Analysis of Electroencephalography in Epilepsy after Transcranial Brain Stimulation using Connectivity Models and Machine Learning

by

Αλεξάνδρα Τσιπουράκη - Alexandra Tsipouraki

A diploma thesis submitted to the
Department of Electrical and Computer Engineering
in conformity with the requirements for
the degree of Electrical and Computer Engineering

Thesis Committee:

Professor Zervakis Michail, Thesis Supervisor, TUC

Professor Liavas Athanasios, TUC

Professor Wolters H. Carsten, University of Münster

Technical University of Crete
Chania, Crete, Greece
March 2024

Abstract

Epilepsy constitutes a neurological disorder affecting approximately 50 million individuals globally, significantly impacting their quality of life. Conventionally, epilepsy symptoms are managed through the administration of antiepileptic drugs or, where feasible, through surgical intervention. However, the frequent cases of refractory focal epilepsy have highlighted the necessity for new, personalized therapeutic approaches. Among these, transcranial Direct Current Stimulation (tDCS) has emerged as a promising potential solution. The present work, focuses on the study of EEG recordings deriving from a proof-of-principle N-of-1 trial study, whose scope was to investigate the effects of multi-channel (mc-) tDCS application on a patient with refractory focal epilepsy. A double-blind sham-controlled stimulation experiment was conducted in a two-week long stimulation trial. Distributed Constrained Maximum Intensity (D-CMI)-based-mc-tDCS and sham stimulation were applied twice every week-day for 20 minutes each. EEG data, was recorded for 1 hour before and after stimulation. Experts, marked a highly significant reduction in interictal spike frequency after the stimulation process, while this was not the case for sham. Our purpose, is to evaluate EEG connectivity patterns, using generalized Partial Directed Coherence (gPDC) before and after stimulation and sham procedures accordingly. The raw EEG recordings are segmented into 3-second long sub-signals, to which then gPDC is applied and studied. We further proceed to the extraction of connectivity and statistical features from this analysis, and provide this information to Machine Learning models, in order to verify and validate our connectivity findings. The final results are promising; the connectivity analysis performed on the EEG data validated the results which had already derived from the trial, the epileptogenic zone was confirmed, as also was the reduction of IEDs after the tDCS. Finally, the ML models' results validated the robustness of our connectivity study, highlighted by the decrease in class separability after the stimulation process but not after sham. This research,

contributes to gaining a deeper understanding of the neural mechanisms underlying epilepsy, utilizing a non-invasive modality, as is EEG. The ability to gather and analyze more extensive EEG data sets over longer periods, can enhance in the future the depth and reliability of our findings, offering a richer understanding of the effects of Transcranial Direct Current Stimulation (tDCS) on epilepsy.

Περίληψη

Η επιληψία, αποτελεί μία από τις πιο κοινές νευρολογικές διαταραχές, πλήττοντας περίπου 50 εκατομμύρια άτομα παγκοσμίως και επηρεάζοντας σημαντικά την ποιότητα ζωής τους. Συμβατικά, τα συμπτώματα της επιληψίας διαχειρίζονται μέσω της χορήγησης αντιεπιληπτικών φαρμάκων ή, όπου είναι εφικτό, μέσω χειρουργικής επέμβασης. Ωστόσο, η συχνή εμφάνιση περιπτώσεων φαρμακο-ανθεκτικής, μη χειρουργήσιμης επιληψίας, μας ωθεί στην αναζήτηση νέων, εξατομικευμένων, μεθόδων θεραπείας. Μεταξύ αυτών, η άμεση Διακρανιακή Ηλεκτρική Διέγερση (tDCS) έχει αναδειχθεί ως πολλά υποσχόμενη λύση. Η παρούσα εργασία εστιάζει στη μελέτη εγκεφαλογραφημάτων (ΗΕΓ) που προέρχονται από μία N-of-1 μελέτη με στόχο την εξέταση της επίδρασης της εφαρμογής πολυκαναλικής tDCS (mc-tDCS) σε ασθενή με φαρμακοανθεκτική επιληψία. Σε αυτό το πλαίσιο, είχε διεξαχθεί ένα Distributed Constrained Maximum Intensity (D-CMI)-based-mc-tDCS and sham stimulation πείραμα συνολικής διάρκειας δύο εβδομάδων. Τόσο το πείραμα πραγματικής διέγερσης όσο και εικονικής εφαρμόστηκαν δύο φορές κάθε εργάσιμη ημέρα της εβδομάδας για 20 λεπτά η κάθε μία. Τα δεδομένα ΗΕΓ καταγράφηκαν για 1 ώρα πριν και μετά τη διέγερση. Οι ειδικοί, διαπίστωσαν μία ιδιαίτερα σημαντική μείωση της συχνότητας εμφάνισης επιληπτικών κρίσεων μετά τη διαδικασία διέγερσης, ενώ δεν παρατηρήθηκε κάτι αντίστοιχο μετά την εικονική διέγερση. Σκοπός μας, είναι να αξιολογήσουμε τα μοτίβα συνδεσιμότητας που προκύπτουν από τη μελέτη των καταγραφών ΗΕΓ, με την εφαρμογή της μεθόδου γενικευμένης Μερικής Κατευθυνόμενης Συνάφειας (gPDC) πριν και μετά την πραγματική και εικονική διέγερση αντίστοιχα. Οι αρχικές καταγραφές, διαιρέθηκαν σε υποσήματα διάρκειας 3 δευτερολέπτων, στα οποία μετέπειτα εφαρμόστηκε και μελετήθηκε η gPDC. Κατόπιν, τα χαρακτηριστικά που προέκυψαν από την προαναφερθείσα μελέτη συνδεσιμότητας καθώς και κάποια

στατιστικά μεγέθη, εισήχθησαν σε μοντέλα Μηχανικής Μάθησης, προκειμένου να επικυρωθούν τα προγενέστερα ευρήματά μας. Τα τελικά αποτελέσματα κρίνονται ενθαρρυντικά, καθώς η ανάλυση συνδεσιμότητας επικύρωσε τα αποτελέσματα που είχαν προκύψει από το αρχικό πείραμα, επιβεβαιώθηκε η επιληπτογόνος περιοχή, όπως και η μείωση των καταγεγραμμένων επιληπτικών κρίσεων μετά την πραγματική ηλεκτρική διέγερση. Τέλος, τα μοντέλα μηχανικής μάθησης επιβεβαίωσαν τη μελέτη συνδεσιμότητας μέσω της εμφανούς μείωσης ποσοστών επιτυχημένου διαχωρισμού των κλάσεων από τα μοντέλα πρίν και μετά την πραγματική διέγερση, ενώ δεν παρατηρήθηκε κάτι αντίστοιχο για την εικονική. Η παρούσα εργασία, επιχειρεί να συμβάλλει στην περαιτέρω διερεύνηση και κατανόηση των νευρικών μηχανισμών και συνδέσεων που υπόκεινται της επιληψίας, χρησιμοποιώντας μία μη επεμβατική μέθοδο καταγραφής δεδομένων. Η δυνατότητα συλλογής και ανάλυσης δεδομένων για μεγαλύτερες χρονικές περιόδους μπορεί να ενισχύσει μελλοντικά το βάθος και την αξιοπιστία των ευρημάτων της συγκεκριμένης ανάλυσης, προσφέροντας μία πιο πλούσια κατανόηση των επιδράσεων της διακρανιακής ηλεκτρικής διέγερσης στην επιληψία.

Acknowledgements

This diploma thesis marks for me the end of a wonderful journey of growth at the Technical University of Crete, both academically and personally. In this institution, I took my first steps as an aspiring engineer, learning valuable lessons along the way which I will carry with me and cherish for life. I met wonderful people, without which this journey wouldn't have been so memorable nor so exciting and, of course, I had the luck to have a truly strong support system from start to finish.

First and foremost, I would like to thank my supervisor, Professor Michail Zervakis, who saw the passion I had for biomedical engineering and supported and guided me along the way, showing true interest into forming the next generation of engineers. Of course, through Prof. Zervakis, I had the honor of meeting and collaborating with Dr. Marios Antonakakis, who I'd like to thank for his valuable guidance, his kind replies to my millions of questions and his push to always aim for the best possible result. I would like to also extend my sincere gratitude to Professor Carsten H. Wolters, who kindly provided his lab's data and his guidance as a member of my thesis committee, and also, to Professor Athanasios Liavas for his valuable evaluation on the present work. Furthermore, I would like to thank my friends at the DISPLAY lab for proofreading this thesis.

Most importantly, I would like to thank my family, my parents Stefanos and Silvia and my brother Alexios, for always being there for me and supporting me constantly, in any way humanly possible.

Finally, I'd like to thank Leonidas and all of my friends, for cheering for me at my best and encouraging me through the difficult moments.

Contents

List of Figures

Chapter 1

Introduction

The human brain, an intricate organ both in structure and functionality, plays a central role in physiological and psychological processes. However, its complexity makes it vulnerable to various disorders that can heavily impact the lives of those affected. Epilepsy, constitutes a neurological brain disorder present in more than 50 million people worldwide, according to the World Health Organization. Characterized by recurrent seizures, this disorder can lead to a variety of symptoms ranging in severity and impact on the patient's life. Therefore, the field of epilepsy research and treatment is constantly evolving, driven by the necessity to address the consequences of this disorder's manifestations. Currently, the most commonly used treatment methods for epilepsy are either anti-epileptic drug treatments or even surgery. However, given that these approaches are not suitable for all patients, there is a pressing need for alternative, personalized, treatment options. Transcranial Direct Current Stimulation (tDCS), emerges as a promising treatment method for the reduction of seizure frequency, especially for patients considered unresponsive to drug treatment and not eligible for resection. In this work, the outcomes of a double-blind sham-controlled N-of-1 clinical trial on a patient with refractory focal epilepsy which has been treated with personalized and dCMI optimized tDCS as well as with active sham (ActiSham) will be studied [1], focusing on the EEG recordings produced and the effective connectivity perspective. These results, will then be validated and highlighted by Machine Learning models.

