042
C3-**F8_STE**-Del-5-Dim-7	18.31	0.0042
O2-**F8_STE**-Del-5-Dim-7	18.21	0.0044
Fp1-F8_STE-Del-5-Dim-7	18.21	0.0044
T5_PermEn	18.18	0.0044
F3-**F8_STE**-Del-5-Dim-7	18.18	0.0044

(d) Channel Measure

Combination	**T^2-statistic**	**p-value**
A4_PermEn	38.50	0.0252

(e) Coefficient Measure

Combination	**t-statistic**	**p-value**
C3-**T6**_D4_**STE**-Del-2-Dim-3	3.83	0.0002
Fp1-Cz_D3_**STE**-Del-2-Dim-3	-3.70	0.0003
Fp1-Cz_D2_**STE**-Del-2-Dim-3	-3.70	0.0003
Fp1-T3_D3_**STE**-Del-2-Dim-3	-3.23	0.0015
T5_A4_PermEn	-3.21	0.0016
F4-**F8_A4_STE**-Del-5-Dim-7	3.16	0.0019
P3-**F8_A4_STE**-Del-5-Dim-7	3.15	0.0019
O2-**F8_A4_STE**-Del-5-Dim-7	3.14	0.0020
O1-**F8_A4_STE**-Del-5-Dim-7	3.14	0.0020
F7-**F8_A4_STE**-Del-5-Dim-7	3.13	0.0020

(f) Channel Coefficient Measure

Starting with the one-way interactions, no single coefficient reached the significance level. However, significant results were observed in combinations arising from multi-channel measures such as STE or SIE. Here, T6, F8, C4, Cz, O2 and Pz are the most prominent. Among the measures, only DispEn and LZC showed significance, which suggests that most measures are highly dependent on specific channels and coefficients.

When incorporating at least one additional variable, p-values decreased significantly. In the channel-coefficient combinations, mostly combinations stemming from multi-channel measures reached the significance threshold, with T5 in combination with A4 also showing a low p-value. A similar trend can be observed in the channel-measure combinations,

where T5 appears again. Notably, almost all top combinations in this category involve STE values, with the F8 channel being the most prominent.

Incorporating all three variables resulted in even lower p-values, thus reaffirming the validity of the proposed workflow. STE remained the most dominant measure, except for the combination of channel T5 in the A4 coefficient using PermEn, which had a similarly low p-value. Channel-wise, F8, Cz and Fp1 stood out the most, while C4 seems to be less important. The combination of channels C3 and T6 in the A4 coefficient yields the lowest p-value of 0.0002.

Key Findings

- **Lower p-values for multiple variables:** Considering the interaction of channel, coefficient, and measure generally results in lower p-values, indicating a strong hierarchical structure and suggesting that the combined effect of these variables better captures the differences between responders and non-responders.
- **Combinations with multi-channel measures preferred:** Multi-channel measure combinations are observed more frequently than single-channel ones, implying that treatment response involves complex, distributed networks rather than isolated brain regions.
- **Frequent channels:** Channels F8, T6, Fp1 and Cz frequently appear in significant combinations.
- **Strong influence of STE:** STE shows up prominently across multiple combinations, indicating its strong influence.
- **Significance of the A4 coefficient:** The A4 coefficient is particularly significant, especially when combined with the F8 channel and STE.
- **Lowest p-value combination:** The combination of channels C3-T6 with the D4 coefficient and STE yields the lowest p-value (0.0002).

6.3. Classification

In this section, we evaluate the performance of various classification models using both individual and a comprehensive set of measures to determine which combination yields the best results. In addition, we examine the effects of different feature selection techniques and identify the features that are selected most frequently by each method under the different settings. To put these results into perspective, we report the results of the benchmark methods described in Section 5.1.

6.3.1. Benchmark results

This section presents the performance outcomes of the benchmark classifier models on our EEG dataset. The models, previously introduced and described, were evaluated based on their accuracy and F1 scores. The data were preprocessed with the same steps outlined in Section 5.1 except for the feature extraction and feature selection steps. We employ a stratified 5-fold CV for each classification model.

Table 6.2.: Classification scores for the benchmark models.

Model	**Training**		**Testing**	
	Accuracy	**F1**	**Accuracy**	**F1**
Random Guessing Classifier	0.483 ± 0.041	0.413 ± 0.042	0.510 ± 0.080	0.443 ± 0.096
Simple CNN	0.887 ± 0.092	0.906 ± 0.079	0.510 ± 0.055	0.639 ± 0.057
EEGNet	0.478 ± 0.005	0.951 ± 0.025	0.473 ± 0.055	0.472 ± 0.056

The scores are reported in Table 6.2. As expected, the random guessing classifier obtains scores related to random guessing in a binary classification problem. On the other hand, EEGNet [Law+18] and the simple CNN [FMI83][LeC+99] model exhibit overfitting as they achieve high accuracy and F1 scores on the training data but fail to generalize, with test scores that are also close to random guessing. The test F1 score of the simple CNN demonstrates a notable increase, however, the overall accuracy remains comparable to that of random guessing. This suggests that the model is overfitting to the extent that it accurately predicts responders while neglecting non-responders.

6.3.2. Feature selection

As mentioned earlier, a comprehensive comparison of the performance of different feature selection techniques is lacking in the current literature in terms of depression EEG data. We therefore run multiple iterations of each experiment, with each iteration using a different feature selection method as the basis of the experiment setup. As mentioned in Chapter 5, we employ ANOVA F-value, mRMR and L1-based feature selection.

6.3.3. Single Measure

To begin, we examine each measure on its own to determine which singular measure yields the best generalizable results. On top of that, we also show which channels and coefficients are selected the most.

Table 6.3 displays the accuracy and F1 scores obtained from a single iteration of the nested CV, as previously described in Section 5.1. Although it is a single iteration, the nature of the nested CV makes the results stable and generalizable. We report the mean and standard deviation of the training and testing scores obtained using the best model

and feature selection technique for each single measure. The experimental setup involved a 5-fold CV in the outer layer, resulting in five sets of scores, which were averaged to present the final scores in the table.

Although training accuracy and F1 scores are generally high, the corresponding test metrics are significantly lower, suggesting that the models tend to overfit the training data. This overfitting persisted despite exploring a wide range of regularization parameter values. Increasing the regularization parameters beyond a certain threshold led to a substantial decline in both the training and test metric values without significantly decreasing overfitting.

However, some techniques were notably effective. Specifically, STE with a delay of 2 and embedding dimension of 3 as well as DispEn, Hjorth complexity and HFD, which utilize mRMR and L1 for feature selection and SVM for classification, demonstrated the highest mean F1 scores on the test data. This result aligns with the statistical analysis presented earlier, in which STE and DispEn had some of the lowest p-values. We report the hyperparameters used in Table A.4. All pipelines share a rather large C value, which helps reduce overfitting. In this setting, classification pipelines using L1 or mRMR feature selection scored the highest most of the time.

Figure 6.3 presents a view of the channel and coefficient relationship. It shows the number of times a channel was selected in one of the top estimators in terms of their mean F1 score on the test folds. Given the large number of estimators, we restricted the visualization to the top ten estimators. Figure 6.4 shows the selection frequency of the coefficients. The A4 coefficient is the most frequently selected coefficient, followed by D3 and D4, with the remaining coefficients having roughly similar frequencies. In terms of channel selection, no single channel stands out as the most frequently selected across all coefficients. Channels T3, T5, T6, Fp2 and F3 are chosen more often than others, but the difference is not substantial. Notably, dual-channel channel combinations such as those in the STE and SIE computations are especially frequently selected in the A4 and D3 coefficients.

In Figure A.1, we present the interplay between feature selection and the best performing models and outline the most important results. For each measure and for each feature selection technique, the best performing models are shown sorted by their frequency. SVM models are the most frequently selected, regardless of the feature selection method. LR models, however, perform best when combined with F-Test or L1 feature selection. Linear SVM models tend to perform best with L1 feature selection.

Table 6.3.: Mean and standard deviation of train and test scores of the best model and feature selection technique per measure. Each run consists of five CV folds. The highest overall F1 score is highlighted in bold.

Measure	**Feature Selection**	**Model**	**Training**		**Testing**	
			Accuracy	**F1**	**Accuracy**	**F1**
BP-alpha	mRMR	SVM	0.878 ± 0.018	0.850 ± 0.026	0.664 ± 0.058	0.583 ± 0.105
BP-beta	mRMR	SVM	0.915 ± 0.040	0.892 ± 0.056	0.660 ± 0.039	0.569 ± 0.075
BP-delta	L1	RF	0.999 ± 0.002	0.999 ± 0.003	0.605 ± 0.056	0.496 ± 0.057
BP-gamma	L1	XGB	1.000 ± 0.000	1.000 ± 0.000	0.621 ± 0.036	0.526 ± 0.055
BP-theta	mRMR	XGB	0.999 ± 0.002	0.999 ± 0.003	0.660 ± 0.045	0.535 ± 0.028
DetFluc	mRMR	SVM	0.983 ± 0.004	0.979 ± 0.005	0.652 ± 0.030	0.585 ± 0.045
DispEn	L1	SVM	0.979 ± 0.008	0.974 ± 0.010	0.700 ± 0.075	**0.635 ± 0.086**
FuzzyEn	F-Test	SVM	0.979 ± 0.004	0.974 ± 0.006	0.668 ± 0.037	0.612 ± 0.058
HP-comp	mRMR	SVM	0.986 ± 0.006	0.983 ± 0.008	0.699 ± 0.058	**0.635 ± 0.082**
HP-mob	L1	SVM	0.960 ± 0.035	0.949 ± 0.045	0.668 ± 0.062	0.596 ± 0.083
HiguchiFD	L1	SVM	0.979 ± 0.008	0.975 ± 0.010	0.695 ± 0.067	**0.634 ± 0.090**
LZC	F-Test	SVM	0.974 ± 0.012	0.967 ± 0.016	0.691 ± 0.066	0.628 ± 0.096
PermEn	mRMR	SVM	0.988 ± 0.003	0.985 ± 0.003	0.668 ± 0.128	0.618 ± 0.123
SASI	L1	XGB	0.947 ± 0.064	0.930 ± 0.087	0.648 ± 0.054	0.536 ± 0.093
SIE-Del-2-Dim-3	L1	LinSVM	0.942 ± 0.079	0.929 ± 0.097	0.668 ± 0.050	0.571 ± 0.072
SIE-Del-3-Dim-5	L1	LR	0.956 ± 0.058	0.946 ± 0.071	0.672 ± 0.068	0.599 ± 0.059
SIE-Del-5-Dim-7	F-Test	XGB	0.996 ± 0.009	0.995 ± 0.011	0.621 ± 0.048	0.496 ± 0.075
STE-Del-2-Dim-3	mRMR	SVM	0.983 ± 0.008	0.980 ± 0.010	0.692 ± 0.086	**0.646 ± 0.079**
STE-Del-3-Dim-5	L1	LinSVM	0.954 ± 0.068	0.943 ± 0.084	0.691 ± 0.054	0.617 ± 0.093
STE-Del-5-Dim-7	mRMR	RF	0.903 ± 0.061	0.877 ± 0.078	0.683 ± 0.053	0.567 ± 0.120
SlopEn	F-Test	SVM	0.982 ± 0.004	0.978 ± 0.005	0.680 ± 0.113	0.606 ± 0.130
SpecEn	mRMR	SVM	0.981 ± 0.015	0.977 ± 0.018	0.691 ± 0.082	0.633 ± 0.098

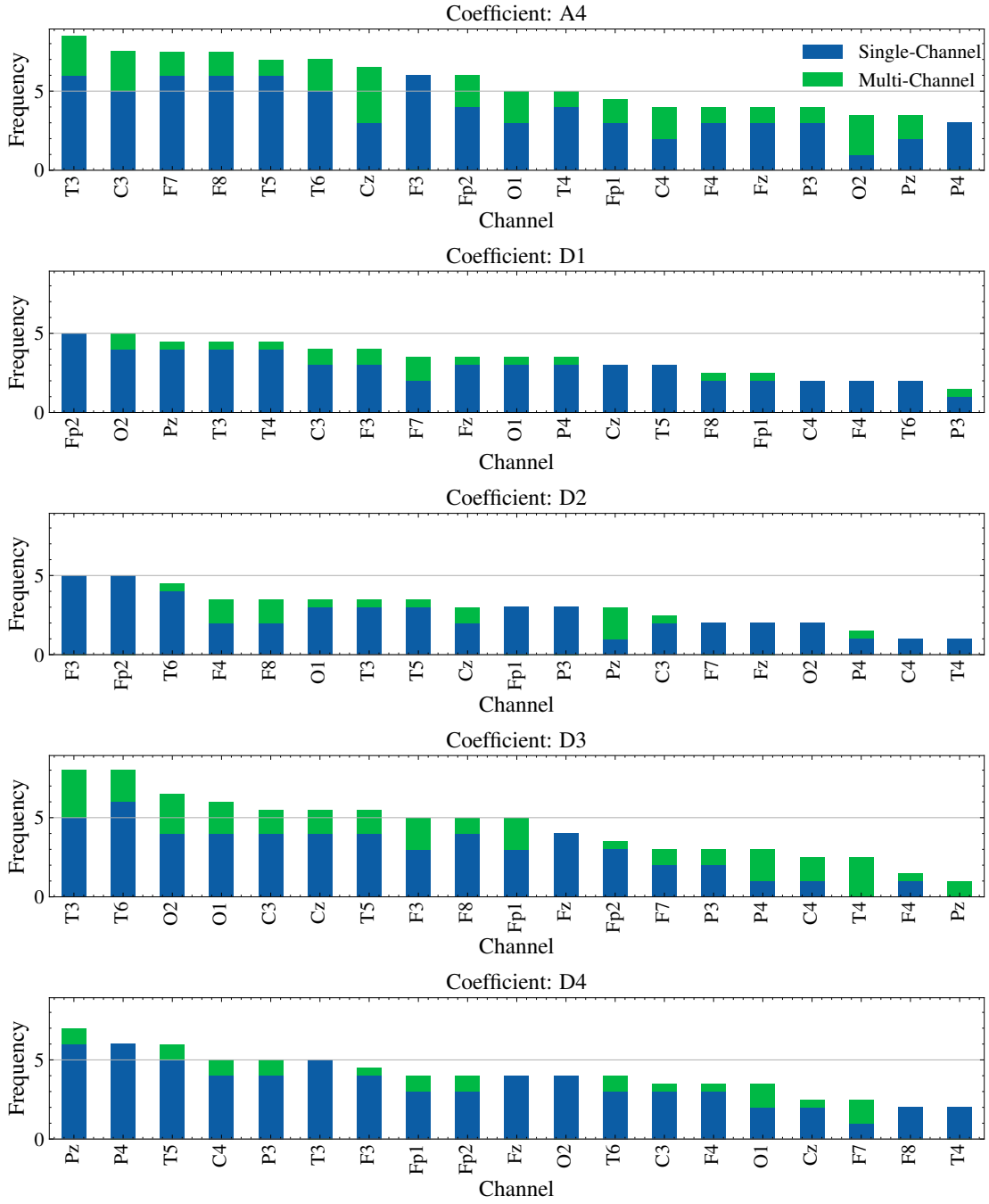


Figure 6.3.: Selection frequency of channels for the top ten performing estimators among all single measure estimators split by coefficients. Multi-Channel channels involve interactions measured by STE or SIE and as such are counted as a half count to avoid double counting.

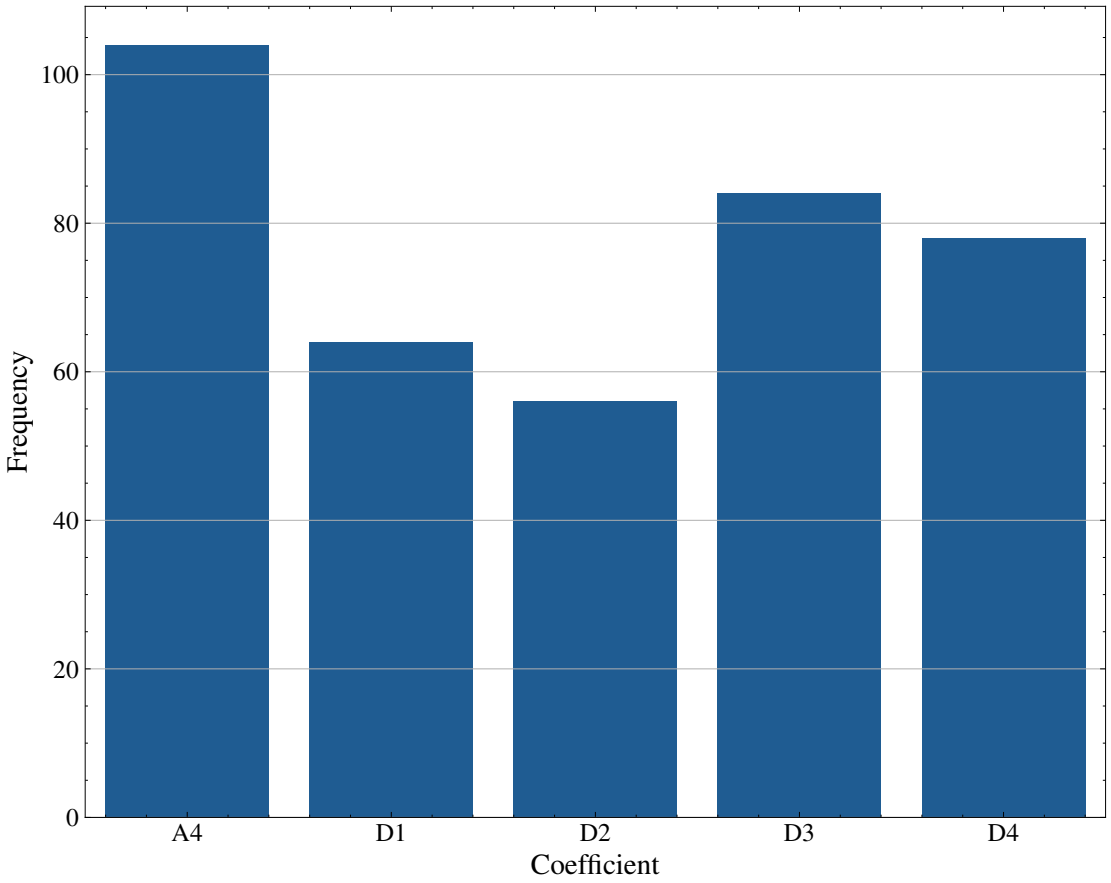


Figure 6.4.: Selection frequency of coefficients for the top ten performing estimators among all single measure estimators.

6.3.4. Multi-Measure

In this section, we present the results of training classification models utilizing all available features and thus incorporating all measures. The objective is to determine whether a combination of multiple features results in improved overall performance compared to using individual measures separately.

For the experiments described in this section, the results were obtained by running the nested CV procedure five times. Although this was not computationally feasible for the single measure approach, as mentioned, nested CV provides robust estimates even when run once. However, when using all measures, where we have a large feature space, running it five times is desirable. This ensures the robustness and reliability of our findings by accounting for the increased complexity and potential overfitting, which provides a more stable and generalizable evaluation of the models' performance.

We again report the accuracy and F1 scores, this time averaged across five runs. Table 6.4 gives the obtained scores. Here, we report the train and test scores for each feature selection and classification method.

Similar to the single measure setting, we observe high training scores with a steep dropoff in test scores. The highest mean testing F1 score achieved in this setting is 0.597, obtained by the RF model with mRMR feature selection. Despite the varied performance across different models and feature selection methods, none of the combinations surpassed the highest F1 scores observed in the single measure setting.

In Figure 6.5, a distinct change in frequency between single- and multi-channels compared to the single measure setting can be observed. Here, the multi-channels are selected the most, with only a few single-channels being selected. We observe that combinations involving C4 and C3 are selected the most, followed by Cz, F8, Fp1 and T5. Figure 6.6 shows that in this setting, the A4 coefficient is by far the most selected one, while D1 is selected the least. There are also a few cases of features being selected that were calculated on the whole signal, which was not divided into wavelet coefficients (see Section 5.2, specifically on bandpower and SASI).

When examining the distribution of selected measures across the five runs for each experiment, Figure 6.7 reveals that SIE with a delay of 3 and an embedding dimension of 5 in combination with F-Test feature selection in the A4 coefficient is the most frequently selected measure by a large margin. In the D3 coefficient a similar, but not as striking trend is observed for STE with a delay of 2 and dimension of 3, but this time in combination with the L1 feature selection. This measure is also consistently among the most selected features across all other coefficients.

Conversely to the single-measure setting, the best-performing models in the combined measure setting, as depicted in Figure A.2, indicate that RF achieves the highest performance for mRMR and F-Test, followed by LR, which is combined with all three feature selection techniques equally often. Linear SVM models demonstrate their best performance when combined with the L1 feature selection technique but is never selected in combination with mRMR. Additionally, NB, which was among the least selected models in the single-measure setting, shows notable improvement and achieves the best results when paired with mRMR feature selection.

Table 6.4.: Mean and standard deviation of train and test scores across five runs, with each run consisting of five CV folds. The best F1 scores for each feature selection method are underscored, while the highest overall F1 score is highlighted in bold.

Feature Selection	**Model**	**Training**		**Testing**	
		Accuracy	**F1**	**Accuracy**	**F1**
L1	LR	0.936 ± 0.091	0.920 ± 0.114	0.621 ± 0.079	0.504 ± 0.104
	LinearSVM	0.962 ± 0.070	0.953 ± 0.087	0.664 ± 0.079	0.562 ± 0.131
	NB	0.799 ± 0.056	0.745 ± 0.071	0.606 ± 0.049	0.493 ± 0.093
	RandomForest	0.950 ± 0.059	0.936 ± 0.076	0.625 ± 0.050	0.484 ± 0.095
	SVM	0.979 ± 0.027	0.972 ± 0.036	0.652 ± 0.056	0.571 ± 0.076
	XGB	0.976 ± 0.055	0.969 ± 0.069	0.574 ± 0.077	0.440 ± 0.107
F-Test	**LR**	0.842 ± 0.050	0.802 ± 0.059	0.684 ± 0.049	**0.589 ± 0.084**
	LinearSVM	0.853 ± 0.034	0.815 ± 0.044	0.590 ± 0.066	0.468 ± 0.066
	NB	0.713 ± 0.032	0.577 ± 0.077	0.629 ± 0.038	0.438 ± 0.092
	RandomForest	0.938 ± 0.069	0.920 ± 0.089	0.586 ± 0.051	0.389 ± 0.087
	SVM	0.982 ± 0.003	0.978 ± 0.003	0.672 ± 0.113	0.567 ± 0.165
	XGB	0.904 ± 0.116	0.861 ± 0.177	0.629 ± 0.080	0.477 ± 0.146
mRMR	LR	0.804 ± 0.027	0.753 ± 0.035	0.625 ± 0.081	0.523 ± 0.106
	LinearSVM	0.820 ± 0.023	0.771 ± 0.035	0.547 ± 0.073	0.443 ± 0.071
	NB	0.770 ± 0.024	0.673 ± 0.050	0.633 ± 0.069	0.437 ± 0.141
	RandomForest	0.966 ± 0.020	0.958 ± 0.025	0.684 ± 0.065	**0.597 ± 0.078**
	SVM	0.983 ± 0.016	0.979 ± 0.020	0.656 ± 0.059	**0.579 ± 0.050**
	XGB	0.961 ± 0.087	0.944 ± 0.124	0.644 ± 0.038	0.512 ± 0.074

Key Findings

- **Overfitting:** The models obtained high training scores but significantly lower test metrics, yet outperformed the benchmarks.
- **Top single measures:** STE (Delay 2, Dimension 3) and DispEn were among the most effective single measures.
- **Linear and nonlinear measures equally effective:** No substantial differences in scores, indicating all measures learned from the data.
- **A4 coefficient dominant:** The A4 coefficient was selected most frequently, especially in dual-channel combinations.
- **Channel selection trends:** Channels T3, T5, T6, Fp2 and F3 were slightly

more frequently chosen in the single-measure setting, while channels C4, C3 were most frequently selected in the multi-measure setting, followed by Cz, F8, Fp1 and T5.

- **Multi-channel features preferred in multi-measure settings:** Multi-channel features were selected more often, with a significant focus on A4 and D3 coefficients.
- **Single vs. multi-measure performance:** Combining multiple measures did not consistently outperform single-measure settings.
- **Top performing classification models:** SVM models performed the best overall; RF and Linear SVM also showed strong results depending on the feature selection method.

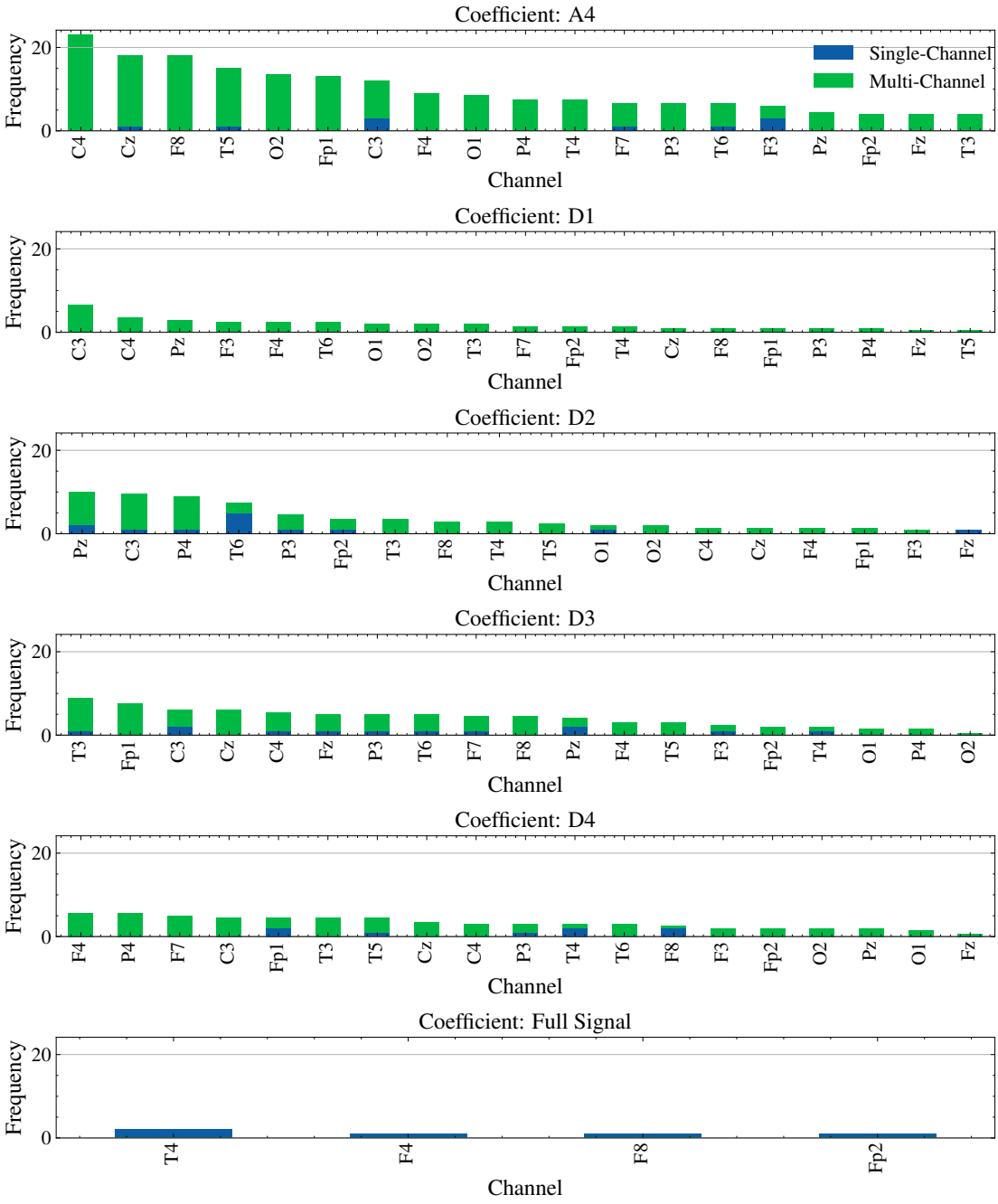


Figure 6.5.: Selection frequency of channels among all estimators in the multi-measure setting split by coefficients. Multi-Channels involve interactions measured by STE or SIE and, as such, are counted as a half count to avoid double counting.

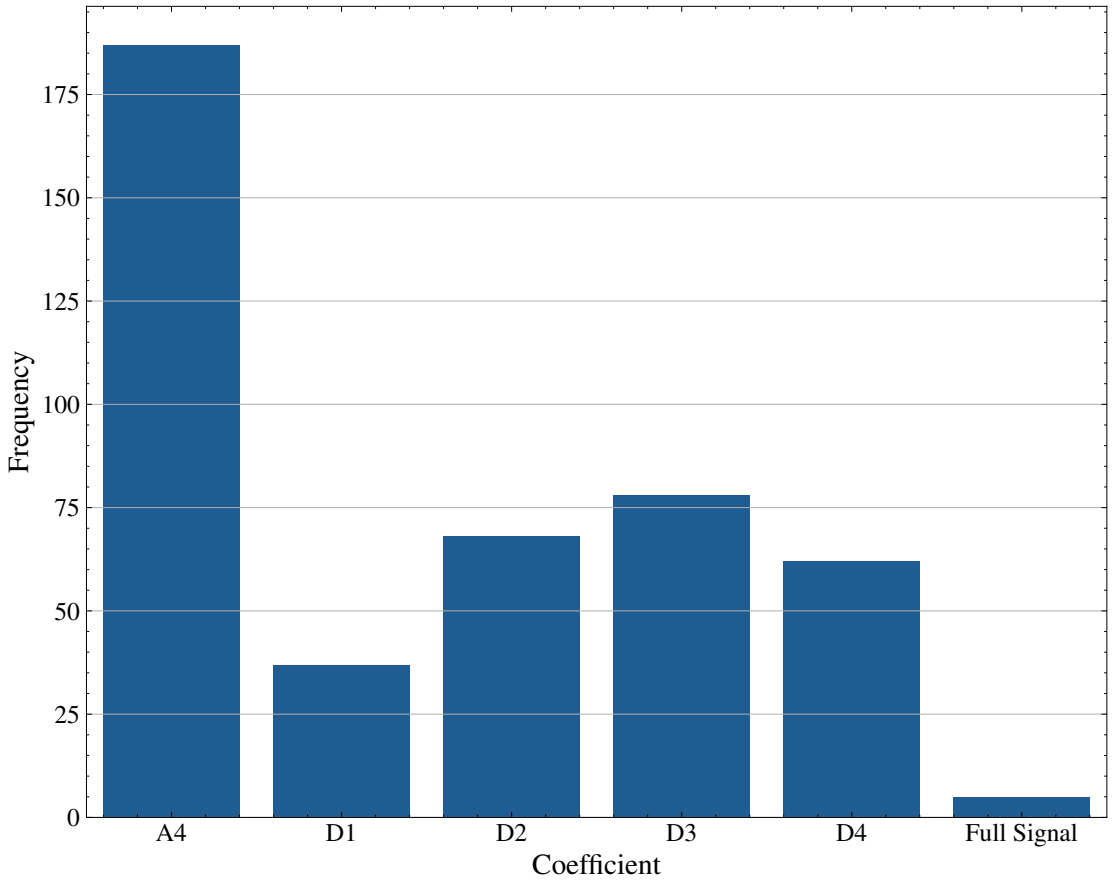


Figure 6.6.: Selection frequency of coefficients across all estimators of the multi-measure setting.

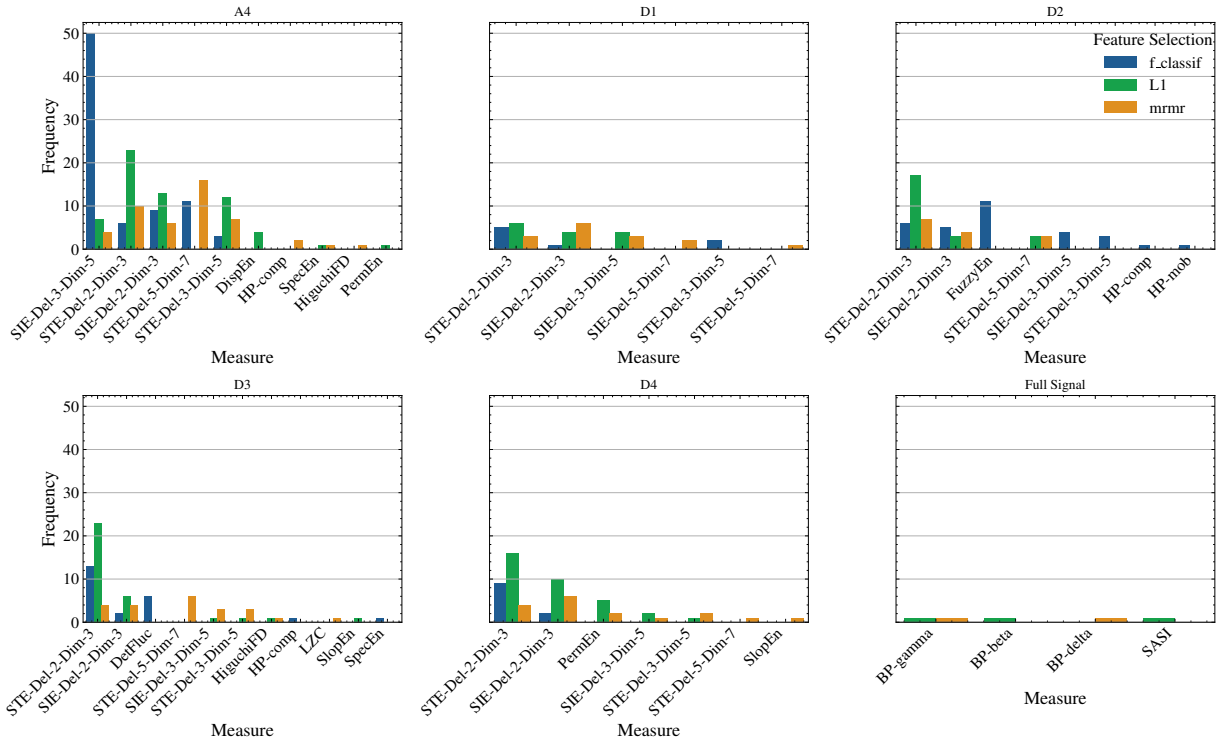


Figure 6.7.: Number of times each measure was selected divided by coefficient and feature selection technique.