1.1 Contributions and Innovations of this Thesis

This work aims to present a comprehensive framework developed in Python, designed to transform raw EEG data into insightful findings on the brain's effective connectivity patterns in cases of refractory focal epilepsy. At the core of this, is a series of steps, beginning with preprocessing, advancing through connectivity analysis, feature extraction and finally the application of Machine Learning algorithms for binary classification of seizure or non-seizure events. The main purpose of this process is perform and validate our analysis utilizing solely the EEG recordings from the trial. The choice of EEG, a non-invasive recording technique, enables the frequent and safe conduct of experiments on patients. Thereby, through this approach we aim to contribute to the field of epilepsy research by demonstrating the effectiveness of utilizing EEG to enhance our understanding on the treatment of epilepsy, and more specifically, on the effects of tDCS in the reduction of epileptic seizures, and study the brain's functionality in general.

1.2 Thesis Outline

In Chapter 2, the reader is introduced to the human brain; it's anatomy, functionality and the concept of neural circuits. The discussion extends to the presentation of the disorders of the human brain, with a special focus on epilepsy. Definitions, terminology, and the current state of epilepsy research, including advanced imaging and signal processing techniques are covered, to establish a solid theoretical background for the study. Finally, an introduction to brain connectivity measures in epilepsy research along with the concept of gPDC are presented, as well as an introduction to the ML algorithms ulitized in our pipeline. In Chapter3, the initial study design on which this work is based on is introduced and explained in depth, along with how its methods and findings will be exploited in our work. Furthermore, this thesis' technical design will be explained in detail. Chapter 4 will include all the findings from the connectivity analysis, demonstrating the impact of tDCS on the patient's neural network through effective connectivity. It also showcases the results obtained from applying machine learning models to the EEG data, offering insights into the predictive power of the extracted features in distinguishing seizure and non-seizure states. Finally, Chapter 5 synthesizes the conclusions deriving from

this study, suggesting also some future work ideas that could further improve our results.

Chapter 2

Exploring the Brain: Structure, Function, and the Path to Epilepsy Solutions

2.1 The Human Brain

The human brain, a pivotal organ in the Central Nervous System (CNS), is a fascinating proof of biological complexity and versatility, as it plays a central role in the orchestration of both physiological and psychological processes. Structurally, it is composed of three main parts; the cerebrum, the cerebrellum and the brainstem [2] (Figure 2.1). As for the brain's functionality, it depends on neurons, the specialized cells responsible for transmitting and receiving nerve impulses. These cells communicate through electrical and chemical signals forming a complex network that underlies all brain activities including thought, emotion and motor controls. Neuronal activity is fundamental to all aspects of brain function, from basic reflexes to complex cognitive processes (Figure 2.2) [3].

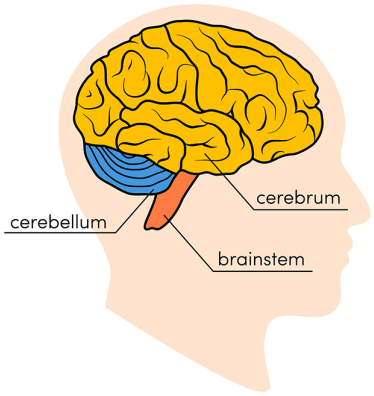


Figure 2.1: The human brain [4]

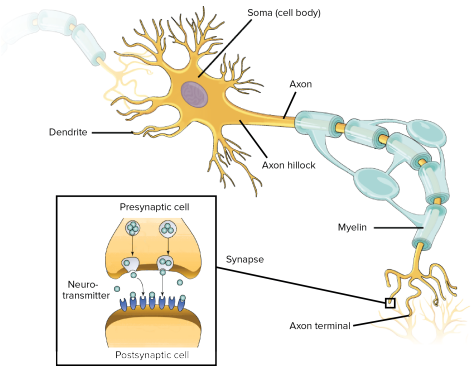


Figure 2.2: Neurons and neuronal activity [5]

2.2 Anatomy/ Gross Anatomy

Let's discuss the human brain in more detail, in order to fully understand its form and functionality.

Physical Structure of the Brain

Weighing about 3 pounds in the average adult, the brain is composed of about 60% fat. The remaining 40% is a combination of water, protein, carbohydrates and salts. The brain itself is not a muscle. It contains blood vessels and nerves, including neurons and glial cells[6]. Beyond its basic composition, the cerebral cortex, which envelopes the outer layer of the cerebrum, varies in thickness and shows a complex arrangement across different brain regions. This variation is indicative of the specialized functions housed in each area. Beneath the cortical layers critical subcortical structures can be found, like the basal ganglia, essential for movement coordination, and the thalamus, which acts as a sensory and motor signal relay station.

The brainstem, which contains the medulla, pons, and midbrain, is the brain part which is responsible for maintaining vital autonomic functions, such as heart rate and respiration.

Adjacent to the brainstem, is located the cerebrellum. Although primarily recognized for its role in motor control, it also has surprising contributions to cognitive functions, including attention and language processing.

Finally, the brain's white matter, comprising nerve fibers covered by myelin, forms an extensive network connecting diverse brain regions, facilitating a robust inter-regional communication. The integrity of the white matter tracts is vital for the harmonious functioning of the brain's various components. These intricate structural details of the brain aim to highlight not only its complexity but also the importance of each component in ensuring the organ's overall functionality [7].

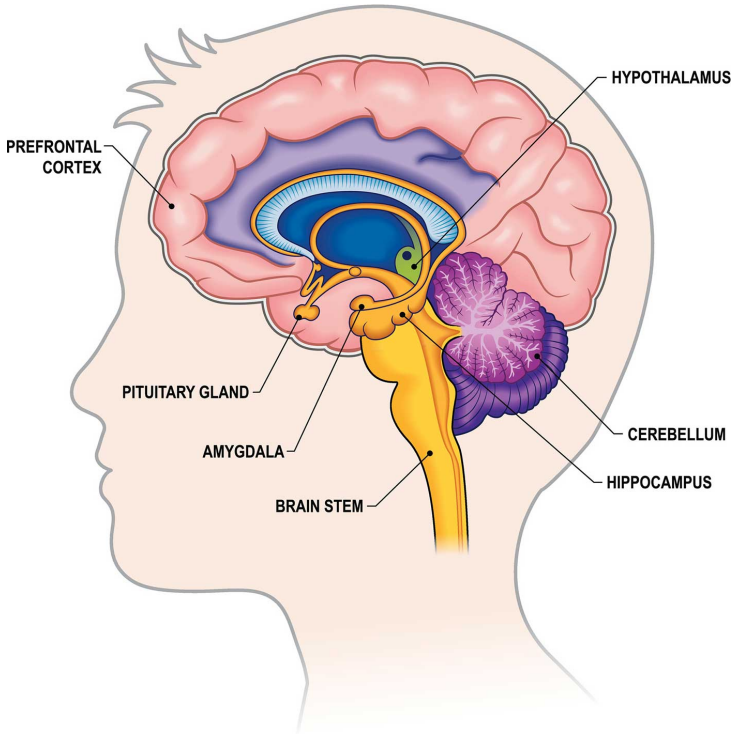


Figure 2.3: A deeper view into the brain's main structure [4]

2.3 Basic Functional Systems of the Brain

Proceeding to the functional systems of the brain, let's discuss about neurons and neural circuits.

2.3.1 Neural circuits and Communication

Neurons never function in isolation; they are organized into groups, namely neural circuits, that process specific kinds of information, forming the basis of neural communication in the brain. These circuits play a crucial role in everything, from basic reflex arcs to advanced functions like memory, learning and decision-making. The mechanisms hidden behind neural communication can be primarily categorized into electrical and chemical signaling. Electrical signals, predominantly in the form of action potentials, are brief electrical impulses that travel along neurons.

Complementing the electrical signaling process is chemical signaling, which occurs at synapses, the junctions between neurons. In the synaptic transmission, neurotransmitters are released from the presynaptic neuron and bind to specific receptors in the postsynaptic neuron. This chemical process is essential for the continuation of the neural signal across the synaptic

gap, allowing for more complex forms of neural processing and communication. This complex interaction between electrochemical signals ensures the efficient and dynamic functioning of the brain's neural circuits.

2.3.2 Major Brain Regions and their Functions

When discussing the brain, it's of great importance to theoretically dissect it and explain it's anatomy. The brain can be separated into two hemispheres; the left and right one, which are parts of the cerebrum. Each hemisphere has four sections, called lobes, namely: frontal, parietal, temporal and occipital, as seen in Figure 2.4. To explain these structures in more detail, we shall start by analyzing the functions of the frontal lobe.

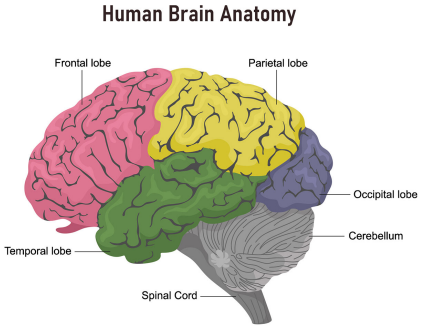


Figure 2.4: The brain dissected to lobes [4]

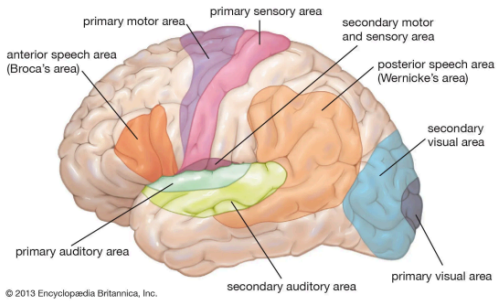


Figure 2.5: Broca's area and more

Frontal Lobe

The frontal is the largest lobe of the brain. Located in the front part of the head, it's crucial for decision-making, voluntary movement, emotional control and problem-solving. It also contains Broca's area (Figure 2.5), which is associated with speech production. Another part of this lobe, the prefrontal cortex, is particularly significant for its role in personality expression and complex cognitive processes [8].

Parietal Lobe

Situated right behind the frontal lobe, the parietal lobe is primarily responsible for processing and integrating sensory information coming from various parts of the body as it is involved in

interpreting pain and touch. It is also essential for understanding spatial navigation and orientation and plays a part in coordinating movement and spatial reasoning. Moreover, this section of the brain also houses Wernicke's area (Figure 2.5), which helps with understanding spoken language.