7. Discussion

This section discusses the results obtained in this thesis, contextualizing them within the existing body of research and highlighting their clinical relevance. Additionally, we propose potential future directions for research to further refine the prediction of treatment response in patients with MDD using EEG data.

7.1. Comparison with related work

This thesis builds on previous research in the areas of depression detection and treatment response prediction for MDD patients using EEG data. Linear methods, such as power spectrum analysis, have been widely used but have limitations in capturing the complex, dynamic nature of EEG signals. Non-linear methods, on the other hand, offer a more nuanced understanding of brain dynamics [Sta05][MJA17]. The use of entropy-based measures, specifically designed to quantify the complexity and unpredictability of EEG signals, represents a significant advancement in this field.

Unlike previous studies that reported high accuracy ranging from 83.42% to 99.5%, our results did not match these findings. However, these results are not directly comparable to this thesis due to differences in datasets especially considering number of patients. This is highlighted by the work of Beliveau et al., who achieved similar scores to those obtained in this thesis when applying workflows that initially yielded high scores on another dataset. This suggests a strong dependence on the dataset itself [Bel+22].

In comparison to the benchmark models, our approach demonstrates more consistent generalization, suffering less from the severe overfitting observed in the EEGNet and Simple CNN models, which failed to generalize beyond random guessing during testing.

Within the parent project of this thesis, various approaches were evaluated using the same dataset as in this thesis. Hlaváčková-Schindler and Pacher et al. [Hla+23] extracted features by computing Granger-causal graphs and trained classifiers to distinguish between male and female patients, achieving test F1 scores of 0.6 and 0.2 for female and male patients, respectively. Kraljevska [Kra24] applied motif-discovery algorithms and achieved a best F1 score of 0.782 on a separate test set outside the CV. Notably, this thesis balances the data in the same manner as the thesis by Kraljevska to address biological sex imbalances in the dataset. Although our approach outperforms the method using Granger-causal graphs, our results are slightly lower than the F1 score obtained using motif-discovery algorithms. Although we observe test F1 scores exceeding the best score of Kraljevska on some folds, the results are not consistent enough to achieve consistently higher scores. Ultimately, both approaches suffer from overfitting and are thus constrained by similar limitations.

7.2. Interpretation of the results

The analysis demonstrated that multi-channel measures generally exhibit higher absolute mean differences between responders and non-responders than single-channel measures. This observation is supported by the key findings of the preliminary feature separation analysis, which highlighted the higher mean differences observed in multi-channel settings. Symbolic measures, such as STE and SIE, along with DispEn, consistently performed well across all analysis types. Specifically, STE showed high mean absolute differences, lower p-values in statistical analysis and frequent selections in classification experiments, reinforcing its robustness and predictive power. Among the single-channel measures, DispEn was the most effective, followed by HFD, LZC, PermEn and Hjorth Complexity, as observed in the classification results. However, no combination of features appears to be able to consistently capture the entire range of underlying non-linear and complex interactions and patterns in the brain required to perform accurate predictions on unseen patients.

Combining the results of the previous chapters, we can make the following additional observations:

- Incorporating multiple channel signals is preferable over single-channel measures.
- Measures that rely on symbolic sequences or dispersion patterns tend to achieve higher F1 scores compared to those that use exact values. For instance, STE, SIE, PermEn, and DispEn do not operate directly on raw signals but instead utilize these symbolic or dispersion-based methods. This improvement in performance is likely due to their reduced sensitivity to noise, which typically affects measures that rely on raw signals.
- Entropy measures consistently yield scores within 0.04 of the best-observed score, demonstrating strong reliability. While non-entropy-based measures like LZC or HFD also perform well, others like bandpower or SASI can show greater variability, with scores ranging from 0.07 to 0.15 points below the best, indicating inconsistency.

The wavelet coefficient analysis revealed that the A4 coefficient, representative of delta frequencies, was frequently selected in both single- and multi-measure settings, underscoring its significance in predicting treatment response. This was evident in the feature separation analysis, where A4 exhibited larger values in multi-channel settings and in the statistical analysis, where it was particularly significant, especially when combined with specific channels like F8 and measures like STE. Interestingly, A4 did not show large difference values for the single-channel measure feature separation analysis. D1 and D2 coefficients also showed higher differences, indicating their relevance, as noted in the feature separation and statistical analyses. The prominence of the A4 coefficient suggests the importance of delta frequencies, but the findings also highlight the need to consider a range of coefficients to capture the full spectrum of brain activity. These findings align with the varying reports and lack of consensus in recent literature on the best-performing frequency range.

When analyzing the EEG channels, no single channel significantly dominates the number of selections in the feature selection analysis. Generally though, patterns in the mean absolute differences and selected channels in the classification problem tend to coincide, with similar top-performing channels identified in both single- and multi-measure settings. Combined with the results of the statistical analysis, channels T3, T5, T6, C4, Cz, F8 and Pz emerge as the most promising for our classification problem, with T6 showing large mean absolute difference values. The identified channels indicate that treatment response in depressed patients involves a complex interplay of sensory, motor, cognitive and executive functions, with a significant focus on the temporal lobe's role in both hemispheres of the cerebrum in processing auditory information, memory and language.

Model-wise SVM in the single-measure setting and RF in the multi-channel setting appear to be the preferred classifiers. In the single-measure setting, the SVM is the clear favorite while in the multi-measure setting, this preference is not as clear and seems to depend on the feature selection technique and therefore also the selected features. Notably, NB performs poorly on the training sets compared to the other models with the effect not being transferred to the testing set because the lack of ability to generalize affects all models similarly. This further highlights that overfitting is the main limitation of this approach, as even models that perform poorly on the training set do not show worse precision on the testing set compared to other models.

No single feature selection technique consistently outperforms the other methods across all settings. However, mRMR and L1-based selection demonstrate reliable performance across various measures in the single-measure setting, which is consistent with their strengths in handling complex, non-linear data and producing sparse solutions. Interestingly, feature selection using the F-test in the multi-measure setting frequently selects SIE, which, according to our statistical analysis, is inferior to STE. This might indicate a limitation of the F-test in capturing multivariate interactions.

Furthermore, the overall effectiveness of each model is hindered by the limited dataset size. While feature selection techniques identified features that performed well on the training set, this success did not translate to the testing set, emphasizing the need for larger, more diverse datasets to enhance generalizability and robustness.

7.3. Limitations and Future Work

Although the dataset used in this study is relatively large within the context of medical research, it is still considered small by machine learning standards, where datasets often contain thousands or even millions of samples necessary for achieving higher accuracy [Vab+19]. This smaller size could limit the generalizability of the findings, a limitation that seems to be at the core of all previous studies on this particular dataset.

In addition, feature values were extracted from the entire signal for each patient and averaged. Although necessary due to the large number of features, this averaging may have obscured important patterns in the data. With a larger dataset, it would be possible to split the signals into more granular segments to capture these nuances, as larger sample sizes allow for a larger feature space to be learned.

Furthermore, the parameters for feature calculation were not chosen optimally. Although many measures share parameters, such as delay or dimension, determining the best parameters for each feature and signal is computationally intensive. In this thesis, optimal parameters were calculated for only a few sample signals and measures and then averaged. These averages may not generalize well across different signals and measures. Future studies could improve parameter optimization with more computational resources and refined methodologies.

Another limitation is the variability in treatments received by patients. In this study, patients received different treatments, sometimes in combination, which was not considered in the classification workflow due to the small dataset size. Future work should focus on a more consistent treatment regimen for all patients, as this could make the results more robust and the effects on EEG signals more consistent across responders.

Additionally, future research should incorporate separate datasets to confirm the validity of the findings. As mentioned earlier, Beliveau et al. achieved comparable scores to those obtained in this thesis by applying workflows that initially performed well on one dataset with 79 patients to a new, separate dataset. This may indicate a strong dependence on the dataset itself [Bel+22]. Consequently, it highlights the importance of validating the results across multiple data sources to ensure that the results are reliable and generalizable.

Key Findings

- **Overfitting as a Central Issue:** Overfitting was identified as the main limitation across various models and analyses, emphasizing the need for larger datasets.
- **Best Performing Measure:** STE outperformed all other measures in terms of the absolute value difference between responders and non-responders, the p-value in statistical analysis, the F1 score in the single-measure setting and the frequency of selection in the multi-measure setting.
- **Multi-channel vs. Single-channel:** Multi-channel measures generally demonstrate better F1 scores and p-values than single-channel measures.
- **Entropy and Symbolic Sequence Measures:** Entropy measures, especially those utilizing symbolic sequences or dispersion patterns rather than raw signals, demonstrated higher F1 scores.
- **Non-linear vs. Linear Measures:** Non-linear measures did not consistently outperform linear measures, indicating the need for careful selection based on the specific context.
- **Multiple Measures vs Single Measures:** Combining multiple measures did not consistently yield better results than using single measures.
- **Best Performing Coefficient:** The A4 wavelet coefficient, representing delta frequencies, was the most significant predictor of treatment response in

the statistical analysis and also showed increased selection in the classification analysis.

- **Best Performing Channels:** The channels T3, T5, T6, C4, Cz, F8 and Pz emerged as the most selected for classification, while also ranking high in both the absolute difference and statistical analyses, highlighting the impact of depression treatment on various brain functions.
- **Future Research Directions:** Future studies should use larger and more diverse datasets to improve generalizability and robustness while addressing overfitting and refining feature parameters.

8. Conclusion

This thesis focuses on the potential of entropy-based methods for predicting treatment outcomes in patients with MDD using EEG data. The primary objective of this thesis was to evaluate the effectiveness of these non-linear techniques in extracting meaningful features from EEG signals and their application in machine learning models for treatment prediction.

The findings of this study have practical implications for the development of predictive models in clinical settings. By identifying key EEG channels, frequency ranges and measures that correlate with treatment response, our approach can guide the design of more effective and personalized treatment plans for patients with MDD.

Contributions To the best of our knowledge, this thesis is the first to consider a larger heterogeneous set of entropy-based measures and evaluate their effectiveness in the context of treatment prediction for patients with MDD. This set includes both single- and multi-channel measures, while also introducing a new type of multi-channel measure, SIE. Additionally, this work evaluates multiple feature-selection algorithms, namely mRMR, F-test and L1 feature selection, a comparison that has been lacking in MDD treatment prediction analysis until now. The use of a new large dataset in the context of EEG depression datasets contributes significantly to the sparse body of research examining how classification methods perform when applied to a larger dataset of individuals with depression.

Results Before extracting the features from the EEG signals by calculating the measure values, we applied a wavelet decomposition that roughly corresponded to the common frequency bands: alpha, beta, gamma, delta and theta. We then use nested CV, which offers a more accurate estimation of a model's generalization while still allowing for model selection and hyperparameter tuning.

Although the results of this thesis outperform the benchmarks of random guessing and EEGNet, a deep-learning approach specifically designed to model EEG data, the approach still suffers from overfitting and does not meet some of the reported scores in the literature. However, our results are comparable to those of other studies using the same dataset, achieving a top accuracy of approximately 0.7 and an F1 score of 0.646. Therefore, the results may be heavily dependent on the dataset and overfitting seems to remain a central issue when working with medical datasets.

In this thesis, several channels were identified as important, particularly T3, T5, T6, C4, Cz, F8 and Pz. For the coefficients, the A4 coefficient, which corresponds to delta frequencies, was selected the most. Regarding the measures, we observed that multi-

channel measures, such as STE and the newly introduced SIE, consistently scored best across all types of analysis, ranging from statistical assessments to classification tasks, with the former generally performing slightly better than the latter. Both linear and non-linear methods produced high-performing features, with entropy measures generally performing near the best-observed score. For single measures, DispEn, HFD, LZC and PermEn provided the best results, demonstrating that both linear and non-linear measures can provide insights into brain dynamics. Including multiple measures did not improve classification precision compared to employing features on their own. The SVM classifiers consistently performed well, particularly in the single-measure setting, while RF classifiers were most effective in the multi-measure setting. However, the preference for RF in multi-measure settings may vary depending on the feature selection technique used.

Limitations Despite the relatively large dataset used in this study (176 patients), it remains small by machine learning standards, potentially limiting the generalizability of the findings. In addition, feature values were extracted from the entire signal for each patient and averaged. Although necessary due to the large number of features, this averaging may have obscured important patterns in the data. In addition, the parameters for feature calculation were not optimally selected due to computational constraints, which led to a potential suboptimal performance.

Future work Future research should focus on using larger datasets to enhance the generalizability and robustness of the models. This includes splitting the signals into more granular segments to capture the nuances and improving parameter optimization with more computational resources. In addition, maintaining a consistent treatment regimen for all patients would make the results more robust and the effects on EEG signals more consistent across responders. Finally, validation across multiple data sources is crucial to ensure the reliability and generalizability of the findings.

In conclusion, this thesis makes significant contributions to the field of EEG-based treatment prediction for MDD by providing a structured, comprehensive framework that utilizes a diverse set of entropy-based measures and applies it to a larger dataset, both of which have been lacking in previous MDD treatment response research. While challenges remain, the insights gained from this study provide a valuable foundation for future research and clinical applications.

Bibliography

[Nyq24] H. Nyquist. 'Certain Factors Affecting Telegraph Speed'. In: *Trans. Am. Inst. Electr. Eng.* XLIII (Jan. 1924), pp. 412–422. ISSN: 0096-3860, 2330-9431. DOI: 10.1109/T-AIEE.1924.5060996. URL: https://ieeexplore.ieee.org/document/5060996/ (visited on 19/01/2024).

[Har28] R. V. L. Hartley. 'Transmission of Information'. In: *Bell System Technical Journal* 7.3 (July 1928), pp. 535–563. ISSN: 00058580. DOI: 10.1002/j.1538-7305.1928.tb01236.x. URL: https://ieeexplore.ieee.org/document/6769394 (visited on 19/01/2024).

[Hot31] Harold Hotelling. 'The Generalization of Student's Ratio'. In: *Ann. Math. Statist.* 2.3 (Aug. 1931), pp. 360–378. ISSN: 0003-4851. DOI: 10.1214/aoms/1177732979. URL: http://projecteuclid.org/euclid.aoms/1177732979 (visited on 26/08/2024).

[Sha48] C. E. Shannon. 'A Mathematical Theory of Communication'. In: *The Bell System Technical Journal* 27.3 (July 1948), pp. 379–423. ISSN: 0005-8580. DOI: 10.1002/j.1538-7305.1948.tb01338.x. URL: https://ieeexplore.ieee.org/abstract/document/6773024 (visited on 30/12/2023).

[Hjo70] Bo Hjorth. 'EEG Analysis Based on Time Domain Properties'. In: *Electroencephalography and Clinical Neurophysiology* 29.3 (Sept. 1970), pp. 306–310. ISSN: 00134694. DOI: 10.1016/0013-4694(70)90143-4. URL: https://linkinghub.elsevier.com/retrieve/pii/0013469470901434 (visited on 25/03/2024).

[LZ76] A. Lempel and J. Ziv. 'On the Complexity of Finite Sequences'. In: *IEEE Trans. Inform. Theory* 22.1 (Jan. 1976), pp. 75–81. ISSN: 0018-9448. DOI: 10.1109/TIT.1976.1055501. URL: http://ieeexplore.ieee.org/document/1055501/ (visited on 25/03/2024).

[FMI83] Kunihiko Fukushima, Sei Miyake and Takayuki Ito. 'Neocognitron: A Neural Network Model for a Mechanism of Visual Pattern Recognition'. In: *IEEE Trans. Syst., Man, Cybern.* SMC-13.5 (Sept. 1983), pp. 826–834. ISSN: 0018-9472, 2168-2909. DOI: 10.1109/TSMC.1983.6313076. URL: https://ieeexplore.ieee.org/document/6313076/ (visited on 23/08/2024).

[Hig88] T. Higuchi. 'Approach to an Irregular Time Series on the Basis of the Fractal Theory'. In: *Physica D: Nonlinear Phenomena* 31.2 (June 1988), pp. 277–283. ISSN: 01672789. DOI: 10.1016/0167-2789(88)90081-4. URL: https://linkinghub.elsevier.com/retrieve/pii/0167278988900814 (visited on 25/03/2024).

[Moc89] Jonas Mockus. *Bayesian Approach to Global Optimization: Theory and Applications. Buch.* Mathematics and Its Applications Soviet Series 37. Dordrecht: Kluwer Acad. Publ, 1989. 254 pp. ISBN: 978-0-7923-0115-8.

[Pin91] S M Pincus. 'Approximate Entropy as a Measure of System Complexity.' In: *Proc. Natl. Acad. Sci. U.S.A.* 88.6 (15th Mar. 1991), pp. 2297–2301. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.88.6.2297. URL: https://pnas.org/doi/full/10.1073/pnas.88.6.2297 (visited on 03/03/2024).

[KK92] J. N. Kapur and H. K. Kesavan. 'Entropy Optimization Principles and Their Applications'. In: *Entropy and Energy Dissipation in Water Resources.* Ed. by V. P. Singh and M. Fiorentino. Red. by V. P. Singh. Vol. 9. Dordrecht: Springer Netherlands, 1992, pp. 3–20. ISBN: 978-94-010-5072-2. DOI: 10.1007/978-94-011-2430-0_1. URL: http://link.springer.com/10.1007/978-94-011-2430-0_1 (visited on 03/03/2024).