Occipital Lobe

Further back in the brain is located the occipital lobe, which is the center for visual processing. It is involved in interpreting all visual information coming from the eyes, including the recognition of shapes, colors and motion.

Temporal Lobe

Finally, the temporal lobes, located on the sides of the brain and beneath the lateral fissure, are key areas for processing auditory information and are critically involved in the comprehension of language and speech. Temporal lobes, are also involved in short-term memory, musical rhythm and some degree of smell recognition.

Limbic System

Besides the lobes, another complex part of the brain worth being mentioned is the limbic system. More precisely, the limbic system is composed of a system of nerves and networks and is involved in controlling mainly emotional states and memory, more specifically, it plays a crucial role in emotional responses, memory formation (mostly in the formation of episodic and declarative memories) and learning process. Its key components are the hippocampus, the amygdala and the hypothalamus [9].

The intricate functionality of all the aforementioned brain regions highlights the significance of their integrity for a person to be considered as overall healthy. Damage or malfunction in any of these areas can lead to a wide range of health problems varying in severity. Understanding the potential consequences that can be caused from damage in each specific brain area is of great importance for diagnosing and treating both brain damage and diseases of all kinds.

The various implications that can be caused, highlight the importance of personalized treatment approaches in neurological care. Each brain disorder, whether it's an injury or any kind of disease, as is epilepsy, requires a tailored approach considering the specific region affected and the unique symptoms/manifestations that can appear for each patient. Effective treatment must not only address the symptoms but also consider underlying causes and the patient's overall health, emphasizing in a holistic patient-centered approach. Last but not least, as will be thoroughly analyzed later on, it has been shown that each patient responds differently to each treatment method [10], hence, focusing on personalized treatment is more and more considered to be very beneficial.

2.3.3 Disorders of the Human Brain

The human brain, being a complex and intricate organ, is susceptible to a variety of disorders that can impact its function and overall health. These disorders can range from congenital conditions to acquired diseases, each affecting the brain in different ways. Brain disorders vary both in etiology and manifestations and have been categorized based on their specific characteristics. First, there exist neurodegenerative disorders [11], including disorders like Alzheimer's and Parkinson's disease, characterized by the progressive degeneration of brain cells. They often lead to symptoms like memory loss, cognitive decline and impaired motor functions. Another category is brain tumors. These can be sub categorized as benign or malignant and they affect brain function by creating pressure or disrupting normal neural pathways. Furthermore, another significant category is Traumatic Brain Injuries (TBI's); resulting from external physical trauma to the head, TBI's can lead to various symptoms, ranging from mild ones such as concussions to even severe brain damage. Strokes are also considered a separate brain disorder category, happening when the blood supply to the brain is interrupted, leading to cell death. They can result in short-term or long-term disabilities, ranging from speech impairment to paralysis and other disabilities. Mental Health Disorders, such as depression, anxiety, bipolar disorder and schizophrenia also constitute another, newly highlighted, category of brain disorders. They primarily affect mood, thinking and behavior, often requiring long-term management. A similar category is developmental disorders in which mainly lie autism spectrum disorder and attention

deficit hyperactivity disorder (ADHD), which are often identified in early childhood, affecting communication, behavior and learning. Finally, a large category of brain disorders with various different symptoms is epilepsy. Epilepsy, is a significant brain disorder characterized mainly by recurrent and unprovoked seizures. This specific disorder affects around 50 million people worldwide, according to the World Health Organization, making it one of the most common neurological diseases globally. Focusing on the purpose and content of this thesis, in which the subject of epilepsy will be tackled, an extended definition of the disorder should be given, along with its causes, consequences and potential treatments.

2.4 Epilepsy

2.4.1 Definitions and Terminology

Throughout its recorded history of at least 3 to 4 millennia, epilepsy has been described through various terms across cultures, with the term 'seizure' currently preferred in medical classification. Even though the characteristics of epileptic attacks vary, whatever word is chosen to define one, there are always similar or borderline non-epileptic attacks to confuse the physician in the differential diagnosis [12]. In his 2012 work, Reynolds explains that the word 'epilepsy' is of Greek origin and means to seize, to take hold of or to attack. The word 'seizure ' is Latin, from 'sacire' i.e. to claim. These words reflect the ancient belief that the sufferer has been seized or claimed by a supernatural power, spirit or god.

Epilepsy in modern medicine can be defined as a chronic neurological disorder characterized by the enduring predisposition to generate epileptic seizures. These seizures are transient occurrences of signs and/or symptoms caused by abnormal excessive or synchronous neuronal activity in the brain. The clinical manifestations of seizures vary, depending on the seizure type and brain region affected by it. In order to be diagnosed, a case of epilepsy has to comprise of at least two unprovoked (or reflex) seizures occurring more than twenty-four hours apart, or after one seizure with a high probability of further seizures that are similar to the general recurrence risk after two unprovoked seizures [13]. From a neurophysiological point of view, epilepsy involves disruptions in the brain's network connectivity, leading to periodic and unpredictable seizure events. This disruption can be caused by many factors, including genetic

predispositions, brain injuries and other brain diseases or even developmental disorders. The study of these network disruptions, particularly through advanced imaging and signal processing techniques, is crucial for understanding the pathophysiology of epilepsy, but also for developing targeted treatment strategies, since every case of epilepsy and hence, every patient, needs to be studied and treated differently due to the variety of factors which can provoke seizures and the different possible brain areas that are affected by them.

Seizure Classification

The seizure classification process is constantly being reevaluated. Initially, classifications were mainly based on all observable symptoms, a practice dating back to ancient times when seizures were often attributed to supernatural phenomena [14]. In the 20th century a pivotal shift was marked, with the advent of electroencephalography (EEG), introduced by Hans Berger in 1929. This innovation revolutionized the understanding of seizures, enabling clinicians to correlate clinical manifestations with electrical activities in the brain [15]. EEG primarily records the electrical activity of the brain with the use of electrodes placed on the scalp. It's considered a non-invasive imaging method with high temporal and low spatial resolution, meaning it provides information on the changes of neural activity over time, facilitating the detection of brain disorders but is usually unable to discover their onset. The integration of EEG findings into seizure classification facilitated a more precise and scientific approach, paving the way for modern classification systems [16]. Today, these historical foundations continue being the basis of classification methodologies, underscoring their critical role in accurate diagnosis and effective treatment planning for epilepsy [17].

The International League Against Epilepsy (ILAE) plays a crucial role in epilepsy research and treatment since its foundation in 1909. The ILAE classification system, is praised and valued by professionals for its methodical approach to seizure categorization based on multiple factors including origin, clinical presentation and EEG findings. Significant changes were made during the 2017 revision, reflecting the advances in the field of neurology and neuroscience, ensuring the system's relevance and efficacy [18]. In the current ILAE system, based on all the aforementioned criteria, the major categories of epilepsy include focal and generalized

seizures. The system also categorizes separately seizures with unknown onset, where the origin of the disorder is unclear, highlighting once again the necessity of refinement in epilepsy diagnostics [17]. Having already established the classification system and the various categories of epilepsy based on it, let us now give a more detailed explanation per category.

- **Focal Seizures**

Focal seizures originate from one specific brain area but can become generalized and spread to other areas. They can be characterized as simple (auras) or complex focal seizures [6]. In some medical journals, these categories are also described as focal aware or focal impaired awareness seizures, based on whether the individual remains conscious or not. Based on the seizure description and the patient's symptoms, physicians may be able to identify the specific part of the brain causing the seizures. The diagnosis, typically involves EEG recordings to identify the brain area affected or even Video-electroencephalography monitoring (VEEG) and other imaging techniques like MRI to detect any brain abnormalities.

- **Generalized Seizures**

Generalized seizures occur when the abnormal electrical activity causing the seizures is generated from both hemispheres of the brain and is not localized in a specific part of it. These usually include absence seizures, characterized by brief awareness loss but also atonic, tonic, clonic, tonic-clonic, myoclonic and febrile seizures. These seizures may result in spasms of the body, stiffening, shaking etc. Diagnosis also begins with tests like EEG recordings, MRI and blood tests too, since these tools can reveal characteristic patterns that differenciate them from focal seizures.

- **Unknown Seizures**

In some cases, the onset of the seizure can't be localized. These seizures are described as unknown. Ongoing research is vital to try to better understand, describe and try to provide more targeted treatment methods for these particular cases.

2.4.2 Epilepsy and Brain Connectivity

Epilepsy is being increasingly recognized as a brain network disorder, as seizures and their effects can't be considered confined to isolated brain regions, but rather involve complex brain networks. In fact, according to Kramer and Cash's work [19], seizures are classically viewed as a hyper-synchronous state, indicative of excessive connectivity or communication among brain regions and neurons. Many researchers used this hypothesis as their baseline, proving it to be quite valid. In fact, as it can be imagined, during seizures, the produced activity can cause temporary but dramatic changes in the brain networks, which are the ones visible through EEG patterns. The rather unusual fact lies in the interictal state, which is the state between seizures, where apparently the brain network in people with epilepsy shows altered connectivity [20], which may also contribute to cognitive and behavioral change. Building on this, it's being considered that ictogenic regions, or otherwise, areas initiating seizures, might demonstrate enhanced correlations with other brain areas, a phenomenon called hypersynchrony. Diverse studies employing various synchrony measures in interictal EEG have shown some success in identifying epileptogenic cortex [21]. However, there's still much to learn about which connectivity measures are most effective in finding ictogenic regions and how these findings can be integrated into clinical evaluations.

2.4.3 Etiology, Diagnosis and Treatments of Epilepsy

The etiologies causing epilepsy contain a variety of factors, ranging from genetic ones, to structural changes in the brain provoked by some injury or disease, immune and metabolic disorders or even developmental conditions. However, in many cases the exact cause for epilepsy remains unknown. As it has been previously mentioned, the diagnosis process involves typically a combination of patient history, neurological examination and also diagnostic tests like EEG, MRI and other brain imaging tests. But for now, let's focus on potential treatments.

The treatment of epilepsy primarily aims at controlling seizures by reducing their severity and their frequency in order to improve the patient's quality of life. Several treatment methods are being used and new ones are constantly being studied and tested. The first-line treatment of epilepsy usually involves medication, most commonly antiepileptic drugs (AEDs).