[DT93] A. De Luca and S. Termini. 'A Definition of a Nonprobabilistic Entropy in the Setting of Fuzzy Sets Theory'. In: *Readings in Fuzzy Sets for Intelligent Systems.* Elsevier, 1993, pp. 197–202. ISBN: 978-1-4832-1450-4. DOI: 10.1016/B978-1-4832-1450-4.50020-1. URL: https://linkinghub.elsevier.com/retrieve/pii/B9781483214504500201 (visited on 10/05/2024).

[Tib96] Robert Tibshirani. 'Regression Shrinkage and Selection Via the Lasso'. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1 (1st Jan. 1996), pp. 267–288. ISSN: 1369-7412, 1467-9868. DOI: 10.1111/j.2517-6161.1996.tb02080.x. URL: https://academic.oup.com/jrsssb/article/58/1/267/7027929 (visited on 04/08/2024).

[Got98] Ian Gotlib. 'EEG Alpha Asymmetry, Depression, and Cognitive Functioning'. In: *Cognition & Emotion - COGNITION EMOTION* 12 (1st May 1998), pp. 449–478. DOI: 10.1080/026999398379673.

[LeC+99] Yann LeCun, Patrick Haffner, Léon Bottou and Yoshua Bengio. 'Object Recognition with Gradient-Based Learning'. In: *Shape, Contour and Grouping in Computer Vision.* Ed. by David A. Forsyth, Joseph L. Mundy, Vito di Gesú and Roberto Cipolla. Berlin, Heidelberg: Springer, 1999, pp. 319–345. ISBN: 978-3-540-46805-9. DOI: 10.1007/3-540-46805-6_19. URL: https://doi.org/10.1007/3-540-46805-6_19 (visited on 23/08/2024).

[Kan+00] Eric R Kandel, James H Schwartz, Thomas M Jessell, Steven Siegelbaum, A James Hudspeth and Sarah Mack. *Principles of Neural Science.* Vol. 4. McGraw-hill New York, 2000.

[Sch00] Thomas Schreiber. 'Measuring Information Transfer'. In: *Phys. Rev. Lett.* 85.2 (10th July 2000), pp. 461–464. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.85.461. arXiv: nlin/0001042. URL: http://arxiv.org/abs/nlin/0001042 (visited on 05/01/2024).

[And+01] Ralph G. Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David and Christian E. Elger. 'Indications of Nonlinear Deterministic and Finite-Dimensional Structures in Time Series of Brain Electrical Activity: Dependence on Recording Region and Brain State'. In: *Phys. Rev. E* 64.6 (20th Nov. 2001), p. 061907. ISSN: 1063-651X, 1095-3787. DOI: 10.1103/PhysRevE.64.061907. URL: https://link.aps.org/doi/10.1103/PhysRevE.64.061907 (visited on 04/07/2024).

[AH01] Roseanne Armitage and Robert F Hoffmann. 'Sleep EEG, Depression and Gender'. In: *Sleep Medicine Reviews* 5.3 (June 2001), pp. 237–246. ISSN: 10870792. DOI: 10.1053/smrv.2000.0144. URL: https://linkinghub.elsevier.com/retrieve/pii/S1087079200901448 (visited on 06/07/2023).

[Kno+01] Verner Knott, Colleen Mahoney, Sidney Kennedy and Kenneth Evans. 'EEG Power, Frequency, Asymmetry and Coherence in Male Depression'. In: *Psychiatry Research: Neuroimaging* 106.2 (Apr. 2001), pp. 123–140. ISSN: 09254927. DOI: 10.1016/S0925-4927(00)00080-9. URL: https://linkinghub.elsevier.com/retrieve/pii/S0925492700000809 (visited on 06/07/2023).

[BP02] Christoph Bandt and Bernd Pompe. 'Permutation Entropy: A Natural Complexity Measure for Time Series'. In: *Phys. Rev. Lett.* 88.17 (11th Apr. 2002), p. 174102. ISSN: 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.88.174102. URL: https://link.aps.org/doi/10.1103/PhysRevLett.88.174102 (visited on 03/03/2024).

[Gol+02] Robert N. Golden, Charles B. Nemeroff, Paul McSorley, Cornelius D. Pitts and Eric M. Dube. 'Efficacy and Tolerability of Controlled-Release and Immediate-Release Paroxetine in the Treatment of Depression'. In: *J. Clin. Psychiatry* 63.7 (15th July 2002), pp. 577–584. ISSN: 0160-6689. DOI: 10.4088/JCP.v63n0707. URL: http://article.psychiatrist.com/?ContentType=START&ID=10001526 (visited on 26/09/2023).

[DP03] C. Ding and H. Peng. 'Minimum Redundancy Feature Selection from Microarray Gene Expression Data'. In: *Computational Systems Bioinformatics. CSB2003. Proceedings of the 2003 IEEE Bioinformatics Conference. CSB2003.* 2003 IEEE Bioinformatics Conference. CSB2003. Stanford, CA, USA: IEEE Comput. Soc, 2003, pp. 523–528. ISBN: 978-0-7695-2000-1. DOI: 10.1109/CSB.2003.1227396. URL: http://ieeexplore.ieee.org/document/1227396/ (visited on 05/04/2024).

[Mac03] David J. C. MacKay. *Information Theory, Inference and Learning Algorithms.* Cambridge University Press, 25th Sept. 2003. 694 pp. ISBN: 978-0-521-64298-9. Google Books: AKuMj4PN_EMC.

[All+04] John J. B. Allen, Heather L. Urry, Sabrina K. Hitt and James A. Coan. 'The Stability of Resting Frontal Electroencephalographic Asymmetry in Depression'. In: *Psychophysiology* 41.2 (Mar. 2004), pp. 269–280. ISSN: 0048-

5772, 1469-8986. DOI: 10.1111/j.1469-8986.2003.00149.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1469-8986.2003.00149.x (visited on 06/07/2023).

[MO04] Roger S McIntyre and Claire O'Donovan. 'The Human Cost of Not Achieving Full Remission in Depression'. In: *Can J Psychiatry* 49 (2004).

[RLM04] Joshua S. Richman, Douglas E. Lake and J.Randall Moorman. 'Sample Entropy'. In: *Methods in Enzymology*. Vol. 384. Elsevier, 2004, pp. 172–184. ISBN: 978-0-12-182789-2. DOI: 10.1016/S0076-6879(04)84011-4. URL: https://linkinghub.elsevier.com/retrieve/pii/S0076687904840114 (visited on 03/03/2024).

[NS05] Ernst Niedermeyer and F. H. Lopes da Silva. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. Lippincott Williams & Wilkins, 2005. 1342 pp. ISBN: 978-0-7817-5126-1. Google Books: tndqYGPHQdEC.

[Pet+05] Timothy Petersen, George I. Papakostas, Michael A. Posternak, Alexis Kant, Wendy M. Guyker, Dan V. Iosifescu, Albert S. Yeung, Andrew A. Nierenberg and Maurizio Fava. 'Empirical Testing of Two Models for Staging Antidepressant Treatment Resistance'. In: *Journal of Clinical Psychopharmacology* 25.4 (Aug. 2005), p. 336. ISSN: 0271-0749. DOI: 10.1097/01.jcp.0000169036.40755.16. URL: https://journals.lww.com/psychopharmacology/abstract/2005/08000/empirical_testing_of_two_models_for_staging.8.aspx (visited on 26/09/2023).

[Sta05] C.J. Stam. 'Nonlinear Dynamical Analysis of EEG and MEG: Review of an Emerging Field'. In: *Clinical Neurophysiology* 116.10 (Oct. 2005), pp. 2266–2301. ISSN: 13882457. DOI: 10.1016/j.clinph.2005.06.011. URL: https://linkinghub.elsevier.com/retrieve/pii/S1388245705002403 (visited on 12/07/2023).

[Ole06] Piotr Olejniczak. 'Neurophysiologic Basis of EEG:' in: *Journal of Clinical Neurophysiology* 23.3 (June 2006), pp. 186–189. ISSN: 0736-0258. DOI: 10.1097/01.wnp.0000220079.61973.6c. URL: http://journals.lww.com/00004691-200606000-00002 (visited on 29/11/2023).

[JKL07] Ole Jensen, Jochen Kaiser and Jean-Philippe Lachaux. 'Human Gamma-Frequency Oscillations Associated with Attention and Memory'. In: *Trends in Neurosciences* 30.7 (1st July 2007), pp. 317–324. ISSN: 0166-2236, 1878-108X. DOI: 10.1016/j.tins.2007.05.001. pmid: 17499860. URL: https://www.cell.com/trends/neurosciences/abstract/S0166-2236(07)00105-1 (visited on 18/03/2024).

[GAD08] S. Ghosh-Dastidar, H. Adeli and N. Dadmehr. 'Principal Component Analysis-Enhanced Cosine Radial Basis Function Neural Network for Robust Epilepsy and Seizure Detection'. In: *IEEE Trans. Biomed. Eng.* 55.2 (Feb. 2008), pp. 512–518. ISSN: 0018-9294, 1558-2531. DOI: 10.1109/TBME.2007.905490.

URL: http://ieeexplore.ieee.org/document/4360124/ (visited on 06/07/2023).

[Li+08] Yingjie Li, Shanbao Tong, Dan Liu, Yi Gai, Xiuyuan Wang, Jijun Wang, Yihong Qiu and Yisheng Zhu. 'Abnormal EEG Complexity in Patients with Schizophrenia and Depression'. In: *Clinical Neurophysiology* 119.6 (June 2008), pp. 1232–1241. ISSN: 13882457. DOI: 10.1016/j.clinph.2008.01.104. URL: https://linkinghub.elsevier.com/retrieve/pii/S1388245708000461 (visited on 25/03/2024).

[SL08] Matthäus Staniek and Klaus Lehnertz. 'Symbolic Transfer Entropy'. In: *Physical review letters* 100 (18th Apr. 2008), p. 158101. DOI: 10.1103/PhysRevLett.100.158101.

[Hin+09] Hiie Hinrikus, Anna Suhhova, Maie Bachmann, Kaire Aadamsoo, Ülle võhma, Jaanus Lass and Viiu Tuulik. 'Electroencephalographic Spectral Asymmetry Index for Detection of Depression'. In: *Medical & biological engineering & computing* 47 (1st Nov. 2009), pp. 1291–9. DOI: 10.1007/s11517-009-0554-9.

[PCL09] Y N Pan, J Chen and X L Li. 'Spectral Entropy: A Complementary Index for Rolling Element Bearing Performance Degradation Assessment'. In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 223.5 (1st May 2009), pp. 1223–1231. ISSN: 0954-4062, 2041-2983. DOI: 10.1243/09544062JMES1224. URL: http://journals.sagepub.com/doi/10.1243/09544062JMES1224 (visited on 10/05/2024).

[Vid+09] Carmen Vidaurre, Nicole Krämer, Benjamin Blankertz and Alois Schlögl. 'Time Domain Parameters as a Feature for EEG-based Brain–Computer Interfaces'. In: *Neural Networks* 22.9 (Nov. 2009), pp. 1313–1319. ISSN: 08936080. DOI: 10.1016/j.neunet.2009.07.020. URL: https://linkinghub.elsevier.com/retrieve/pii/S0893608009001786 (visited on 25/03/2024).

[Ste+10] Jennifer L. Stewart, Andrew W. Bismark, David N. Towers, James A. Coan and John J. B. Allen. 'Resting Frontal EEG Asymmetry as an Endophenotype for Depression Risk: Sex-specific Patterns of Frontal Brain Asymmetry.' In: *Journal of Abnormal Psychology* 119.3 (Aug. 2010), pp. 502–512. ISSN: 1939-1846, 0021-843X. DOI: 10.1037/a0019196. URL: http://doi.apa.org/getdoi.cfm?doi=10.1037/a0019196 (visited on 06/07/2023).

[Ped+11] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg et al. 'Scikit-Learn: Machine Learning in Python'. In: *the Journal of machine Learning research* 12 (2011), pp. 2825–2830.

[Tel+11] Qawi K. Telesford, Sean L. Simpson, Jonathan H. Burdette, Satoru Hayasaka and Paul J. Laurienti. 'The Brain as a Complex System: Using Network Science as a Tool for Understanding the Brain'. In: *Brain Connect* 1.4 (2011), pp. 295–308. ISSN: 2158-0022. DOI: 10.1089/brain.2011.0055. pmid: 22432419.

[Arn+12] Martijn Arns, Wilhelmus H. Drinkenburg, Paul B. Fitzgerald and J. Leon Kenemans. 'Neurophysiological Predictors of Non-Response to rTMS in Depression'. In: *Brain Stimulation* 5.4 (Oct. 2012), pp. 569–576. ISSN: 1935861X. DOI: 10.1016/j.brs.2011.12.003. URL: https://linkinghub.elsevier.com/retrieve/pii/S1935861X11001720 (visited on 18/01/2024).

[Ash12] Robert B Ash. *Information Theory*. Courier Corporation, 2012. ISBN: 0-486-14145-4.

[Gra12] Peter Grassberger. *Randomness, Information, and Complexity*. 15th Aug. 2012. arXiv: 1208.3459 [physics]. URL: http://arxiv.org/abs/1208.3459 (visited on 19/01/2024). Pre-published.

[KB12] J. Satheesh Kumar and P. Bhuvaneswari. 'Analysis of Electroencephalography (EEG) Signals and Its Categorization–A Study'. In: *Procedia Engineering* 38 (2012), pp. 2525–2536. ISSN: 18777058. DOI: 10.1016/j.proeng.2012.06.298. URL: https://linkinghub.elsevier.com/retrieve/pii/S1877705812022114 (visited on 29/11/2023).

[Mén+12] María Andreina Méndez, Pilar Zuluaga, Roberto Hornero, Carlos Gómez, Javier Escudero, Alfonso Rodríguez-Palancas, Tomás Ortiz and Alberto Fernández. 'Complexity Analysis of Spontaneous Brain Activity: Effects of Depression and Antidepressant Treatment'. In: *J Psychopharmacol* 26.5 (May 2012), pp. 636–643. ISSN: 0269-8811, 1461-7285. DOI: 10.1177/0269881111408966. URL: http://journals.sagepub.com/doi/10.1177/0269881111408966 (visited on 25/03/2024).

[NG12] Nicoletta Nicolaou and Julius Georgiou. 'Detection of Epileptic Electroencephalogram Based on Permutation Entropy and Support Vector Machines'. In: *Expert Systems with Applications* 39.1 (Jan. 2012), pp. 202–209. ISSN: 09574174. DOI: 10.1016/j.eswa.2011.07.008. URL: https://linkinghub.elsevier.com/retrieve/pii/S0957417411009705 (visited on 27/04/2023).

[Pan+12] Mayur Pandya, Murat Altinay, Donald A. Malone and Amit Anand. 'Where in the Brain Is Depression?' In: *Curr Psychiatry Rep* 14.6 (Dec. 2012), pp. 634–642. ISSN: 1523-3812. DOI: 10.1007/s11920-012-0322-7. pmid: 23055003. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3619732/ (visited on 03/03/2024).

[Bac+13] Maie Bachmann, Jaanus Lass, Anna Suhhova and Hiie Hinrikus. 'Spectral Asymmetry and Higuchi's Fractal Dimension Measures of Depression Electroencephalogram'. In: *Computational and Mathematical Methods in Medicine* 2013 (2013), pp. 1–8. ISSN: 1748-670X, 1748-6718. DOI: 10.1155/2013/251638. URL: http://www.hindawi.com/journals/cmmm/2013/251638/ (visited on 20/07/2023).

[HMR13] Behshad Hosseinifard, Mohammad Hassan Moradi and Reza Rostami. 'Classifying Depression Patients and Normal Subjects Using Machine Learning Techniques and Nonlinear Features from EEG Signal'. In: *Computer Methods and Programs in Biomedicine* 109.3 (1st Mar. 2013), pp. 339–345. ISSN: 0169-2607. DOI: 10.1016/j.cmpb.2012.10.008. URL: https://www.sciencedirect.com/science/article/pii/S0169260712002507 (visited on 27/04/2023).

[Kas13] Nikola Kasabov. *Springer Handbook of Bio-/Neuroinformatics*. 1st edition. New York: Springer, 2013. ISBN: 978-3-642-30573-3.

[SC13] Saeid Sanei and Jonathon A. Chambers. *EEG Signal Processing*. John Wiley & Sons, 28th May 2013. 308 pp. ISBN: 978-1-118-69123-6. Google Books: f44hLefOz6UC.

[YT13] Albert C. Yang and Shih-Jen Tsai. 'Is Mental Illness Complex? From Behavior to Brain'. In: *Progress in Neuro-Psychopharmacology and Biological Psychiatry* 45 (Aug. 2013), pp. 253–257. ISSN: 02785846. DOI: 10.1016/j.pnpbp.2012.09.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S0278584612002539 (visited on 19/01/2024).

[Arn+14] Martijn Arns, Alexander Cerquera, Rafael M. Gutiérrez, Fred Hasselman and Jan A. Freund. 'Non-Linear EEG Analyses Predict Non-Response to rTMS Treatment in Major Depressive Disorder'. In: *Clinical Neurophysiology* 125.7 (1st July 2014), pp. 1392–1399. ISSN: 1388-2457. DOI: 10.1016/j.clinph.2013.11.022. URL: https://www.sciencedirect.com/science/article/pii/S1388245713012194 (visited on 18/01/2024).