The choice of medication depends on the patient's diagnosis. When medications fail to control seizures after even a second AED trial, the patient is categorized as having drug resistent epilepsy (DRE), according to the ILAE, or otherwise named refractory epilepsy, and surgery is considered. Procedures like resective surgery, which involves removing the area where seizures originate, are considered effective but are also quite invasive. Another surgical intervention is the implantation of a vagus nerve stimulator (VNS), which sends regular mild pulses of electrical energy to the brain via the vagus nerve. Also, the ketogenic diet, meaning a dietary therapy treatment, has been proved effective in specific cases. Finally, ongoing research is exploring gene therapy, stem cell therapy and new drug treatments. It is worth mentioning once more, that as the years pass and research on epilepsy dives deeper into this disorder, the need for patient-centered therapy seems to be increasingly critical. Tailored treatment approaches that consider the individual's specific type and severity of seizures, lifestyle, and other health conditions are becoming more important for effective management and improved quality of life. This shift towards personalized medicine in epilepsy care reflects of a broader shift in healthcare, which is now focusing on studying the unique aspects of each patient's condition and response to treatment.

2.5 Advanced Imaging and Signal Processing in Epilepsy Research

The field of epilepsy research has been revolutionized by the discovery and utilization of advanced imaging and signal processing techniques. These technologies not only offer valuable insights into the human brain's complex neural mechanisms which cause epileptic seizures, but they also highlight their effects. Advanced imaging, which includes methods like magnetic resonance imaging (MRI), functional MRI (fMRI) and positron emission tomography (PET), allows for detailed visualization of the brain's structure and function, enabling researchers to pinpoint epileptic zones and deepen their understanding of the disorder. A particularly groundbreaking discovery in this domain was electroencephalography (EEG). This tool stands out for its ability to capture electrical brain activity with high temporal resolution and, this particular fact, makes it invaluable for monitoring neural activity both during seizures and in interictal periods [22]. The application of signal processing techniques to EEG data, such as performing connectiv-

ity analysis and applying machine learning models, methods which were implemented in the present work and will be discussed later on, has furthered our ability to understand and interpret neural signals, aiding the accurate diagnosis and effective treatment of epilepsy but also paving the way for potential seizure prediction.

Furthermore, the emergence of Transcranial Brain Stimulation (TBS) as a treatment modality has offered a new avenue for managing epilepsy. This therapy method, which involves non-invasive stimulation of the brain, has shown promise in modifying neural activity and potentially reducing seizure frequency. Its combination with EEG monitoring allows for a precise assessment of the treatment's impact on brain activity, further enhancing our understanding of epilepsy and opening new possibilities for personalized therapeutic strategies per patient. A subcategory of TBS worth mentioning due to its innovative approach in the epilepsy treatment field is transcranial Direct Current Stimulation (tDCS). This therapeutic approach has gained attention in the epilepsy research, particularly for its potential in managing refractory focal epilepsy [23] . Being a non-invasive neuromodulation technique, it has been explored in various studies over the past few decades which focused on reducing seizure frequency and EEG epileptiform discharges in epilepsy patients. The majority of these studies involved applying cathodal tDCS, targeted at areas showing maximal EEG abnormalities. Results from these studies are promising: a significant portion reported a reduction in seizure frequency and a decline in EEG epileptiform discharge rates, with no serious adverse events presented [24] [25]. However, it's important to note that these studies vary in their methodology and sample populations. There is a consensus on the need for more extensive, sham-controlled randomized trials with well-informed stimulation protocols to further understand and establish tDCS's role in epilepsy management [26]. From this necessity, stemmed the trial on which the present work's results are based on, which will be discussed in detail in Chapter 3.

2.6 Theoretical Background

In order to proceed with a deeper explanation of the methodology followed in the present thesis, it is of great importance to state the fundamentals and lay the basis on which this work was built on. Let's begin with stating once again the purpose, and connecting it with all the in-

formation stated above about epilepsy. Epilepsy stands as a severe neurological condition that significantly impacts the quality of life for those affected. For patients with refractory epilepsy, the use of alternative treatments is crucial. Transcranial Direct Current Stimulation (tDCS) presents as a promising avenue for such cases, offering a non-invasive, innovative approach to managing epilepsy. In the present work, whose pipeline will be thoroughly explained in the following chapter, our purpose is to study the effects of an N-of-1 trial on a patient with refractory focal epilepsy, focusing on the brain connectivity patterns and their changes throughout the trial process, using dynamic connectivity measures like generalized Partial Directed Coherence (gPDC), and then highlight the relevance of the connectivity measures with the help of Machine Learning classifiers.

2.6.1 Brain Connectivity Measures in Epilepsy Research

Introduction to Brain Connectivity

As has been known since the nineteenth century, the neuronal elements of the brain form a complicated structural network. As Bullmore and Sporns highlighted in their 2009 study [27], it has been increasingly recognized since the twentieth century that the anatomical foundation of the brain facilitates the dynamic development of coherent physiological activities, like phase-locked high-frequency electromagnetic oscillations. These activities extend across various spatially separate brain regions, forming a functional network. Brain connectivity can be subdivided into neuroanatomical (or structural), functional, and effective connectivity [28]. Neuroanatomical connectivity refers to structural links such as synapses or fiber pathways at the microscopic scale of neurons [29]. This connectivity type is often mapped with the use of techniques such as diffusion tensor imaging (DTI), with which the pathways of white matter tracts in the brain can be visualised. These structural pathways form the anatomical framework for neuronal communication. Functional connectivity describes the statistical dependencies and temporal correlations between spatially separated neuronal units or brain regions. It is concerned with the patterns of co-activation and synchronization between different brain areas, even if these areas are not directly structurally connected. This connectivity type, is frequently measured through non-invasive imaging techniques like functional magnetic resonance imaging (fMRI) or elec-

troencephalography (EEG), in order to reveal dynamic patterns of interaction while the subject is either resting or performing a task. Finally, effective connectivity goes beyond describing the patterns of correlation or association, as in functional connectivity, to infer directional influence and causal interactions between brain regions. It aims to understand how neuronal systems are correlated and examines the mechanism of these interactions. Effective connectivity is assessed through models, estimating the influence that one neural unit exerts over another, accounting for the direction and causal effect of these interactions. Techniques such as Granger causality analysis or dynamic causal modeling (DCM) are often employed to explore effective connectivity, providing insights into the operational architecture of the brain's functional networks.

In summary, neuroanatomical (structural) connectivity maps the brain's physical wiring, functional connectivity reveals temporal correlation patterns between brain regions, and effective connectivity offers insights into the directional and causal relationships underlying these correlations. Together, these connectivity categories provide a comprehensive framework for understanding the complex interplay between the brain's structure and function, as well as the dynamic processes that underlie human cognition and behavior.

In this thesis, we will mainly focus on effective connectivity and specifically analyze the EEG data collected with the generalized Partial Directed Coherence (gPDC) method.

The importance of understanding effective connectivity in the field of neuroscience is paramount, especially when dealing with complex neurological disorders such as epilepsy. This concept, not only illuminates the patterns of communication between different brain regions but also sheds light on both the directionality and influence these regions have over one another. Furthermore, effective connectivity analysis can pinpoint the specific brain regions initiating epileptic activity, namely the epileptogenic zones. Furthermore, by understanding the directional flow of neural information, clinicians and researchers can identify the source of seizures more accurately, which is crucial for both surgical planning and other targeted interventions. Also, as the patterns revealed by effective connectivity analysis may vary between different types of epilepsy and seizure manifestations, studying them helps in classifying seizure types more precisely, while at the same time contributing to a more personalized approach to both diagnosis and treatment. In particular, monitoring changes in effective connectivity pat-

terns over time can enable clinicians to assess how well a treatment is performing on a patient and make any adjustments to the procedure accordingly. This is particularly valuable in evaluating the outcomes of surgical interventions or ongoing drug therapy and definitely facilitates the process of creating targeted, personalised therapies. Last but not least, effective connectivity analysis benefits from integration with other neuroimaging techniques, such as structural MRI or PET scans, to provide a more comprehensive view of the brain's architecture and function. In summary, the analysis of effective connectivity is of critical importance in understanding, diagnosing, and treating epilepsy, as it offers a window into the dynamic and complex interactions within the brain, enabling both clinicians and scientists to identify epileptogenic zones with better accuracy and follow tailored treatments while gaining deeper insights into the mechanisms underlying epilepsy. As research and technology in this area continue to evolve, the potential to improve patient care and outcomes through effective connectivity analysis becomes increasingly significant.

The concept of Partial Directed Coherence and generalized Partial Directed Coherence

Partial Directed Coherence (PDC) is a sophisticated frequency-domain measure deriving from the concept of Granger causality, designed to explore directional interactions between multiple time series such as EEG signals. It provides insights into the directional flow of information between different brain regions. Building on this foundation, PDC was introduced as a method to precisely characterize the directed linear relationships between pairs of time series, such as $x_i(n)$ and $x_j(n)$, while considering their interactions within a larger network of time series. The concept of PDC was driven by its ability to reveal key elements of functional connectivity, especially in the field of neuroscience. The significance, stems from the critical function of neural rhythms (namely alpha (α) , beta (β), gamma (γ) etc), in understanding physiologic relevance. Conceptually, PDC is a generalization to the case of multiple time-series of Saito and Harashima's 'directed coherence' (DC). DC as an approach, provides insights into the functional linkage between two studied entities by indicating if and how they are connected. Unlike traditional coherence, that mainly considers the entities and their concurrent activities, DC highlights the nature of their interaction by categorizing it into 'feedforward' and 'feedback'

mechanisms. This approach was particularly important at the time it was first proposed, because much of our understanding of structural and functional neural connections was based on post-mortem anatomical studies of experimental animals, but studies did not illuminate whether the connections between structures were active during specific brain processes that lead to certain behaviors. Then, a novel approach, namely PDC, was first introduced by Luiz A. Baccalá and Koichi Sameshima in their year 2000 paper 'Partial directed coherence: a new concept in neural structure determination' [30], where the writers outlined the mathematical formulation of PDC and distinguished it from existing coherence measures while also demonstrating its utility in identifying directed neural connectivity patterns.

Firstly, the concept of Granger causality should be introduced, which is a statistical concept that helps determine whether one time series can predict another. It's named after Clive Granger, who developed the idea that if one time series can predict another, then it can be said to "Granger-cause" the other. The traditional Granger causality test relies on linear prediction models and looks at whether including past values of one time series can improve the prediction of another series. The writers of the PDC paper introduce a new approach to Granger causality, which is based on frequency-domain analysis rather than time-domain analysis, to now study how the different frequencies within the signals relate to each other, rather than just their values at different times. So the idea is to understand the directed interactions between signals, seeing how one signal's frequency components can predict those of another signal. To introduce PDC correctly, the Direct Transfer Function (DTF) should also be determined.

DTF is a measure that examines how well past values of one time series can predict current values of another. Mathematically, it's represented by the ratio of the cross-spectrum of two signals to the autospectrum of the signal being predicted, after considering the influences of all other signals in the system.

$$\mathrm{DTF}_{ij}(f) = \frac{H_{ij}(f)}{\sqrt{\sum_k \left|H_{kj}(f)\right|^2}} \tag{2.1}$$

PDC can be considered as a normalized version of DTF. It is computed by taking the elements of a matrix A(f), which contains coefficients that represent the linear interaction

between signals at frequency f, and dividing by the sum of the squares of the coefficients associated with the predicting signal. This results in a measure that reflects the relative influence of one signal on another, after factoring out the effects of all other signals.

Mathematically, the writers explained this as:

Definition 1: The partial directed coherence factor (PDCF) from j to i is given by

$$\pi_{ij}(f) = \frac{A_{ij}(f)}{\sqrt{A_{jj}(f)\,\Sigma^{-1}\,a_{ij}(f)}} \tag{2.2}$$

where $A_{ij}(f)$ is the i, jth element of $A(f)$.

It follows immediately that the partial coherence between i and j is given by

$$\kappa_{ij}(f) = \pi_{ih}(f) \sum^{-1} \pi_{ij}(f) \tag{2.3}$$

for $\pi_{ij}(f) = \left[\pi_{1j}(f), \quad \ldots, \quad \pi_{Nj}(f)\right]^T$, whence the motivation for $\pi_{ij}(f)$'s name.

Because

$$A_{ij}(f) = \begin{cases} 1 - \sum_{r=1}^{p} a_{ij}(r)e^{-2\pi f r}, & \text{if } i = j \\ -\sum_{r=1}^{p} a_{ij}(r)e^{-2\pi f r}, & \text{otherwise} \end{cases} \tag{2.4}$$

Definition 2: The PDC form j to i is given by

$$\pi_{ij}(f) = \frac{\hat{A}_{ij}(f)}{\sqrt{\hat{a}_{jj}^H(f)\hat{a}_{ij}(f)}} \tag{2.5}$$

where also hold these normalization properties:

$$0 \leq |\pi_{ij}(f)|^2 \leq 1 \tag{2.6}$$

$$\sum_{l=1}^{N} |\pi_{lj}(f)|^2 = 1, \quad \text{for all } 1 \leq j \leq N. \tag{2.7}$$

By establishing a clear method for calculating PDC and illustrating its application through empirical examples, the paper aims to provide neuroscientists with a powerful tool for uncovering the dynamic structural organization of neural networks, thereby contributing

to a deeper understanding of complex brain functions and disorders. PDC offers a method to quantify the directional flow of information between brain regions, which is crucial for identifying the pathways along which epileptic activity spreads during a seizure. By applying PDC to EEG data, researchers can potentially map out the seizure propagation patterns, identifying both the sources and targets of epileptic discharges. Given that different frequency bands may play distinct roles in the generation and propagation of epileptic seizures, the frequency-domain approach of PDC allows for the analysis of directional interactions within specific frequency bands. This specificity can help in identifying frequency-dependent patterns of neural connectivity that are relevant to the initiation and spread of seizures. In 2007, Baccalá and Sameshima decide to revisit and improve the PDC concept, introducing the concept of generalized Partial Directed Coherence (gPDC). gPDC improves upon the original PDC, as it is a refined method in the frequency domain that quantifies the direct influence of one time series on another within a multivariate dataset. This generalization makes gPDC a more versatile tool for analyzing complex neural systems, as it can accommodate a wider range of signal relationships and is not limited to specific model assumptions inherent to the original PDC formulation. The novelty of this method is that it takes into account the statistical properties of the estimated spectral matrix, making it more robust and consistent in various scenarios, especially when dealing with datasets that have a complex structure or when the data suffer from certain types of contamination or noise. The new partial directed coherence estimator as presented by the original paper of Baccalá, Sameshima and Takahashi is defined as:

$$\pi_{ij}^{(w)}(f) = \frac{\frac{1}{\sigma_i}\hat{A}_{ij}(f)}{\sqrt{\sum_{k=1}^{N}\frac{1}{\sigma_k^2}\hat{A}_{kj}(f)\hat{A}_{kj}^{*}(f)}} \tag{2.8}$$

The above represents a weighted version of the PDC factor, where $\hat{A}_{ij}(f)$ is the estimated complex coefficient from time series j to i at frequency f, σ_i, σ_k are the standard deviations associated with the respective time series and finally $\hat{A}_{kj}^{*}(f)$ is the complex conjugate of $\hat{A}_{kj}(f)$, summed over all N time series present in the system being studied.

2.6.2 Machine Learning

Machine Learning (ML) has emerged as a discipline within the Artificial Intelligence (AI) domain, to provide us with a new approach for solving complex problems. ML is defined as the field of study that grants computers the ability to learn without being explicitly programmed, as it focuses on the development of algorithms which can learn from data and make predictions based on them. Through Machine Learning, systems are not merely coded to execute specific tasks but rather to adapt and improve their performance as they are exposed to more data over time. This particular characteristic, has allowed the handling of many tasks which were previously considered impractical or even impossible to solve with fixed program instructions.

The most commonly highlighted example used to explain how ML is designed to work, is it's application to the recognition of handwritten digits into their corresponding numerical values. Let's consider some handwritten numbers. Each number can be considered as a single digit image. This digit image, can be represented as a vector of real numbers, serving as the input to a machine learning model. The complexity of this task, originates from the diversity of individual handwriting, which makes the accurate identification (or classification) of handwritten digits a difficult task to be performed through conventional programming methods. But what if, we had an entire dataset filled with digit images written in diverse handwriting styles? Then, we could consider this our ML training set, which will be used to fine-tune the parameters of an adaptable model. This model, after being trained with the dataset, will be capable of recognizing new, previously unseen digit images by generalizing from it's gained experience. This example, showcases in a simple but effective way, the power of ML in pattern recognition and prediction.

The previous problem, is considered an example of supervised learning. But, the field of Machine Learning is split into three types of learning methods; namely, Supervised, Unsupervised and Reinforcement Learning (RL).

Supervised Learning, which was illustrated with the handwritten digit problem example, involves learning a function that maps an input to an output based on example input-output pairs. It operates under the guidance of labeled data, aiming to predict the output associated with new inputs.

Unsupervised Learning, unlike Supervised Learning, deals with data that does not have labeled responses. The goal in this case is to discover inherent patterns or groupings in the data, such as clustering similar examples together. In the context of the digit images example, unsupervised learning could be used to group images based on visual similarities without actually knowing the true digit each image represents.

Finally, Reinforcement Learning (RL) focuses on making sequences of decisions. It involves an agent that learns to achieve a certain goal in a complex, uncertain, environment by performing actions and assessing the results. RL is distinguished by its emphasis on learning through trial and error, guided by the environment's feedback on the agent's actions.

2.6.3 ML Classifiers : SVM and RandomForest

In the context of the present thesis we will be dealing with a binary classification problem. More precisely, our data can be considered as labeled, as we will use events marked by epileptologists as our guides. The events marked will be considered labeled as True, whereas all others will be considered labeled as False. Hence, as we have learned from the handwritten digit problem, we will have to utilize some Supervised Learning algorithms.

Support Vector Machines (SVM)

Support Vector Machines belong to the supervised learning models and are commonly used to resolving classification tasks or regression problems, by finding the optimal hyperplane which separates different classes/features.

Let's first describe the concept behind SVMs. In an effort to design a system that filters out unwanted spam emails, we might look to the principles of Support Vector Machines (SVMs), which provide a way to classify and separate data. To simplify, consider the task of distinguishing between regular emails and spam emails from a persistent sender.

Imagine plotting all received emails on a graph, with the axes representing different characteristics of the emails, such as frequency of certain words, time of sending, etc. In this graph, we want to draw a line (or hyperplane in higher dimensions) that separates the spam

emails from the regular ones.

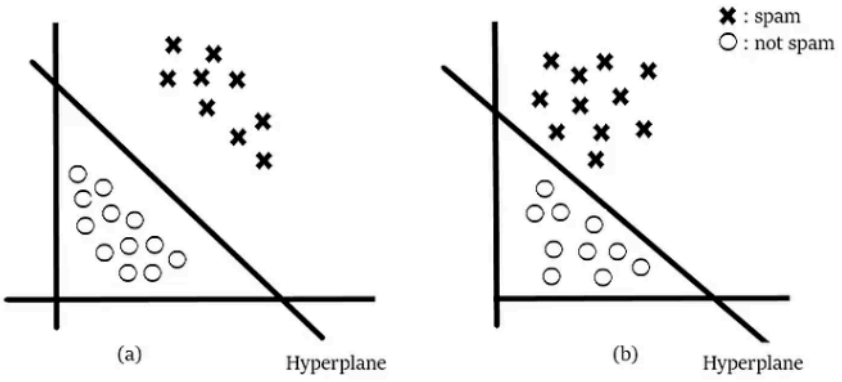


Figure 2.6: SVM hyperplane choice concept [31]

The provided figures (Figure 2.6) demonstrate two potential hyperplanes that could perform this task. To decide which hyperplane is more suitable, we must consider which one creates the widest distance between the spam and non-spam emails, as this margin helps in reducing misclassification.