[Fau+14] Oliver Faust, Peng Chuan Alvin Ang, Subha D. Puthankattil and Paul K. Joseph. 'Depression Diagnosis Support System Based on Eeg Signal Entropies'. In: *J. Mech. Med. Biol.* 14.03 (June 2014), p. 1450035. ISSN: 0219-5194. DOI: 10.1142/S0219519414500353. URL: https://www.worldscientific.com/doi/abs/10.1142/S0219519414500353 (visited on 25/07/2023).

[JPB14] Robert Jenke, Angelika Peer and Martin Buss. 'Feature Extraction and Selection for Emotion Recognition from EEG'. In: *IEEE Trans. Affective Comput.* 5.3 (1st July 2014), pp. 327–339. ISSN: 1949-3045. DOI: 10.1109/TAFFC.2014.2339834. URL: http://ieeexplore.ieee.org/document/6858031/ (visited on 25/03/2024).

[Kjæ+14] Marie Kjærgaard, Catharina Elisabeth Arfwedson Wang, Knut Waterloo and Rolf Jorde. 'A Study of the Psychometric Properties of the Beck Depression Inventory- II , the Montgomery and Åsberg Depression Rating Scale, and the Hospital Anxiety and Depression Scale in a Sample from a Healthy Population'. In: *Scandinavian J Psychology* 55.1 (Feb. 2014), pp. 83–89. ISSN: 0036-5564, 1467-9450. DOI: 10.1111/sjop.12090. URL: https://onlinelibrary.wiley.com/doi/10.1111/sjop.12090 (visited on 13/02/2024).

[Lef+14] Jean-Pascal Lefaucheur, Nathalie André-Obadia, Andrea Antal, Samar S. Ayache, Chris Baeken, David H. Benninger, Roberto M. Cantello, Massimo Cincotta, Mamede De Carvalho, Dirk De Ridder, Hervé Devanne, Vincenzo Di Lazzaro, Saša R. Filipović, Friedhelm C. Hummel, Satu K. Jääskeläinen, Vasilios K. Kimiskidis, Giacomo Koch, Berthold Langguth, Thomas Nyffeler, Antonio Oliviero, Frank Padberg, Emmanuel Poulet, Simone Rossi, Paolo Maria Rossini, John C. Rothwell, Carlos Schönfeldt-Lecuona, Hartwig R. Siebner, Christina W. Slotema, Charlotte J. Stagg, Josep Valls-Sole, Ulf Ziemann, Walter Paulus and Luis Garcia-Larrea. 'Evidence-Based Guidelines on the Therapeutic Use of Repetitive Transcranial Magnetic Stimulation (rTMS)'. In: *Clinical Neurophysiology* 125.11 (Nov. 2014), pp. 2150–2206. ISSN: 13882457. DOI: 10.1016/j.clinph.2014.05.021. URL: https://linkinghub.elsevier.com/retrieve/pii/S138824571400296X (visited on 06/08/2024).

[Ach+15] U. Rajendra Acharya, Vidya K. Sudarshan, Hojjat Adeli, Jayasree Santhosh, Joel E.W. Koh and Amir Adeli. 'Computer-Aided Diagnosis of Depression Using EEG Signals'. In: *Eur Neurol* 73.5-6 (2015), pp. 329–336. ISSN: 0014-3022, 1421-9913. DOI: 10.1159/000381950. URL: https://www.karger.com/Article/FullText/381950 (visited on 06/07/2023).

[Aka+15] Saime Akdemir Akar, Sadık Kara, Sümeyra Agambayev and Vedat Bilgiç. 'Nonlinear Analysis of EEGs of Patients with Major Depression during Different Emotional States'. In: *Computers in Biology and Medicine* 67 (Dec. 2015), pp. 49–60. ISSN: 00104825. DOI: 10.1016/j.compbiomed.2015.09.019. URL: https://linkinghub.elsevier.com/retrieve/pii/S0010482515003297 (visited on 18/07/2023).

[Bac+15] Maie Bachmann, Kaia Kalev, Anna Suhhova, Jaanus Lass and Hiie Hinrikus. 'Lempel Ziv Complexity of EEG in Depression'. In: *IFMBE Proceedings*. Vol. 45. 1st Jan. 2015, pp. 58–61. ISBN: 978-3-319-11127-8. DOI: 10.1007/978-3-319-11128-5_15.

[BR15] Thomas Burns and Ramesh Rajan. 'Combining Complexity Measures of EEG Data: Multiplying Measures Reveal Previously Hidden Information'. In: *F1000Res* 4 (1st June 2015), p. 137. ISSN: 2046-1402. DOI: 10.12688/f1000research.6590.1. pmid: 26594331. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4648221/ (visited on 13/07/2023).

[Erg+15] Turker Tekin Erguzel, Serhat Ozekes, Selahattin Gultekin, Nevzat Tarhan, Gokben Hizli Sayar and Ali Bayram. 'Neural Network Based Response Prediction of rTMS in Major Depressive Disorder Using QEEG Cordance'. In: *Psychiatry Investig* 12.1 (Jan. 2015), pp. 61–65. ISSN: 1738-3684. DOI: 10.4306/pi.2015.12.1.61. pmid: 25670947. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4310922/ (visited on 25/07/2023).

[Lia+15] Zhenhu Liang, Yinghua Wang, Xue Sun, Duan Li, Logan J. Voss, Jamie W. Sleigh, Satoshi Hagihira and Xiaoli Li. 'EEG Entropy Measures in Anesthesia'. In: *Frontiers in Computational Neuroscience* 9 (2015). ISSN: 1662-5188. URL: https://www.frontiersin.org/articles/10.3389/fncom.2015.00016 (visited on 27/04/2023).

[Sto+15] Miodrag Stokić, Dragan Milovanović, Miloš R. Ljubisavljević, Vanja Nenadović and Milena Čukić. 'Memory Load Effect in Auditory–Verbal Short-Term Memory Task: EEG Fractal and Spectral Analysis'. In: *Exp Brain Res* 233.10 (1st Oct. 2015), pp. 3023–3038. ISSN: 1432-1106. DOI: 10.1007/s00221-015-4372-z. URL: https://doi.org/10.1007/s00221-015-4372-z (visited on 12/07/2023).

[De +16] Alejandro De La Torre-Luque, Xavier Bornas, Maria Balle and Aina Fiol-Veny. 'Complexity and Nonlinear Biomarkers in Emotional Disorders: A Meta-Analytic Study'. In: *Neuroscience & Biobehavioral Reviews* 68 (Sept. 2016), pp. 410–422. ISSN: 01497634. DOI: 10.1016/j.neubiorev.2016.05.023. URL: https://linkinghub.elsevier.com/retrieve/pii/S014976341630094X (visited on 19/01/2024).

[ET16] Turker Erguzel and K. Tarhan. *Machine Learning Approaches to Predict Repetitive Transcranial Magnetic Stimulation Treatment Response in Major Depressive Disorder*. 20th Sept. 2016.

[LC16] Fabien Lotte and Marco Congedo. 'EEG Feature Extraction'. In: *Brain–Computer Interfaces 1*. Ed. by Maureen Clerc, Laurent Bougrain and Fabien Lotte. 1st ed. Wiley, 29th July 2016, pp. 127–143. ISBN: 978-1-119-14497-7. DOI: 10.1002/9781119144977.ch7. URL: https://onlinelibrary.wiley.com/doi/10.1002/9781119144977.ch7 (visited on 26/09/2023).

[MLP16] Jesus Minguillon, Miguel A. Lopez-Gordo and Francisco Pelayo. 'Stress Assessment by Prefrontal Relative Gamma'. In: *Frontiers in Computational Neuroscience* 10 (2016). ISSN: 1662-5188. URL: https://www.frontiersin.org/articles/10.3389/fncom.2016.00101 (visited on 25/07/2023).

[RA16] Mostafa Rostaghi and Hamed Azami. 'Dispersion Entropy: A Measure for Time-Series Analysis'. In: *IEEE Signal Process. Lett.* 23.5 (May 2016), pp. 610–614. ISSN: 1070-9908, 1558-2361. DOI: 10.1109/LSP.2016.2542881. URL: http://ieeexplore.ieee.org/document/7434608/ (visited on 03/03/2024).

[MJA17] Wolfram Müller, Alexander Jung and Helmut Ahammer. 'Advantages and Problems of Nonlinear Methods Applied to Analyze Physiological Time Signals: Human Balance Control as an Example'. In: *Sci Rep* 7.1 (1 26th May 2017), p. 2464. ISSN: 2045-2322. DOI: 10.1038/s41598-017-02665-5. URL: https://www.nature.com/articles/s41598-017-02665-5 (visited on 06/07/2023).

[Mum+17] Wajid Mumtaz, Likun Xia, Mohd Azhar Mohd Yasin, Syed Saad Azhar Ali and Aamir Saeed Malik. 'A Wavelet-Based Technique to Predict Treatment Outcome for Major Depressive Disorder'. In: *PLoS ONE* 12.2 (2nd Feb. 2017). Ed. by Dewen Hu, e0171409. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0171409. URL: https://dx.plos.org/10.1371/journal.pone.0171409 (visited on 29/12/2023).

[Bac+18] Maie Bachmann, Laura Päeske, Kaia Kalev, Katrin Aarma, Andres Lehtmets, Pille Ööpik, Jaanus Lass and Hiie Hinrikus. 'Methods for Classifying Depression in Single Channel EEG Using Linear and Nonlinear Signal Analysis'. In: *Computer Methods and Programs in Biomedicine* 155 (Mar. 2018), pp. 11–17. ISSN: 01692607. DOI: 10.1016/j.cmpb.2017.11.023. URL: https://linkinghub.elsevier.com/retrieve/pii/S016926071631166X (visited on 19/07/2023).

[Bai+18] N.W. Bailey, K.E. Hoy, N.C. Rogasch, R.H. Thomson, S. McQueen, D. Elliot, C.M. Sullivan, B.D. Fulcher, Z.J. Daskalakis and P.B. Fitzgerald. 'Responders to rTMS for Depression Show Increased Fronto-Midline Theta and Theta Connectivity Compared to Non-Responders'. In: *Brain Stimulation* 11.1 (Jan. 2018), pp. 190–203. ISSN: 1935861X. DOI: 10.1016/j.brs.2017.10.015. URL: https://linkinghub.elsevier.com/retrieve/pii/S1935861X17309506 (visited on 02/08/2023).

[Law+18] Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung and Brent J. Lance. 'EEGNet: A Compact Convolutional Network for EEG-based Brain-Computer Interfaces'. In: *J. Neural Eng.* 15.5 (1st Oct. 2018), p. 056013. ISSN: 1741-2560, 1741-2552. DOI: 10.1088/1741-2552/aace8c. arXiv: 1611.08024 [cs, q-bio, stat]. URL: http://arxiv.org/abs/1611.08024 (visited on 02/07/2024).

[Sha+18] Reza Shalbaf, Colleen Brenner, Christopher Pang, Daniel M. Blumberger, Jonathan Downar, Zafiris J. Daskalakis, Joseph Tham, Raymond W. Lam, Faranak Farzan and Fidel Vila-Rodriguez. 'Non-Linear Entropy Analysis in EEG to Predict Treatment Response to Repetitive Transcranial Magnetic Stimulation in Depression'. In: *Frontiers in Pharmacology* 9 (2018). ISSN: 1663-9812. URL: https://www.frontiersin.org/articles/10.3389/fphar.2018.01188 (visited on 18/01/2024).

[Bai+19] Nw Bailey, Ke Hoy, Nc Rogasch, Rh Thomson, S McQueen, D Elliot, Cm Sullivan, Bd Fulcher, Zj Daskalakis and Pb Fitzgerald. 'Differentiating Responders and Non-Responders to rTMS Treatment for Depression

after One Week Using Resting EEG Connectivity Measures'. In: *Journal of Affective Disorders* 242 (Jan. 2019), pp. 68–79. ISSN: 01650327. DOI: 10.1016/j.jad.2018.08.058. URL: https://linkinghub.elsevier.com/retrieve/pii/S0165032718304725 (visited on 02/08/2023).

[CCS19] Jie Chen, Ben M. Chen and Jian Sun. 'Complex System and Intelligent Control: Theories and Applications'. In: *Frontiers Inf Technol Electronic Eng* 20.1 (1st Jan. 2019), pp. 1–3. ISSN: 2095-9230. DOI: 10.1631/FITEE.1910000. URL: https://doi.org/10.1631/FITEE.1910000 (visited on 19/01/2024).

[Cue19] David Cuesta-Frau. 'Slope Entropy: A New Time Series Complexity Estimator Based on Both Symbolic Patterns and Amplitude Information'. In: *Entropy* 21.12 (12 Dec. 2019), p. 1167. ISSN: 1099-4300. DOI: 10.3390/e21121167. URL: https://www.mdpi.com/1099-4300/21/12/1167 (visited on 03/03/2024).

[HMR19] Fatemeh Hasanzadeh, Maryam Mohebbi and Reza Rostami. 'Prediction of rTMS Treatment Response in Major Depressive Disorder Using Machine Learning Techniques and Nonlinear Features of EEG Signal'. In: *Journal of Affective Disorders* 256 (Sept. 2019), pp. 132–142. ISSN: 01650327. DOI: 10.1016/j.jad.2019.05.070. URL: https://linkinghub.elsevier.com/retrieve/pii/S0165032719305543 (visited on 24/05/2023).

[MHM19] Yousef Mohammadi, Mojtaba Hajian and Mohammad Hassan Moradi. 'Discrimination of Depression Levels Using Machine Learning Methods on EEG Signals'. In: *2019 27th Iranian Conference on Electrical Engineering (ICEE)*. 2019 27th Iranian Conference on Electrical Engineering (ICEE). Apr. 2019, pp. 1765–1769. DOI: 10.1109/IranianCEE.2019.8786540.

[Vab+19] Andrius Vabalas, Emma Gowen, Ellen Poliakoff and Alexander J. Casson. 'Machine Learning Algorithm Validation with a Limited Sample Size'. In: *PLoS ONE* 14.11 (7th Nov. 2019). Ed. by Enrique Hernandez-Lemus, e0224365. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0224365. URL: https://dx.plos.org/10.1371/journal.pone.0224365 (visited on 23/08/2024).

[Zhu+19] Jing Zhu, Ying Wang, Rong La, Jiawei Zhan, Junhong Niu, Shuai Zeng and Xiping Hu. 'Multimodal Mild Depression Recognition Based on EEG-EM Synchronization Acquisition Network'. In: *IEEE Access* 7 (2019), pp. 28196–28210. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2019.2901950.

[Cuk+20] Milena Cukic, David Pokrajac, Miodrag Stokic, slobodan Simic, Vlada Radivojevic and Milos Ljubisavljevic. 'EEG Machine Learning with Higuchi Fractal Dimension and Sample Entropy as Features for Successful Detection of Depression'. In: *Cogn Neurodyn* 14.4 (Aug. 2020), pp. 443–455. ISSN: 1871-4080, 1871-4099. DOI: 10.1007/s11571-020-09581-x. arXiv: 1803.05985 [cs, q-bio, stat]. URL: http://arxiv.org/abs/1803.05985 (visited on 24/05/2023).

[IA20] Tomas Iešmantas and Robertas Alzbutas. 'Convolutional Neural Network for Detection and Classification of Seizures in Clinical Data'. In: *Med Biol Eng Comput* 58.9 (Sept. 2020), pp. 1919–1932. ISSN: 0140-0118, 1741-0444. DOI: 10.1007/s11517-020-02208-7. URL: https://link.springer.com/10.1007/s11517-020-02208-7 (visited on 12/07/2023).

[Ras20] Sebastian Raschka. *Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning*. 10th Nov. 2020. DOI: 10.48550/arXiv.1811.12808. arXiv: 1811.12808 [cs, stat]. URL: http://arxiv.org/abs/1811.12808 (visited on 26/03/2024). Pre-published.

[SS20] Dalwinder Singh and Birmohan Singh. 'Investigating the Impact of Data Normalization on Classification Performance'. In: *Applied Soft Computing* 97 (Dec. 2020), p. 105524. ISSN: 15684946. DOI: 10.1016/j.asoc.2019.105524. URL: https://linkinghub.elsevier.com/retrieve/pii/S1568494619302947 (visited on 22/05/2024).

[Sup+20] Supriya Supriya, Siuly Siuly, Hua Wang and Yanchun Zhang. 'Automated Epilepsy Detection Techniques from Electroencephalogram Signals: A Review Study'. In: *Health Inf Sci Syst* 8.1 (12th Oct. 2020), p. 33. ISSN: 2047-2501. DOI: 10.1007/s13755-020-00129-1. pmid: 33088489. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7550618/ (visited on 15/04/2023).

[Uth20] Abicumaran Uthamacumaran. *A Brief Synthesis on Complex Systems*. 19th Aug. 2020. DOI: 10.20944/preprints202008.0400.v1. Preprints: 2020080400. URL: https://www.preprints.org/manuscript/202008.0400/v1 (visited on 19/01/2024). Pre-published.