The most effective hyperplane is the one that not only separates the emails cleanly but also maintains the greatest distance from the nearest points of any class (in this case, the spam emails and the non-spam emails). By selecting the hyperplane with the largest margin, we ensure that our email filter is robust and less likely to incorrectly label new emails as spam or vice versa. Thus, in choosing between the two hyperplanes, the preference should be given to the one that offers this optimal separation.

The points which are closest to the hyperplane, are called support vectors. These data points are the ones which influence the final position and orientation of the hyperplane. Using these support vectors, with SVMs, we aim at maximizing the margin of the classifier.

After gaining a clear view on what SVMs are, let's try to explain the mathematics explaining this concept [32].

Consider l training examples $\{x_i, y_i\}, i = 1, \ldots, l$, where each example has d inputs ($x_i \in \mathbb{R}^d$), and a class label with one of two values ($y_i \in \{-1, 1\}$). Now, all hyperplanes in $\mathbb{R}^d$ are parameterized by a vector (w), and a constant (b), which can be written as:

$$w \cdot x + b = 0 \tag{2.9}$$

Where w is the vector orthogonal to the hyperplane. Given such a hyperplane (w, b) that separates the data, this gives the function

$$f(x) = \text{sign}(w \cdot x + b) \tag{2.10}$$

which correctly classifies the training data and hopefully other "testing" data it hasn't been trained on yet. However, a given hyperplane represented by (w, b) is equally expressed by all pairs $\{\lambda w, \lambda b\}$ for $\lambda \in \mathbb{R}^+$. So we define the canonical hyperplane to be that which separates the data by a "distance" of at least 1. That is, we consider those that satisfy:

$$x_i \cdot w + b \geq +1 \text{ when } y_i = +1 \tag{2.11}$$

$$x_i \cdot w + b \leq -1 \text{ when } y_i = -1 \tag{2.12}$$

or more compactly:

$$y_i(x_i \cdot w + b) \geq 1 \quad \forall i \tag{2.13}$$

All hyperplanes have a "functional distance" ≥ 1. For a given hyperplane (w, b), all pairs $\{\lambda w, \lambda b\}$ define the exact same hyperplane, but each has a different functional distance to a given data point. To obtain the geometric distance from the hyperplane to a data point, we must normalize by the magnitude of w. This distance is simply:

$$d((w, b), x_i) = \frac{y_i(x_i \cdot w + b)}{\|w\|} = \frac{1}{\|w\|} \tag{2.14}$$

Intuitively, we want the hyperplane that maximizes the geometric distance to the clos-

est data points. (See Figure 2.6.) So finally SVMs solve:

$$\max_{w,b} \frac{1}{\|w\|} \quad \text{subject to} \quad y_i(x_i \cdot w + b) \geq 1, \ i = 1, \ldots, l. \tag{2.15}$$

Random Forest

The Random Forest algorithm represents an advanced ensemble, supervised learning technique commonly used for both regression and classification tasks. The methodology is based on the construction of multiple decision trees during the training phase, where the final decision derives by averaging the outcomes of these individual trees. This approach leverages the power of multiple predictive models to achieve a final decision, thereby enhancing the accuracy and robustness of predictions.

The foundation of the Random Forest model is built upon three key concepts: decision trees, ensemble learning, and bootstrapping. Decision Trees are characterized by their tree-like structure, as their name states. These models start their decision-making process at the root node. In the context of classification, they iteratively split the data space into binary outcomes based on the predictive variables' values. This recursive partitioning process, continues until a terminal node, commonly referenced as 'leaf' is attained and the classification outcome is defined. Ensemble learning augments the predictive strength by synthesizing the results of various models trained on the same dataset and, finally, bootstrapping contributes to this synthesis by facilitating the random sampling of subsets from the dataset, which are then used to construct individual trees whose results are aggregated to produce a final prediction.

Random Forest models are appreciated for their ability to handle high-dimensional datasets and produce outcomes that are not only accurate but also less prone to overfitting compared to individual decision trees. This robustness stems from the diversity of the decision trees within the forest and the randomness introduced through bootstrapping, which ensures that the models generalize well to new, unseen data. A more detailed explanation on its implementation in our study will be given in Chapter 3, along with scikit-learn library's [33] implementations of both SVMs and RandomForest models.

Chapter 3

Materials & Methods

3.1 Study Design

As it was previously mentioned, the present work is a N-of-1 study of a patient with refractory focal epilepsy. The study took place in Münster, Germany and the trial was performed at the University Clinic Münster. The following information on the nature of the study as also the equipment, its calibration and other hardware information was kindly provided by Prof. Carsten H. Wolters and his master's student, Mr. Fabian Kaiser, who worked on the trial himself and wrote his thesis on the subject [34] [1].

3.1.1 Ethics Statement & Patient (Case) Description

A local ethics committee approved the research protocol, and a 23-year-old patient participated, giving written informed consent before the beginning of the study process. Initially diagnosed with epilepsy at fourteen (14) years old, which was characterized by disrupted thought and speech but without motor or awareness issues, they experienced seizures multiple, to be precise an average of four, times per day. The condition, which was proven unresponsive to various anti-epileptic drugs (AEDs), was classified as a case of drug-resistant epilepsy (DRE). Despite mild semiology, the seizures significantly affected the patient's quality of life. Presurgical assessments, including videoEEG and FDG-PET scans, indicated seizure origins in the left frontal area, near an area responsible for speech, raising surgery risks. Invasive EEG (iEEG) was also applied, and missed the epileptogenic zone, but imaging and EEG source analysis later identified a focal cortical dysplasia near Broca's area, precluding resection.

Follow-up tests and seizure tracking continued amidst unsuccessful medication trials. In 2019, another MEEG recording was executed, finding numerous Interictal Epileptiform Discharges (IEDs), while AED therapy was paused, as to proceed with source estimation. The patient kept a seizure diary throughout the duration of the study.

3.1.2 Study Design

This double-blind sham-controlled clinical trial, firstly aimed to evaluate the impact of individually tailored and dCMI-optimized transcranial Direct Current Stimulation (tDCS) on the frequency of Interictal Epileptiform Discharges (IEDs). The patient was examined for one-hour long sessions, consisting of twenty (20) minutes of stimulation, followed by a 20-minute break, and another 20 minutes of stimulation, daily, over the course of five consecutive days.

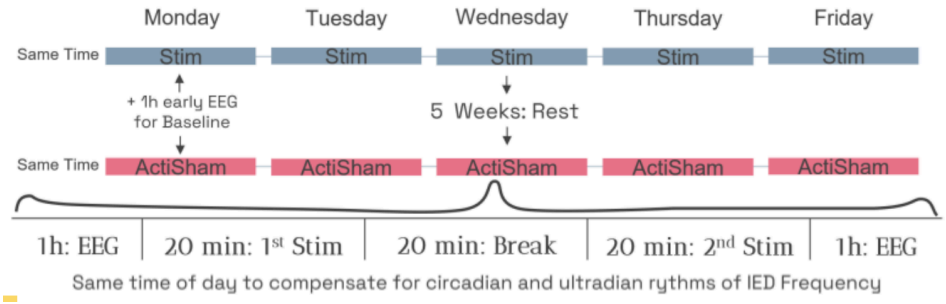


Figure 3.1: Study Design[35]

To serve as a control, an ActiSham stimulation protocol was designed, using a montage with adjacent cathodes and anodes placed identically to the dCMI-optimized setup, to ensure only the sensation of stimulation was felt without significant cerebral effects (Figure 3.4. Considering the patient's history of refractory epilepsy and their anticipation of a genuine treatment experience, the utilization of any kind of scalp anesthesia was avoided. It is worth noting that the adverse effects associated with tDCS are generally considered minimal, particularly when compared with the challenges of living with epilepsy.

After five weeks of rest, which were considered mandatory in order to take into account any residual effects of stimulation from the first phase of the trial, the ActiSham protocol was applied for five consecutive days over a week, following a similar approach to the first one. EEG recordings were performed immediately before and after each day's stimulation session.

It is worth mentioning that extended two-hour EEG sessions were conducted on the first day of both the Stimulation and ActiSham weeks, to establish a baseline, which was condensed to one-hour averages for analysis purposes. Monitoring the consistency of IED recording times was crucial due to the potential variability in IED frequencies influenced by circadian rhythms [36]. The ActiSham also controlled for any potential discrepancies in ultradian rhythms that might have arisen day-to-day [37].

Finally, it is important to mention that both EEG and stimulation sessions were carried out at the University Clinic Münster, as mentioned, under medical oversight by staff which was unaware of the study's design. Kaiser, had personally configured the tDCS management and control software in advance, ensuring that stimulation parameters remained concealed from the medical personnel that was administering the tDCS treatments.

3.1.3 tDCS Hardware

As noted in the previously mentioned master thesis, a StarStim® device developed by Neuroelectrics Barcelona SL [38] was used for tDCS, along with the NE019 Neoprene Headcap from the same company, with 39 predefined electrode positions based on a subset of the 10-10 EEG system (Figure 3.2). Automatic impedance checking was applied before every stimulation and ramp in and ramp out time was 60 seconds each.

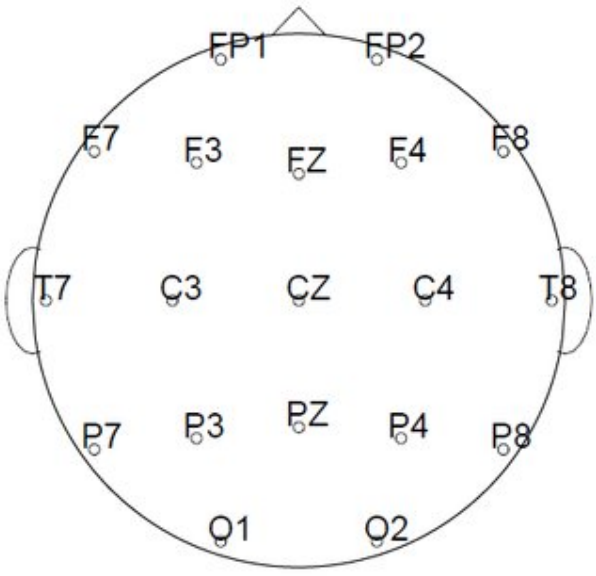


Figure 3.2: This system features all the electrode locations placed on the scalp, according to a subset of the 10-10 system.