[Alz+21] Laith Alzubaidi, Jinglan Zhang, Amjad J. Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J. Santamaría, Mohammed A. Fadhel, Muthana Al-Amidie and Laith Farhan. 'Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions'. In: *Journal of Big Data* 8.1 (31st Mar. 2021), p. 53. ISSN: 2196-1115. DOI: 10.1186/s40537-021-00444-8. URL: https://doi.org/10.1186/s40537-021-00444-8 (visited on 19/06/2024).

[Baj21] Nikesh Bajaj. 'Wavelets for EEG Analysis'. In: *Wavelet Theory*. Ed. by Somayeh Mohammady. IntechOpen, 24th Feb. 2021. ISBN: 978-1-83881-947-7. DOI: 10.5772/intechopen.94398. URL: https://www.intechopen.com/books/wavelet-theory/wavelets-for-eeg-analysis (visited on 17/04/2024).

[Ins21] Institute of Health Metrics and Evaluation. 'Global Health Data Exchange (GHDx)'. In: (2021).

[Jad21] 'Frequencies and Rhythms'. In: *How to Read an EEG*. Ed. by Neville M. Jadeja. Cambridge: Cambridge University Press, 2021, pp. 32–39. ISBN: 978-1-108-82564-1. DOI: 10.1017/9781108918923.008. URL: https://www.

cambridge.org/core/books/how-to-read-an-eeg/frequencies-and-rhythms/A8CD5BF72715B1283780DEAF144EC210 (visited on 25/12/2023).

[MM21] Yousef Mohammadi and Mohammad Hassan Moradi. 'Prediction of Depression Severity Scores Based on Functional Connectivity and Complexity of the EEG Signal'. In: *Clin EEG Neurosci* 52.1 (1st Jan. 2021), pp. 52–60. ISSN: 1550-0594. DOI: 10.1177/1550059420965431. URL: https://doi.org/10.1177/1550059420965431 (visited on 18/01/2024).

[Reb+21] Nancy Rebout, Jean-Christophe Lone, Arianna De Marco, Roberto Cozzolino, Alban Lemasson and Bernard Thierry. 'Measuring Complexity in Organisms and Organizations'. In: *R. Soc. open sci.* 8.3 (Mar. 2021), rsos.200895, 200895. ISSN: 2054-5703. DOI: 10.1098/rsos.200895. URL: https://royalsocietypublishing.org/doi/10.1098/rsos.200895 (visited on 19/01/2024).

[Aza+22] Hamed Azami, Luca Faes, Javier Escudero, Anne Humeau-Heurtier and Luiz E.V. Silva. 'ENTROPY ANALYSIS OF UNIVARIATE BIOMEDICAL SIGNALS: REVIEW AND COMPARISON OF METHODS'. In: Willi Freeden and M Zuhair Nashed. *Contemporary Mathematics and Its Applications: Monographs, Expositions and Lecture Notes*. Vol. 04. WORLD SCIENTIFIC, Oct. 2022, pp. 233–286. ISBN: 9789811259395. DOI: 10.1142/9789811259401_0009. URL: https://www.worldscientific.com/doi/10.1142/9789811259401_0009 (visited on 15/04/2023).

[ASR22] Hamed Azami, Saeid Sanei and Tarek Rajji. 'Ensemble Entropy: A Low Bias Approach for Data Analysis'. In: *Knowledge-Based Systems* (1st Sept. 2022), p. 109876. DOI: 10.1016/j.knosys.2022.109876.

[Bel+22] Vincent Beliveau, Ella Hedeboe, Patrick M. Fisher, Vibeke H. Dam, Martin B. Jørgensen, Vibe G. Frokjaer, Gitte M. Knudsen and Melanie Ganz. 'Generalizability of Treatment Outcome Prediction in Major Depressive Disorder Using Structural MRI: A NeuroPharm Study'. In: *NeuroImage: Clinical* 36 (1st Jan. 2022), p. 103224. ISSN: 2213-1582. DOI: 10.1016/j.nicl.2022.103224. URL: https://www.sciencedirect.com/science/article/pii/S2213158222002893 (visited on 16/01/2024).

[22] 'Global, Regional, and National Burden of 12 Mental Disorders in 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019'. In: *The Lancet Psychiatry* 9.2 (1st Feb. 2022), pp. 137–150. ISSN: 2215-0366, 2215-0374. DOI: 10.1016/S2215-0366(21)00395-3. pmid: 35026139. URL: https://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(21)00395-3/fulltext (visited on 26/09/2023).

[Lau+22] Zen Juen Lau, Tam Pham, SH Chen and Dominique Makowski. 'Brain Entropy, Fractal Dimensions and Predictability: A Review of Complexity Meas-

ures for EEG in Healthy and Neuropsychiatric Populations'. In: *European Journal of Neuroscience* 56 (19th Aug. 2022). DOI: 10.1111/ejn.15800.

[Mar+22] M. Margarette Sanchez, L. Borden, N. Alam, A. Noroozi, M. Ravan, P. Flor-Henry and G. Hasey. 'A Machine Learning Algorithm to Discriminating Between Bipolar and Major Depressive Disorders Based on Resting EEG Data'. In: *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). July 2022, pp. 2635–2638. DOI: 10.1109/EMBC48229.2022.9871453.

[Wat+22] Devon Watts, Rafaela Fernandes Pulice, Jim Reilly, Andre R. Brunoni, Flávio Kapczinski and Ives Cavalcante Passos. 'Predicting Treatment Response Using EEG in Major Depressive Disorder: A Machine-Learning Meta-Analysis'. In: *Transl Psychiatry* 12.1 (12th Aug. 2022), p. 332. ISSN: 2158-3188. DOI: 10.1038/s41398-022-02064-z. URL: https://www.nature.com/articles/s41398-022-02064-z (visited on 24/05/2023).

[Zit+22] M. Sami Zitouni, Shu Lih Oh, Jahmunah Vicnesh, Ahsan Khandoker and U. Rajendra Acharya. 'Automated Recognition of Major Depressive Disorder from Cardiovascular and Respiratory Physiological Signals'. In: *Frontiers in Psychiatry* 13 (2022). ISSN: 1664-0640. URL: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.970993 (visited on 19/07/2023).

[23] *10-20-System*. In: *Wikipedia*. 23rd Oct. 2023. URL: https://de.wikipedia.org/w/index.php?title=10-20-System&oldid=238439510 (visited on 07/07/2024).

[Bri23] The Editors of Encyclopaedia Britannica. *Cerebrum*. In: *Encyclopedia Britannica*. 24th Oct. 2023. URL: https://www.britannica.com/science/cerebrum (visited on 19/01/2024).

[Hla+23] Kateřina Hlaváčková-Schindler, Christina Pacher, Claudia Plant, Mykola Lazarenko, Milan Paluš, Jaroslav Hlinka, Aditi Kathpalia and Martin Brunovský. 'Pattern Discovery in an EEG Database of Depression Patients: Preliminary Results'. In: *2023 14th International Conference on Measurement*. 2023 14th International Conference on Measurement. May 2023, pp. 80–83. DOI: 10.23919/MEASUREMENT59122.2023.10164584. URL: https://ieeexplore.ieee.org/document/10164584/?arnumber=10164584 (visited on 29/07/2024).

[Joy23] Robert Joynt. *Brainstem*. In: *Encyclopedia Britannica*. 16th Dec. 2023. URL: https://www.britannica.com/science/brainstem.

[Khu23] Rani Khushaba. 'Transfer Entropy Illustrated'. 16th Oct. 2023. URL: https://www.youtube.com/watch?v=3AassSBKRmY (visited on 13/03/2024).

[MA23] Kenia A. Maldonado and Khalid Alsayouri. 'Physiology, Brain'. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2023. pmid: 31869182. URL: http://www.ncbi.nlm.nih.gov/books/NBK551718/ (visited on 29/11/2023).

[SIF23] Nazmus Sakib, Md Kafiul Islam and Tasnuva Faruk. 'Machine Learning Model for Computer-Aided Depression Screening among Young Adults Using Wireless EEG Headset'. In: *Computational Intelligence and Neuroscience* 2023 (31st May 2023). Ed. by Abdul Rehman Javed, pp. 1–23. ISSN: 1687-5273, 1687-5265. DOI: 10.1155/2023/1701429. URL: https://www.hindawi.com/journals/cin/2023/1701429/ (visited on 25/03/2024).

[Akr24] Karin Akre. *Cerebral Cortex*. In: *Encyclopedia Britannica*. 2nd July 2024. URL: https://www.britannica.com/science/cerebral-cortex (visited on 06/07/2024).

[Bri24a] The Editors of Encyclopaedia Britannica. *Brain*. In: *Encyclopedia Britannica*. 31st July 2024. URL: https://www.britannica.com/science/brain (visited on 06/08/2024).

[Bri24b] The Editors of Encyclopaedia Britannica. *Cerebellum*. In: *Encyclopedia Britannica*. 11th Jan. 2024. URL: https://www.britannica.com/science/cerebellum (visited on 19/01/2024).

[Kra24] Melanija Kraljevska. 'MASTERARBEIT/MASTER'S THESIS'. In: (2024).

[ZZL24] Zhi Zhang, Sheng-hua Zhong and Yan Liu. 'TorchEEGEMO: A Deep Learning Toolbox towards EEG-based Emotion Recognition'. In: *Expert Systems with Applications* 249 (1st Sept. 2024), p. 123550. ISSN: 0957-4174. DOI: 10.1016/j.eswa.2024.123550. URL: https://www.sciencedirect.com/science/article/pii/S0957417424004159 (visited on 02/07/2024).

Acronyms

ApEn approximate entropy. 12, 16, 32, 33

APV alpha power variability. 11

AUC area under the curve. 14

CD correlation dimension. 12, 14–17, 39

CE conditional entropy. 13

CNN Convolutional Neural Network. 31, 55, 67

CV cross validation. iii, v, ix, xi, 28, 29, 55–57, 60, 61, 67, 73, 105

CWT continuous wavelet transform. 8

DFA Detrended Fluctuation Analysis. 15, 39, 43

DispEn dispersion entropy. 35, 36, 53, 56, 61, 68, 74

DistEn distribution entropy. 13

DT Decision Tree. 15, 17

DWT discrete wavelet transform. xi, 8

EEG electroencephalography. iii, v, vi, ix, 1, 2, 5–9, 11–15, 19, 21, 27, 28, 31, 38, 39, 43, 44, 55, 67, 69, 70, 73, 74, 93, 94, 107

FuzzyEn fuzzy entropy. 13, 16, 33, 34, 43

HFD Higuchi's fractal dimension. 12, 15, 38, 39, 56, 68, 74

ICA independent component analysis. 11

KC Kolmogorov complexity. 15

KFD Katz fractal dimension. 12, 14–17

KNN K-Nearest Neighbors. 16, 17

Kol Kolmogorov entropy. 12, 16

LASSO least absolute shrinkage and selection operator. 29

LLE Lyapunov exponent. 12, 14–17

LR Logistic Regression. 15, 17, 30, 31, 56, 60, 107

LZC Lempel-Ziv complexity. 12, 14–17, 36, 38, 53, 68, 74

MADRS Montgomery-Åsberg Depression Rating Scale. xi, 19–23

MDD major depressive disorder. ix, 1, 2, 11, 13–17, 20, 67, 73, 74

MLP Multilayer Perceptron. 15, 16

mRMR minimum Redundancy Maximum Relevance. iii, v, 29, 55, 56, 60, 69, 73

NB Naive Bayes. 15, 16, 30, 31, 60, 69

PCA principal component analysis. 11

PermEn permutation entropy. 12–14, 16, 34–37, 54, 68, 74

RF Random Forest. 15, 16, 30, 60, 62, 69, 74, 107

RGP relative gamma power. 11

rTMS repetitive transcranial magnetic stimulation. 13, 14

SampEn sample entropy. 12, 15, 16, 32, 33

SASI spectral asymmetry index. 11, 15, 38, 60, 68

ShannonEn Shannon entropy. 12, 13, 15, 16, 32, 35

SIE symbolic influence entropy. iii, v, xi, xii, 38, 43, 44, 48–50, 53, 56, 58, 60, 63, 68, 69, 73, 74, 108

SlopEn slope entropy. 36

SMOTE Synthetic Minority Over-sampling Technique. 28

SpectralEn spectral entropy. 16, 34

STE symbolic transfer entropy. iii, v, xi, xii, 13, 16, 37, 38, 43–45, 48–50, 53, 54, 56, 58, 60, 61, 63, 68–70, 74, 108

SVM Support Vector Machine. iii, v, 15–17, 30, 56, 60, 62, 69, 74, 106, 107

XGB eXtreme Gradient Boosting. 16, 30

A. Appendix

A.1. Feature separation analysis

The two tables below present the results of two metrics used to assess the connectivity of electrodes under different measurement settings and coefficients. The first metric, Weighted Connection Count, balances the number of connections with their corresponding weights. The second metric, Top Percentile Mean, emphasizes the weights of the top 10% of connections, resulting in higher values for electrodes that have stronger, though possibly fewer, connections.

Table A.1.: Weighted connection count for key EEG channels across multiple wavelet coefficients, highlighting channels with significant connectivity differences between responders and non-responders.

Channel	Measure	Coefficient	Weighted Connection Count
T6	SIE-Del-3-Dim-5	D2	0.065390
O2	SIE-Del-3-Dim-5	D2	0.065108
O2	STE-Del-5-Dim-7	D2	0.058287
F4	SIE-Del-3-Dim-5	D2	0.056031
F3	STE-Del-2-Dim-3	D1	0.052087
T6	STE-Del-3-Dim-5	D2	0.051874
F3	SIE-Del-3-Dim-5	D2	0.051508
T6	STE-Del-2-Dim-3	D4	0.051351
O1	SIE-Del-3-Dim-5	D2	0.047121
C3	SIE-Del-3-Dim-5	D2	0.046805
C3	STE-Del-3-Dim-5	D1	0.044178
F8	STE-Del-5-Dim-7	A4	0.042995
C4	STE-Del-2-Dim-3	A4	0.042194
C4	STE-Del-2-Dim-3	D4	0.039769
T5	STE-Del-5-Dim-7	D3	0.037824

Table A.2.: Top percentile mean of transfer/influence entropy differences for key EEG channels, identifying channels with the most substantial connectivity variations between treatment responders and non-responders.

Channel	Measure	Coefficient	Top Percentile Mean
C4	STE-Del-2-Dim-3	D4	0.376772
Fp2	STE-Del-2-Dim-3	D4	0.376772
C4	STE-Del-2-Dim-3	A4	0.373488
Fp1	STE-Del-2-Dim-3	A4	0.373488
C3	STE-Del-2-Dim-3	D4	0.368492
T6	STE-Del-2-Dim-3	D4	0.368492
F3	STE-Del-3-Dim-5	D1	0.352291
C3	STE-Del-3-Dim-5	D1	0.352291
F4	STE-Del-2-Dim-3	D4	0.350769
C3	STE-Del-2-Dim-3	A4	0.341062
F7	STE-Del-2-Dim-3	D3	0.333795
O1	STE-Del-2-Dim-3	D3	0.333795
C4	SIE-Del-2-Dim-3	D1	0.326852
T3	SIE-Del-2-Dim-3	D1	0.326852
T5	SIE-Del-2-Dim-3	A4	0.324681

A.2. Statistical analysis full results

Here we report all the combinations of variables that lead to a p-value below the 0.05 threshold in the Hotelling T^2 and t-test.

Table A.3.: Full results of the t-test and Hotelling's T^2 test for all three-way interactions between channels, coefficients, and measures.

Channel		
C3-T6	52.45	0.0162
C4-F8	50.91	0.0212
T4-Pz	49.45	0.0272
C4-O2	48.68	0.0311
C4-T4	48.25	0.0335
Fp1-Cz	46.41	0.0454
O2-F8	46.05	0.0482
Fp2-F4	45.90	0.0493
Measure		
DispEn	313.98	0.0280

Continued on next page

Table A.3 – continued from previous page

Combination	t-stat	p-value
LZC	306.38	0.0359
Channel Coeff		
Fp1-C4_A4	24.42	0.0004
T5_A4	29.05	0.0019
P4-Pz_D2	20.13	0.0021
O1-F8_A4	18.61	0.0038
C3-T6_D4	17.85	0.0050
Fp1-Cz_D2	17.02	0.0069
O2-F8_A4	15.35	0.0127
F4-O1_D4	15.23	0.0133
C4-T6_A4	15.18	0.0136
C3-O1_D1	14.83	0.0154
C3_D4	21.60	0.0191
Fp1-Cz_D3	14.13	0.0200
C3-T6_D1	14.07	0.0205
P4-T5_D2	13.90	0.0218
C4-O2_A4	13.82	0.0224
C4-F8_A4	13.50	0.0252
C3-O2_D1	13.49	0.0253
Fp2-O2_D3	13.40	0.0261
P4_D3	20.39	0.0272
Fp1-C3_D3	13.16	0.0286
C3-F8_A4	13.11	0.0290
O1_D4	20.08	0.0298
F4-T5_A4	13.04	0.0298
F8-T5_D2	12.99	0.0303
F4-F8_A4	12.98	0.0305
P3-Cz_A4	12.98	0.0305
T4-Pz_D4	12.96	0.0307
Fp1-T3_D3	12.89	0.0315
T6_D3	19.85	0.0318
P4-F8_A4	12.85	0.0320
Fp2-F8_A4	12.63	0.0345
O1_D3	19.24	0.0379
C3-C4_A4	12.33	0.0385
O2_D3	19.13	0.0391
F3-O2_D1	12.27	0.0394
P3-O2_D3	12.23	0.0399
P4-O1_D4	12.23	0.0400

Continued on next page

Table A.3 – continued from previous page

Combination	t-stat	p-value
F3-P3_D1	12.14	0.0412
F4-Pz_D3	12.11	0.0417
P3-F8_A4	12.10	0.0419
Fp2-O1_D2	12.10	0.0419
C4-T4_A4	11.94	0.0444
C4-T6_D1	11.81	0.0465
C3-Pz_D1	11.78	0.0469
T4-Pz_D1	11.75	0.0475
T4-Pz_D3	11.67	0.0488
F3-C4_A4	11.67	0.0488
O2-T5_A4	11.64	0.0494
Channel Measure		
Fp1-Cz_STE-Del-2-Dim-3	26.26	0.0002
C3-T6_STE-Del-2-Dim-3	20.89	0.0016
C4-F8_STE-Del-5-Dim-7	18.33	0.0042
P3-F8_STE-Del-5-Dim-7	18.33	0.0042
F7-F8_STE-Del-5-Dim-7	18.33	0.0042
C3-F8_STE-Del-5-Dim-7	18.31	0.0042
O2-F8_STE-Del-5-Dim-7	18.21	0.0044
Fp1-F8_STE-Del-5-Dim-7	18.21	0.0044
T5_PermEn	18.18	0.0044
F3-F8_STE-Del-5-Dim-7	18.18	0.0044
O1-F8_STE-Del-5-Dim-7	18.12	0.0045
P4-F8_STE-Del-5-Dim-7	17.99	0.0047
F4-F8_STE-Del-5-Dim-7	17.97	0.0048
Fp2-F8_STE-Del-5-Dim-7	17.95	0.0048
P3-O2_STE-Del-3-Dim-5	16.97	0.0070
C4-T4_STE-Del-2-Dim-3	16.17	0.0094
C4-O1_STE-Del-3-Dim-5	15.94	0.0103
Fp1-C4_STE-Del-2-Dim-3	15.55	0.0118
C3-T6_SIE-Del-3-Dim-5	15.36	0.0127
C4-T4_SIE-Del-2-Dim-3	15.23	0.0133
C4-F8_STE-Del-2-Dim-3	14.22	0.0194
C3-O2_SIE-Del-3-Dim-5	14.08	0.0204
T6_PermEn	13.71	0.0233
F3-Cz_STE-Del-2-Dim-3	13.62	0.0241
Fp1-T4_SIE-Del-2-Dim-3	13.59	0.0244
Fp1-C3_STE-Del-2-Dim-3	13.29	0.0272
T6_DispEn	12.90	0.0314

Continued on next page

Table A.3 – continued from previous page

Combination	**t-stat**	**p-value**
Fp1-T3_STE-Del-2-Dim-3	12.83	0.0322
O2_PermEn	12.78	0.0328
T6_SpecEn	12.48	0.0366
Fp2-C3_STE-Del-3-Dim-5	12.29	0.0391
C3_BP-gamma	2.07	0.0397
C3-Fz_STE-Del-2-Dim-3	12.18	0.0407
C3-T4_SIE-Del-3-Dim-5	12.15	0.0411
P3-F7_STE-Del-3-Dim-5	12.05	0.0426
F3-C3_SIE-Del-3-Dim-5	12.05	0.0426
F3-O1_STE-Del-3-Dim-5	11.98	0.0438
F7-T6_SIE-Del-2-Dim-3	11.95	0.0441
C4-O2_STE-Del-3-Dim-5	11.84	0.0460
T6_HP-mob	11.82	0.0462
Fp1-T6_SIE-Del-2-Dim-3	11.74	0.0476
F8_SASI	-1.98	0.0493
Coeff Measure		
A4_PermEn	38.50	0.0252
Channel Coeff Measure		
C3-T6_D4_STE-Del-2-Dim-3	3.83	0.0002
Fp1-Cz_D3_STE-Del-2-Dim-3	-3.70	0.0003
Fp1-Cz_D2_STE-Del-2-Dim-3	-3.70	0.0003
Fp1-T3_D3_STE-Del-2-Dim-3	-3.23	0.0015
T5_A4_PermEn	-3.21	0.0016
F4-F8_A4_STE-Del-5-Dim-7	3.16	0.0019
P3-F8_A4_STE-Del-5-Dim-7	3.15	0.0019
O2-F8_A4_STE-Del-5-Dim-7	3.14	0.0020
O1-F8_A4_STE-Del-5-Dim-7	3.14	0.0020
F7-F8_A4_STE-Del-5-Dim-7	3.13	0.0020
Fp2-F8_A4_STE-Del-5-Dim-7	3.13	0.0021
C3-F8_A4_STE-Del-5-Dim-7	3.13	0.0021
F3-F8_A4_STE-Del-5-Dim-7	3.13	0.0021
Fp1-F8_A4_STE-Del-5-Dim-7	3.11	0.0022
P4-F8_A4_STE-Del-5-Dim-7	3.11	0.0022
P4-Pz_D2_STE-Del-2-Dim-3	-3.11	0.0022
C4-F8_A4_STE-Del-5-Dim-7	3.11	0.0022
Fp1-C4_A4_STE-Del-2-Dim-3	-3.08	0.0024
Fp1-C3_D3_STE-Del-2-Dim-3	-3.06	0.0026
C3-O2_D2_SIE-Del-3-Dim-5	2.92	0.0039

Continued on next page

Table A.3 – continued from previous page

Combination	t-stat	p-value
O1_D1_DispEn	-2.90	0.0042
C4-T6_A4_SIE-Del-2-Dim-3	-2.87	0.0046
C4-T4_D1_STE-Del-2-Dim-3	2.86	0.0047
C3-Pz_D1_SIE-Del-2-Dim-3	2.86	0.0048
C4-T6_D1_STE-Del-2-Dim-3	2.85	0.0049
Fp1-C4_A4_SIE-Del-3-Dim-5	2.84	0.0050
F4-T5_A4_SIE-Del-2-Dim-3	-2.81	0.0055
Fp1-T4_A4_SIE-Del-2-Dim-3	-2.80	0.0057
C4-O2_A4_SIE-Del-3-Dim-5	2.79	0.0059
C3-T6_D2_SIE-Del-3-Dim-5	2.78	0.0059
C4-T4_D3_SIE-Del-2-Dim-3	-2.77	0.0062
Fp2-O2_D2_SIE-Del-3-Dim-5	2.76	0.0064
Fz-Cz_D3_STE-Del-2-Dim-3	-2.74	0.0068
T6-Pz_D1_SIE-Del-3-Dim-5	2.71	0.0075
F3-Cz_D3_STE-Del-2-Dim-3	-2.69	0.0078
T6-Cz_D3_STE-Del-2-Dim-3	-2.69	0.0079
C3-T4_D2_SIE-Del-3-Dim-5	2.69	0.0079
O2_D1_DispEn	-2.65	0.0087
C3-T5_D2_SIE-Del-3-Dim-5	2.65	0.0087
T6_D3_HP-comp	2.65	0.0087
F7-T6_D4_STE-Del-2-Dim-3	2.65	0.0088
O2-T5_A4_SIE-Del-3-Dim-5	2.65	0.0088
O1-F7_D3_STE-Del-2-Dim-3	-2.64	0.0089
P4-T5_D2_STE-Del-2-Dim-3	2.64	0.0090
C3-O1_D1_STE-Del-3-Dim-5	-2.64	0.0091
T6_D2_HP-mob	2.62	0.0096
O2_D1_FuzzyEn	-2.60	0.0102
O1_D1_FuzzyEn	-2.58	0.0107
F3-T5_D1_STE-Del-2-Dim-3	2.57	0.0109
C3_D2_SlopEn	-2.57	0.0109
O2_D3_SpecEn	2.57	0.0111
O2-Cz_D4_STE-Del-5-Dim-7	-2.56	0.0112
F3-C3_D1_SIE-Del-3-Dim-5	-2.56	0.0113
F4-Cz_D4_STE-Del-5-Dim-7	-2.56	0.0113
O1-F8_A4_STE-Del-2-Dim-3	-2.56	0.0114
O1_D3_HP-comp	2.55	0.0115
F3-P3_D1_STE-Del-2-Dim-3	2.55	0.0115
Fz-Cz_D4_STE-Del-5-Dim-7	-2.55	0.0116
F7-T6_D1_SIE-Del-2-Dim-3	-2.55	0.0116

Continued on next page

Table A.3 – continued from previous page

Combination	**t-stat**	**p-value**
O1_D1_HP-comp	-2.55	0.0116
P3-Cz_A4_SIE-Del-3-Dim-5	2.55	0.0117
T6_D2_FuzzyEn	-2.55	0.0117
F3-Cz_D4_STE-Del-5-Dim-7	-2.55	0.0118
O1-Cz_A4_STE-Del-2-Dim-3	-2.54	0.0121
O2_D3_HP-comp	2.53	0.0123
F8-T5_D2_STE-Del-2-Dim-3	2.53	0.0123
O2_D1_HP-comp	-2.53	0.0125
C4-Cz_D4_STE-Del-5-Dim-7	-2.52	0.0126
T6_D1_DispEn	-2.52	0.0127
T6_A4_DispEn	-2.52	0.0127
P4-Cz_D4_STE-Del-5-Dim-7	-2.51	0.0128
F7-Cz_D4_STE-Del-5-Dim-7	-2.51	0.0130
O1-T4_A4_STE-Del-2-Dim-3	-2.51	0.0132
T6_D2_PermEn	-2.50	0.0133
T4-Cz_D4_STE-Del-5-Dim-7	-2.50	0.0133
Fp1-Cz_D4_STE-Del-5-Dim-7	-2.50	0.0134
Fz_D3_HP-comp	2.49	0.0135
C4-F8_D1_STE-Del-2-Dim-3	2.49	0.0135
F8-Cz_D4_STE-Del-5-Dim-7	-2.49	0.0137
T5-Cz_D4_STE-Del-5-Dim-7	-2.49	0.0137
Fp2-Cz_D4_STE-Del-5-Dim-7	-2.49	0.0137
Cz_D1_HiguchiFD	2.49	0.0137
F7-T3_D4_SIE-Del-2-Dim-3	2.49	0.0138
T3-Cz_D4_STE-Del-5-Dim-7	-2.49	0.0138
T6-Cz_D4_STE-Del-5-Dim-7	-2.49	0.0139
F3-C3_D1_STE-Del-3-Dim-5	-2.47	0.0143
O1_D2_DispEn	-2.46	0.0147
P3-Cz_D4_STE-Del-5-Dim-7	-2.46	0.0147
C3-Cz_D4_STE-Del-5-Dim-7	-2.46	0.0148
O1-Cz_D4_STE-Del-5-Dim-7	-2.45	0.0151
O2_D2_PermEn	-2.45	0.0152
C4-T4_A4_STE-Del-2-Dim-3	2.45	0.0152
C3-C4_A4_STE-Del-2-Dim-3	-2.45	0.0152
C3-O2_A4_SIE-Del-3-Dim-5	2.45	0.0153
O1_D2_HP-mob	2.45	0.0153
C4_A4_SpecEn	-2.45	0.0154
F3-C4_A4_STE-Del-2-Dim-3	-2.44	0.0155
C3-T6_D1_STE-Del-3-Dim-5	-2.44	0.0156

Continued on next page

Table A.3 – continued from previous page

Combination	t-stat	p-value
T6_D2_SpecEn	-2.44	0.0157
T6_D2_DispEn	-2.44	0.0157
Fp1-O1_D2_SIE-Del-2-Dim-3	-2.44	0.0158
C3-Fz_A4_STE-Del-2-Dim-3	-2.43	0.0162
T6_D3_SpecEn	2.43	0.0163
O1_D3_DispEn	2.42	0.0165
P3-O2_D4_STE-Del-2-Dim-3	2.42	0.0168
Fp1-T6_A4_SIE-Del-2-Dim-3	-2.41	0.0170
T6_A4_SpecEn	-2.41	0.0170
F3-O1_D2_SIE-Del-3-Dim-5	2.39	0.0180
T3-T6_D4_STE-Del-2-Dim-3	2.38	0.0182
O2_D2_DispEn	-2.38	0.0182
O2_D2_HP-mob	2.38	0.0182
C4-O1_D1_STE-Del-3-Dim-5	-2.38	0.0183
C4-O1_D2_STE-Del-3-Dim-5	-2.38	0.0185
T6_D3_FuzzyEn	2.37	0.0187
Fp2-C3_D4_STE-Del-2-Dim-3	-2.37	0.0187
F3-C3_D2_STE-Del-3-Dim-5	-2.37	0.0190
O1_D3_SpecEn	2.36	0.0192
T6_D1_HP-comp	-2.36	0.0195
F3-O1_D4_STE-Del-3-Dim-5	-2.36	0.0195
C3_D1_FuzzyEn	-2.35	0.0198
P4-O1_D1_STE-Del-3-Dim-5	-2.35	0.0198
C4-O2_D1_STE-Del-2-Dim-3	2.35	0.0200
O1_D2_FuzzyEn	-2.35	0.0200
T6-Fz_D3_STE-Del-2-Dim-3	-2.35	0.0201
C3-O1_D2_STE-Del-3-Dim-5	-2.34	0.0203
F8-T6_D3_SIE-Del-2-Dim-3	-2.34	0.0203
O2_D3_FuzzyEn	2.34	0.0204
Fp2_D1_SlopEn	2.34	0.0205
Fz_D3_DispEn	2.33	0.0208
Fz_D3_SpecEn	2.33	0.0211
O1_D2_PermEn	-2.33	0.0212
C4-Cz_A4_SIE-Del-3-Dim-5	2.33	0.0212
F3-F7_D2_SIE-Del-2-Dim-3	-2.32	0.0213
C3-F8_D4_STE-Del-3-Dim-5	-2.32	0.0214
C3-O1_D2_SIE-Del-3-Dim-5	2.32	0.0215
T4-T6_D4_STE-Del-2-Dim-3	2.32	0.0216
C3_D3_HP-comp	2.32	0.0217

Continued on next page

Table A.3 – continued from previous page

Combination	t-stat	p-value
Fp2-C3_D2_STE-Del-3-Dim-5	-2.32	0.0217
P4-Cz_A4_STE-Del-2-Dim-3	-2.32	0.0217
T4-T5_A4_SIE-Del-3-Dim-5	2.31	0.0218
C4-Fz_D1_STE-Del-2-Dim-3	2.31	0.0220
F3-O2_D2_SIE-Del-3-Dim-5	2.31	0.0222
T6_D1_FuzzyEn	-2.31	0.0223
F3-O2_D1_STE-Del-2-Dim-3	2.30	0.0228
F3-P4_D1_STE-Del-2-Dim-3	2.29	0.0232
F4-C3_D2_STE-Del-3-Dim-5	-2.29	0.0233
O2_D2_HiguchiFD	-2.29	0.0233
C4-T3_D3_SIE-Del-2-Dim-3	-2.28	0.0240
F3-T6_D1_STE-Del-2-Dim-3	2.27	0.0245
F4-P4_D4_SIE-Del-2-Dim-3	-2.27	0.0246
Fp1-C3_D2_STE-Del-3-Dim-5	-2.26	0.0248
C4-T5_D3_STE-Del-5-Dim-7	-2.25	0.0256
T4_D4_PermEn	-2.25	0.0260
Fp2-T5_D3_STE-Del-5-Dim-7	-2.24	0.0263
T3-T5_D3_STE-Del-5-Dim-7	-2.24	0.0263
T6_D2_HP-comp	-2.24	0.0266
F7-T6_A4_SIE-Del-2-Dim-3	-2.24	0.0266
F4-T5_D3_STE-Del-5-Dim-7	-2.24	0.0267
F3-T5_D3_STE-Del-5-Dim-7	-2.23	0.0268
Fz_D1_FuzzyEn	-2.23	0.0268
F7-Cz_A4_STE-Del-2-Dim-3	-2.23	0.0269
O2-T5_D3_STE-Del-5-Dim-7	-2.23	0.0271
P4-T5_D3_STE-Del-5-Dim-7	-2.23	0.0272
T4-T5_D3_STE-Del-5-Dim-7	-2.23	0.0274
O2_D2_FuzzyEn	-2.22	0.0274
F7-T5_D3_STE-Del-5-Dim-7	-2.22	0.0274
F3-Cz_D1_STE-Del-2-Dim-3	2.22	0.0275
Fp2-C3_D1_STE-Del-3-Dim-5	-2.22	0.0277
O2-T6_D2_SIE-Del-3-Dim-5	2.22	0.0277
F4-O2_D2_SIE-Del-3-Dim-5	2.21	0.0281
O1-F8_A4_SIE-Del-3-Dim-5	2.21	0.0281
Fp1-T5_D3_STE-Del-5-Dim-7	-2.21	0.0283
F8-Pz_D4_SIE-Del-2-Dim-3	-2.21	0.0285
F4-Cz_D3_STE-Del-2-Dim-3	-2.21	0.0286
T5-T6_D1_STE-Del-3-Dim-5	-2.20	0.0288
O1_D2_HiguchiFD	-2.20	0.0291