3.1.4 dCMI optimized montage

The treatment-focused Distributed Constrained Maximum Intensity (dCMI) configuration was developed by Dr. Antonakakis utilizing a pipeline explained in detail in his 2021 dissertation [39]. This approach integrates automated segmentation of six head tissue categories identified through T1- and T2-weighted MRI scans with hand-delineated segmentation of surgical burr holes derived from CT scans. It also incorporates considerations for the differing conductivities of white and gray matter observed in diffusion tensor imaging (DTI) scans. Additionally, it calibrates skull conductivity referencing SEP/SEF (somatosensory evoked potential or field) source estimations for accuracy and concludes with the computation of a tDCS setup (Figure 3.3), which is refined through dCMI methods as documented by Khan and colleagues in their 2019 and 2022 studies [40] [41]. For this thesis, only the EEG recordings were utilized.

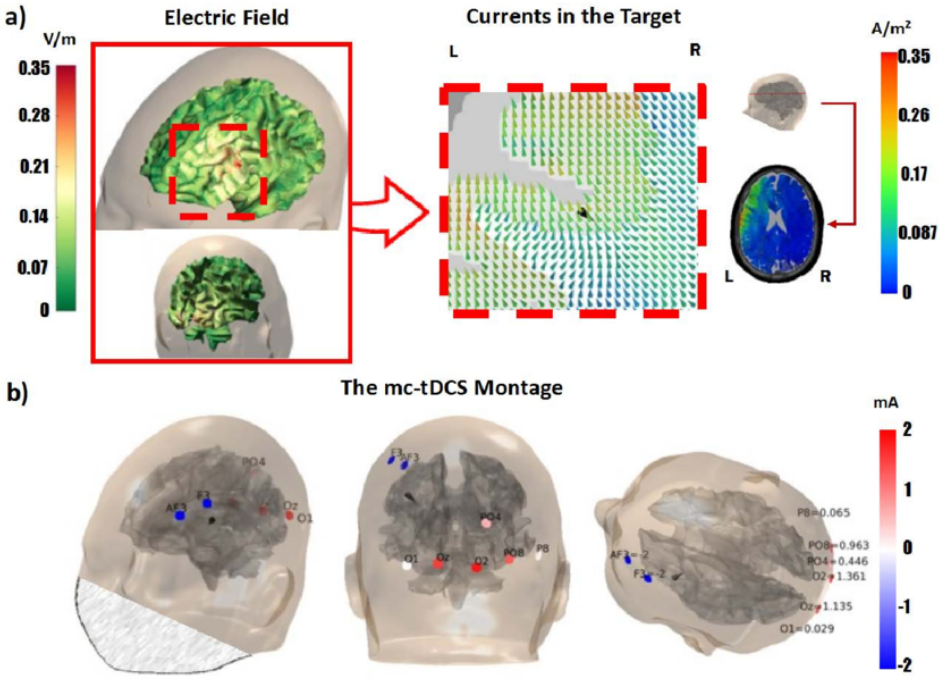


Figure 3.3: EMEG targeted and DCMI optimized 8-channel mc-TDCS montage. a) Distribution of the simulated electric field over the patient's cortex (left subfigure) and zoomed view of the injected current (color-coded cones) in the target region (black cone) (right subfigure). b) Three different views show the D-CMI optimized 8-channel mc-tDCS montage, targeted at the EMEG centroid at spike onset, used for the D-CMI stimulation condition. The sum of absolute values of all currents is 8 mA with a limitation of max 2 mA per electrode [42]. In the present work only EEG recordings performed before and after the stimulation/sham process were used. This figure, aims to illustrate the initial trial setup.

3.1.5 ActiSham montage

As for the ActiSham montage, in terms of electrode placement it mirrors dCMI's. However, the applied current differs, having values : AF3: -2 mA, F3: 2 mA, O2: 2 mA, Oz: -2 mA and the rest: 0 mA.

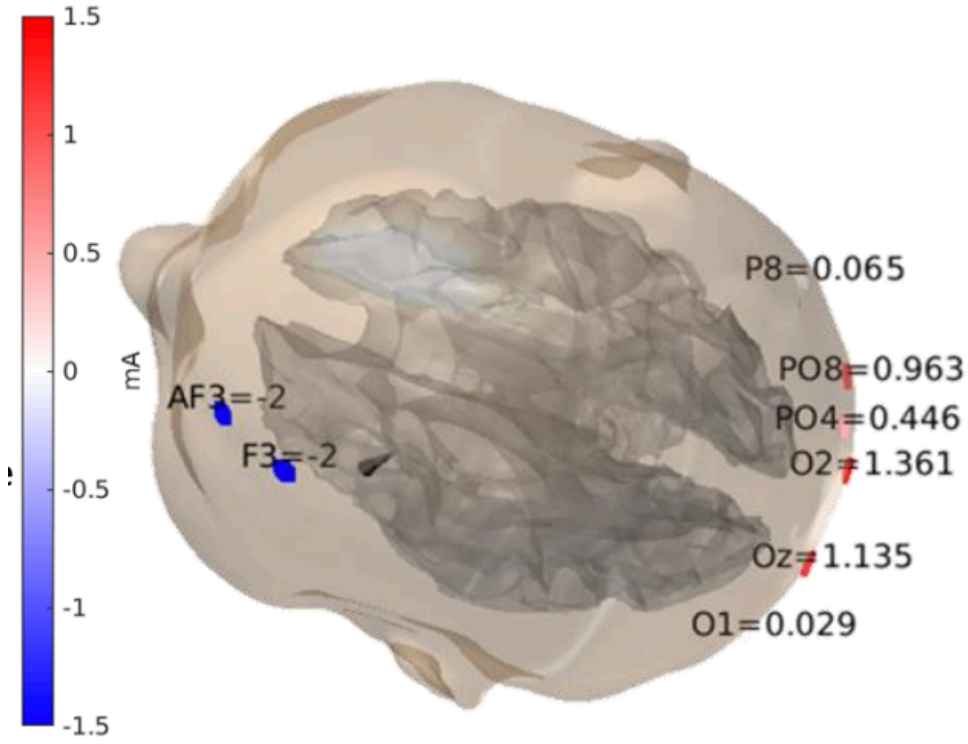


Figure 3.4: Shows the dCMI optimized montage with current in A. The ActiSham montage is the same as the dCMI optimized montage when it comes to electrode positions[35].

3.1.6 EEG Data Acquisition

For the EEG data collection within the study, a Nihon Kohden EEG system from Tokyo, Japan was utilized, recording at a sampling rate of $f_s = 200Hz$. 19 electrodes were positioned in the tDCS headcap, according to the 10-10 system.

3.1.7 Interictal Epileptiform Discharge (IED) Detection

Three experienced epileptologists were asked to mark the EEG recordings in order to identify the Interictal Epileptiform Discharges (IEDs). The data was segmented into one-hour (.edf as for European Data Format (EDF)files / recordings, which was then anonymised and assigned

pseudonyms to ensure the epileptologists were unaware of the timing or sequence of each data segment they were marking, in order to maintain the concept of a blind analysis. Secure transfer of the data was facilitated through encrypted .zip files via Sciebo , adhering to the data management protocols set by the ERA PerMed endorsed PerEpi plan. The actual process of marking IEDs was conducted visually, utilizing the BESA software, as referenced by Scherg and colleagues in their 2018 publication [43].

3.1.8 The patient's point of view

In order to closely monitor the patient's subjective experience and any potential side effects throughout the duration of the trial, a daily questionnaire was administered.The responses provided a detailed account of the patient's experiences during both the stimulation and the sham sessions of the trial, where the patient had no information on which week corresponded to which part of the trial. The following summary aims to present the key findings as they were gathered and written down by Mr. Kaiser. Firstly, no negative outcomes were observed during or after the personalized,optimized dCMI tDCS or ActiSham treatments, and none of the treatment sessions were discontinued. Based on a questionnaire empolying a 5-point numerical rating scale , where 1 signified no sensation and 5 an extreme one, the patient reported experiencing an itching sensation at the back of their head with an average intensity of 3.4 ± 0.55 (ranging from 3 to 4 out of 5) during the stimulation sessions and 2.8 ± 0.45 during the sham sessions. Additional sensations reported included pain located under electrode F3, with values being 4.2 ± 0.84 for stimulation and 3.4 ± 0.55 for sham , a burning sensation was evaluated by the patient as 2 ± 1 for stim and 2 ± 1.41 for sham, a feeling of warmth was noted with 2.2 ± 0.84 for stim and 1.4 ± 0.89 for sham, tiredness or decreased attention was evaluated with 10 [consistent at 1] for stim and sham too and dizziness , with "Stim": 3.2 ± 0.45 and "Sham": 1 ± 0 [consistent at 1]). The intensity of dizziness was the only sensation that showed a significant difference between the stimulation and sham sessions. Finally, the patient was overall uncertain about whether they were receiving the "Stim" or "Sham" treatment on most days, confidently identifying the "Sham" session correctly once during "Stim" and once during "Sham", and mistaking the "Sham" for "Stim" on one occasion.

3.2 Code Implementation Overview

The main goal of the present work is to exploit the information contained in the EEG data provided by the aforementioned trial, for an effective brain connectivity study, focusing on the IEDs marked by the epileptologists and their characteristics, and then proceed to a supervised Machine Learning analysis. The main pipeline steps included firstly the preprocessing of the raw EEG data, followed by an effective connectivity analysis and then by feature extraction and ML analysis.

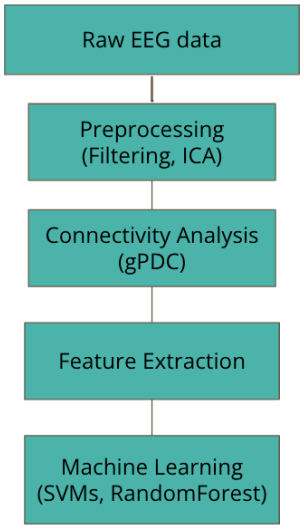


Figure 3.5: A brief overview of our pipeline.