Continued on next page

Table A.3 – continued from previous page

Combination	t-stat	p-value
C3_D2_FuzzyEn	-2.20	0.0292
C3_D1_DispEn	-2.20	0.0293
P3-T5_D3_STE-Del-5-Dim-7	-2.20	0.0293
O1-T5_D3_STE-Del-5-Dim-7	-2.20	0.0294
C3_A4_FuzzyEn	-2.20	0.0294
T4-Pz_D1_STE-Del-2-Dim-3	-2.19	0.0296
P4-O1_D2_STE-Del-3-Dim-5	-2.19	0.0297
T4-T6_A4_SIE-Del-3-Dim-5	2.19	0.0297
O2_D2_HP-comp	-2.19	0.0298
C3-T5_D3_STE-Del-5-Dim-7	-2.19	0.0299
T6_D3_DispEn	2.19	0.0300
Fp1-T6_A4_SIE-Del-3-Dim-5	2.18	0.0303
Fz_D3_FuzzyEn	2.18	0.0303
F7-T6_D2_SIE-Del-3-Dim-5	2.18	0.0306
C3_D3_SpecEn	2.18	0.0306
F3_D3_HP-mob	-2.18	0.0309
C3-Fz_D2_SIE-Del-3-Dim-5	2.17	0.0310
Fz_D2_HiguchiFD	-2.17	0.0312
F8-T5_D3_STE-Del-5-Dim-7	-2.17	0.0313
T3-Fz_A4_SIE-Del-2-Dim-3	2.17	0.0314
F4-O1_D1_STE-Del-3-Dim-5	-2.17	0.0316
F3-O1_D1_STE-Del-3-Dim-5	-2.16	0.0321
Fz_D1_DispEn	-2.16	0.0322
O1-T6_D1_STE-Del-3-Dim-5	-2.16	0.0323
Cz-Pz_D1_STE-Del-2-Dim-3	2.16	0.0323
Fz_D3_HP-mob	-2.14	0.0334
O1-T5_D2_STE-Del-3-Dim-5	-2.14	0.0335
C3-Fz_D1_STE-Del-2-Dim-3	2.14	0.0335
F3-T5_D2_SIE-Del-3-Dim-5	2.14	0.0338
O2_D1_HP-mob	2.14	0.0338
C3_D3_HP-mob	-2.14	0.0338
O1_D2_HP-comp	-2.14	0.0339
F4-T6_D1_STE-Del-3-Dim-5	-2.14	0.0340
Fp2-O1_D1_STE-Del-5-Dim-7	2.13	0.0342
F3_D2_HP-comp	-2.13	0.0342
Fp1-O1_D1_STE-Del-5-Dim-7	2.13	0.0345
F7-T3_A4_STE-Del-3-Dim-5	2.12	0.0352
P4-O1_D1_STE-Del-5-Dim-7	2.12	0.0353
C4-O1_D1_STE-Del-5-Dim-7	2.12	0.0353

Continued on next page

Table A.3 – continued from previous page

Combination	**t-stat**	**p-value**
F3_D3_DispEn	2.12	0.0354
F3-F4_D2_SIE-Del-3-Dim-5	2.12	0.0356
F3-O1_D1_STE-Del-5-Dim-7	2.12	0.0356
P3-T6_D1_STE-Del-3-Dim-5	-2.12	0.0358
P3-O2_D2_STE-Del-3-Dim-5	-2.11	0.0360
P4-O2_D2_SIE-Del-2-Dim-3	-2.11	0.0360
F3_D2_SpecEn	-2.11	0.0360
C3_D2_DispEn	-2.11	0.0362
Fp2-O2_D3_SIE-Del-3-Dim-5	-2.11	0.0363
C3-O1_D1_STE-Del-5-Dim-7	2.11	0.0364
C4_D3_HP-comp	2.11	0.0365
C3_D2_HP-comp	-2.11	0.0366
T4-Pz_D4_SIE-Del-3-Dim-5	-2.10	0.0371
T6_D2_HiguchiFD	-2.10	0.0373
F8-T6_D2_SIE-Del-2-Dim-3	-2.10	0.0375
F4-F7_D1_STE-Del-2-Dim-3	2.10	0.0375
F4-O1_D1_STE-Del-5-Dim-7	2.10	0.0375
Fp1-F3_D3_STE-Del-2-Dim-3	-2.09	0.0377
F3_D1_SlopEn	2.09	0.0377
P3-O1_D1_STE-Del-5-Dim-7	2.09	0.0378
F4-O1_D2_SIE-Del-3-Dim-5	2.09	0.0378
F3-P4_D2_SIE-Del-3-Dim-5	2.09	0.0378
T5_D2_SlopEn	-2.09	0.0382
T6_D1_SpecEn	-2.08	0.0388
O2-T5_D1_STE-Del-3-Dim-5	-2.08	0.0389
P3-Cz_A4_SIE-Del-2-Dim-3	-2.08	0.0391
O2-T4_A4_SIE-Del-3-Dim-5	2.08	0.0392
T5-Pz_A4_SIE-Del-3-Dim-5	2.08	0.0394
C3-O1_D3_SIE-Del-3-Dim-5	-2.07	0.0396
O1-T6_D2_SIE-Del-3-Dim-5	2.07	0.0397
C3_all_BP-gamma	2.07	0.0397
Fz_D2_SpecEn	-2.07	0.0398
T5_D2_HiguchiFD	-2.07	0.0398
C3-C4_A4_STE-Del-5-Dim-7	2.07	0.0402
O2_D3_DispEn	2.07	0.0403
Fp2-F4_D1_STE-Del-3-Dim-5	-2.07	0.0403
F3-T6_D1_STE-Del-3-Dim-5	-2.07	0.0403
F4-F8_D4_STE-Del-3-Dim-5	-2.06	0.0406
F4-C4_A4_STE-Del-5-Dim-7	2.06	0.0407

Continued on next page

Table A.3 – continued from previous page

Combination	t-stat	p-value
C4-F7_A4_SIE-Del-3-Dim-5	2.06	0.0410
O1-O2_D3_SIE-Del-3-Dim-5	-2.06	0.0412
T3-T6_D2_SIE-Del-2-Dim-3	-2.06	0.0413
P4_A4_SpecEn	-2.05	0.0415
P4-O1_D4_STE-Del-3-Dim-5	-2.05	0.0415
Fp2_D1_DispEn	-2.05	0.0416
O2-F7_D3_STE-Del-3-Dim-5	2.05	0.0417
O1-O2_D2_SIE-Del-3-Dim-5	2.05	0.0417
T3-T6_D1_STE-Del-3-Dim-5	-2.05	0.0420
C3-Cz_D1_STE-Del-3-Dim-5	-2.05	0.0421
F3-T5_D3_SIE-Del-3-Dim-5	-2.05	0.0422
O1_D2_SpecEn	-2.04	0.0428
C4_D1_FuzzyEn	-2.04	0.0428
Fp2_D2_FuzzyEn	-2.04	0.0429
Fp2-T6_D2_SIE-Del-3-Dim-5	2.04	0.0432
P3_D1_LZC	2.03	0.0435
T3-T6_D2_SIE-Del-3-Dim-5	2.03	0.0436
Fp2-T5_D2_SIE-Del-3-Dim-5	2.03	0.0437
Fp1-F3_D1_STE-Del-5-Dim-7	2.03	0.0437
Fp2-F3_D1_STE-Del-5-Dim-7	2.03	0.0440
Fp2-F4_D2_STE-Del-3-Dim-5	-2.03	0.0443
C4-O2_A4_SIE-Del-2-Dim-3	-2.02	0.0444
F3-C4_A4_STE-Del-5-Dim-7	2.02	0.0445
Fp2-C4_A4_STE-Del-5-Dim-7	2.02	0.0445
Fp2-O1_D2_STE-Del-3-Dim-5	-2.02	0.0446
C3-Cz_D1_STE-Del-2-Dim-3	2.02	0.0447
F4_D3_HP-comp	2.02	0.0451
T5_D3_HP-comp	2.02	0.0452
O1-F7_D2_SIE-Del-2-Dim-3	-2.02	0.0453
T5_D2_HP-mob	2.02	0.0453
P3-T5_D4_STE-Del-2-Dim-3	2.02	0.0454
C4-O2_D2_STE-Del-3-Dim-5	-2.01	0.0461
F4-T6_D2_SIE-Del-3-Dim-5	2.01	0.0465
Fp1-C4_A4_STE-Del-5-Dim-7	2.00	0.0465
F7-T6_D1_STE-Del-3-Dim-5	-2.00	0.0470
Fp1-O1_D4_STE-Del-3-Dim-5	-2.00	0.0472
Fp1-T6_D1_SIE-Del-2-Dim-3	-2.00	0.0472
P4-T6_D4_STE-Del-2-Dim-3	2.00	0.0474
P3_D1_SlopEn	2.00	0.0475

Continued on next page

Table A.3 – continued from previous page

Combination	t-stat	p-value
P4-O1_D1_STE-Del-2-Dim-3	2.00	0.0475
O2-Cz_D3_STE-Del-2-Dim-3	-2.00	0.0476
T4_D3_DispEn	1.99	0.0477
F3_D1_LZC	1.99	0.0478
C3-O2_D1_STE-Del-3-Dim-5	-1.99	0.0479
C4_D1_HP-mob	1.99	0.0480
T5_D2_PermEn	-1.99	0.0482
C3_D2_SpecEn	-1.99	0.0483
F4-T3_D4_STE-Del-2-Dim-3	-1.99	0.0484
F3-Cz_A4_SIE-Del-2-Dim-3	-1.99	0.0484
Fz_D2_DispEn	-1.98	0.0489
P4-T6_D2_SIE-Del-2-Dim-3	-1.98	0.0490
Fp2-T6_D2_STE-Del-3-Dim-5	-1.98	0.0492
Fz_D2_HP-comp	-1.98	0.0492
F8_all_SASI	-1.98	0.0493
O1_D3_HiguchiFD	-1.98	0.0495
F4-O1_D2_STE-Del-3-Dim-5	-1.98	0.0496
P3-O2_D3_STE-Del-3-Dim-5	-1.98	0.0496

A.3. Model selection results

In this section, we present additional plots that illustrate results related to the key objectives of the nested CV approach, specifically model selection and hyperparameter tuning.

A.3.1. Single measure setting

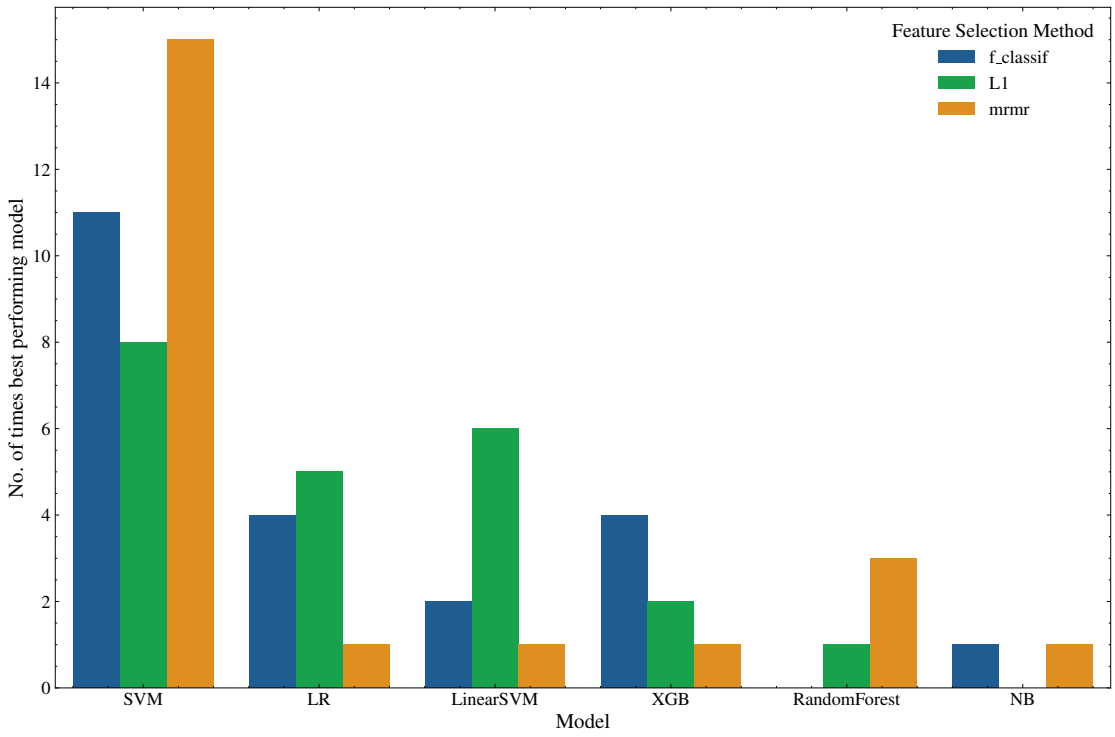


Figure A.1.: Number of times each model scored the best F1 score on the test fold for each feature selection technique for the single measure setting split by feature selection technique.

Table A.4.: Hyperparameters of the classification pipelines with the highest mean F1 scores in the single measure setting.

Measure	Feature Selection	k	Model	Model parameters
STE-Del-2-Dim-3	mRMR	40	SVM	{C=27.7, kernel=RBF}
HP-comp	mRMR	50	SVM	{C=32.3, kernel=RBF}
HiguchiFD	L1	49	SVM	{C=33.8, kernel=RBF}
DispEn	L1	25	SVM	{C=58.0, kernel=RBF}

A.3.2. Multi measure setting

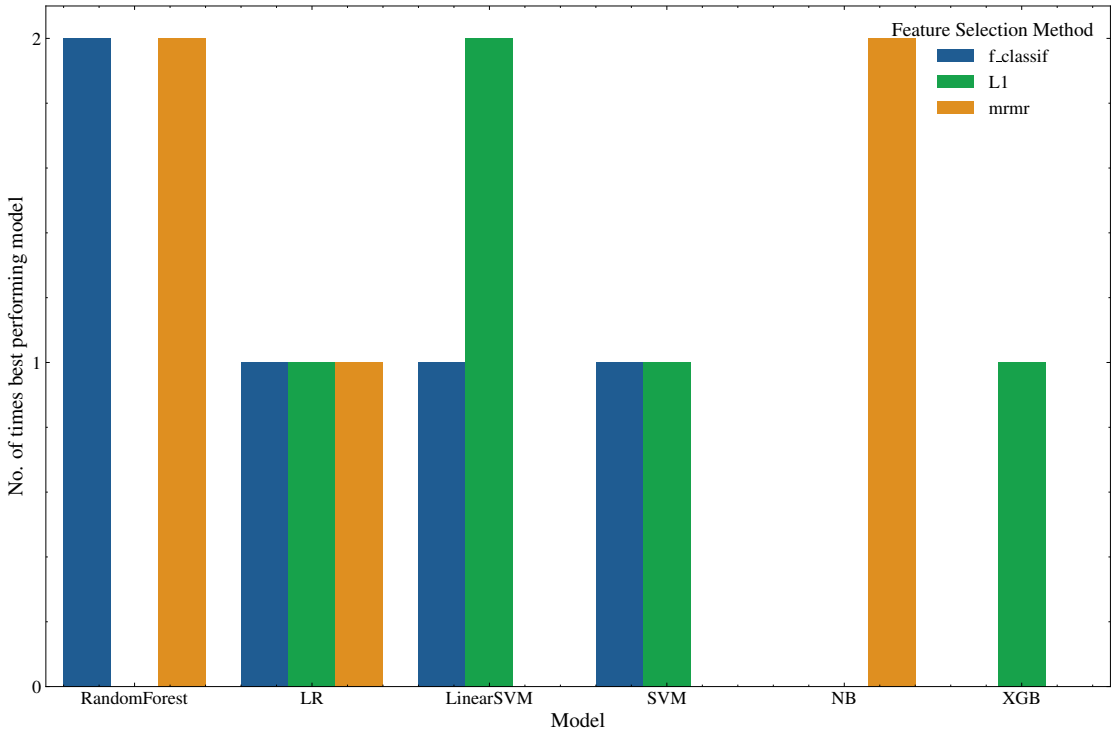


Figure A.2.: Number of times each model scored the best F1 score on the test set for the multi-measure setting split by feature selection technique.

Table A.5.: Hyperparameters of the classification pipelines with the highest mean F1 scores in the multi measure setting.

Feature Selection	k	Model	Model parameters
mRMR	5	RF	{max_depth=5, min_samples_split=22, n_estimators=410}
mRMR	45	SVM	{C=98.0, kernel=RBF}
F-Test	42	LR	{C=28.67, solver=liblinear}

A.4. Proof of Concept: Seizure Dataset

To demonstrate that the presented workflow is effective for EEG data, we provide results from a relatively simpler problem in EEG data analysis: seizure detection. For this purpose, we use the seizure dataset from [And+01], which comprises five sets of 100 single-channel EEG recordings. Sets A and B correspond to healthy individuals, sets C

and D are interictal, and set E is recorded during seizures. Our approach, outlined in Section 5.1, is applied to the raw data. However, since the data is single-channel, we cannot apply multi-channel measures such as STE or SIE. We evaluate two scenarios:

A Determining seizure episodes (set E vs sets A-D).

B Distinguishing seizure patients from healthy patients (training on sets A and C and testing on sets B and D).

The results are shown in Table A.6.

Table A.6.: Results from applying our workflow on the Bonn seizure dataset.

Setting	**Accuracy**	**F1-score**
A)	0.972	0.936
B)	0.96	0.959