It is worth mentioning that for the development and execution of the computational analyses presented in this thesis, Python 3, specifically version 3.11, was utilized as the primary programming language. This choice was made due to Python's extensive library support and its newly widespread adoption in scientific computing and neuroscience (e.g. MNE - python library [44], scikit-learn [33]for ML) The compatibility and performance features offered by Python 3.11 were instrumental in facilitating the complex data processing and analysis tasks required for this study. Finally, all code was written and tested in this environment, using Spyder editor, to ensure the consistency and reproducibility of results.

3.2.1 Data Preprocessing

As it has been previously mentioned, EEG recordings are a pivotal source of data in neuroscience as they capture essential brain information, more specifically, the electrical activity of the brain, with high temporal resolution and without being considered an invasive procedure on the human brain. However, as most experimental data, the raw data captured from EEG recordings is typically contaminated by signals of non-neural origin, mainly various sources of noise and artifacts, including electrical interference, muscle movements and eye blinks. Preprocessing these recordings is therefore essential to isolate genuine important neural activity and proceed with a robust analysis which will ultimately produce accurate results.

Visual Inspection and Filtering

The first preprocessing step when dealing with EEG recordings, usually involves the visual inspection of the raw signal. During this process, bad channels are dropped, meaning individual channels which malfunction and provide data that is too noisy to be usable.

Afterwards, follows the filtering of the data. Human EEG, largely comprises signal power in a range of frequencies from 1–30 Hz. This narrow band contains the majority of neurophysiological information relevant to most research questions, and particularly in event-related potential (ERP) studies, as this one. While there is evidence suggesting the presence of meaningful information beyond this range, extending to 100 Hz[45], the focal point of investigation typically remains closely within these confines [46]. There are five widely recognised brain waves, namely delta (δ), theta(θ), alpha (α), beta(β), and gamma(γ), whose frequency ranges and characteristics are presented in Figure 3.6.

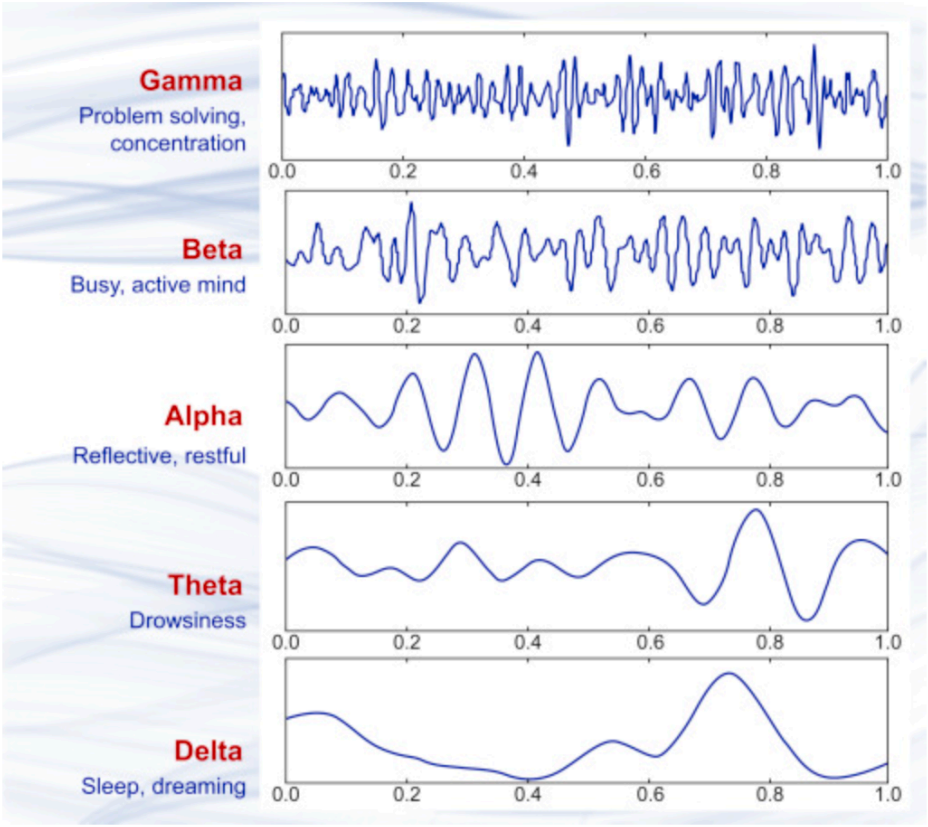


Figure 3.6: Brain wave samples with dominant frequencies belonging to beta, alpha, theta, and delta bands and gamma waves[47]

Our EEG recordings, were sampled at a frequency of 200 Hz, therefore, after applying baseline correction, in order to maintain all useful neural information and based on the studies stated above about neural brain wave frequencies, we decided to exploit MNE-Python's 'filter' method and apply on each 1-hour EEG raw data a finite impulse response (FIR) band-pass filter, with a lower cutoff frequency of 0.05 Hz and an upper cutoff frequency set at 45 Hz. The filter was designed using the 'firwin' method provided by MNE, which is a window-based approach to create a filter with specific ripple characteristics within the frequency bands set. This band-pass filter was implemented to avoid undesirable low-frequency drift and high-frequency noise, while preserving the EEG signal components of interest in the given range.

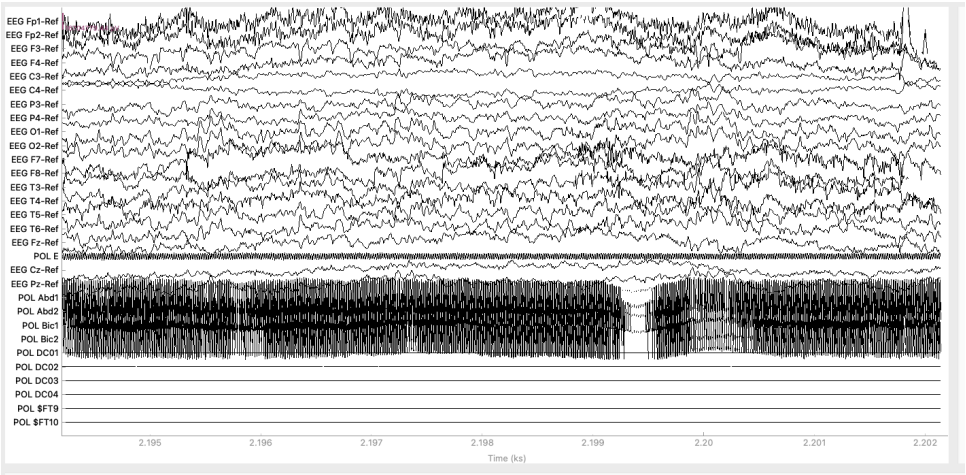


Figure 3.7: Raw EEG signal before preprocessing

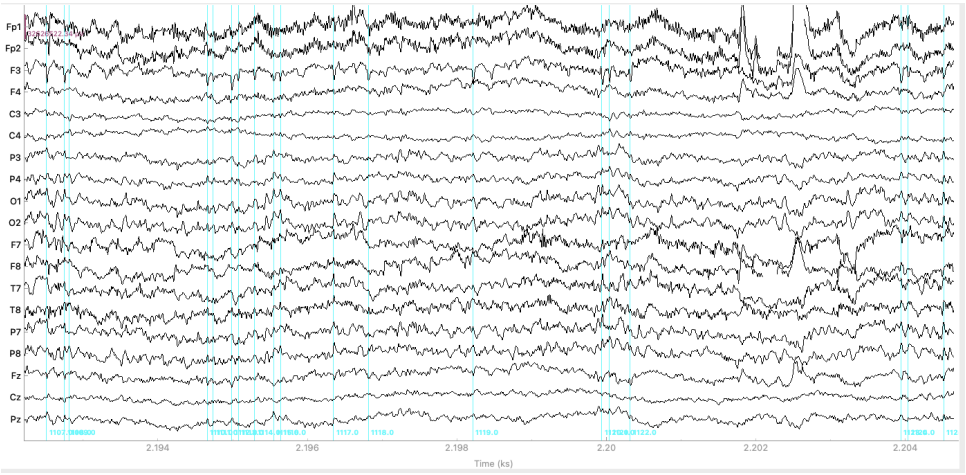


Figure 3.8: Raw EEG signal after preprocessing, with the IEDs marked on the signal in cyan color

Artifact and Noise Detection : Independent Component Analysis (ICA)

Another critical point of this analysis, involved the application of Independent Component Analysis (ICA). Artifact independent components (ICs) can be identified by experts with visual inspection, via plotting the components, their properties and corresponding signals and identifying any components that should be dropped from the analysis. However, sometimes even experts can disagree on a component's categorization. Hence, some research has been produced in order to provide specific guidelines on the visual inspection process. One of them, includes the article by M. Chaumon et al.[48] which was used as the main guideline for our ICA analysis. During the ICA application process, it was fairly noticeable that all the EEG recordings were

subject to blink component artifacts and horizontal eye movement components as well as some muscle components too in fewer cases.

Finally, during the preprocessing phase, the proper EEG montage, which was a subset of the 10-10 system, was configured, in order for the scalp topology to be set for the connectivity analysis. All the preprocessing phase figure representations can be also found in Chapter 4, Results, where they will be analyzed further.

3.2.2 Connectivity Analysis

Data segmentation

As it was mentioned in Paragraph 3.1.7, the IEDs present in the EEG recordings were marked by three different epileptologists. These so called 'events' were marked for a specific time point/ millisecond in time and do not refer to a time period (Figure 3.8).

Since we aimed to perform a connectivity study, it was essential to analyze the connections over time, rather than at particular moments. This necessity stemmed from the understanding, that connectivity is a dynamic process reflecting the brain's ongoing and evolving activity. Hence, even though IEDs were precisely marked by the clinicians at specific milliseconds, our analysis required the investigation of data across broader time windows, in order for us to capture the temporal relationships and the change of brain interactions over time. As of that, we decided to split the raw, preprocessed signal into 3 second windows with a 30% overlap. Since the sampling frequency was of 200Hz , every window contained 600 time points, on which our analysis was performed. Finally, when splitting the signal into windows, we also labeled them based on whether they contained an event or not, to then focus mostly on the windows containing events.

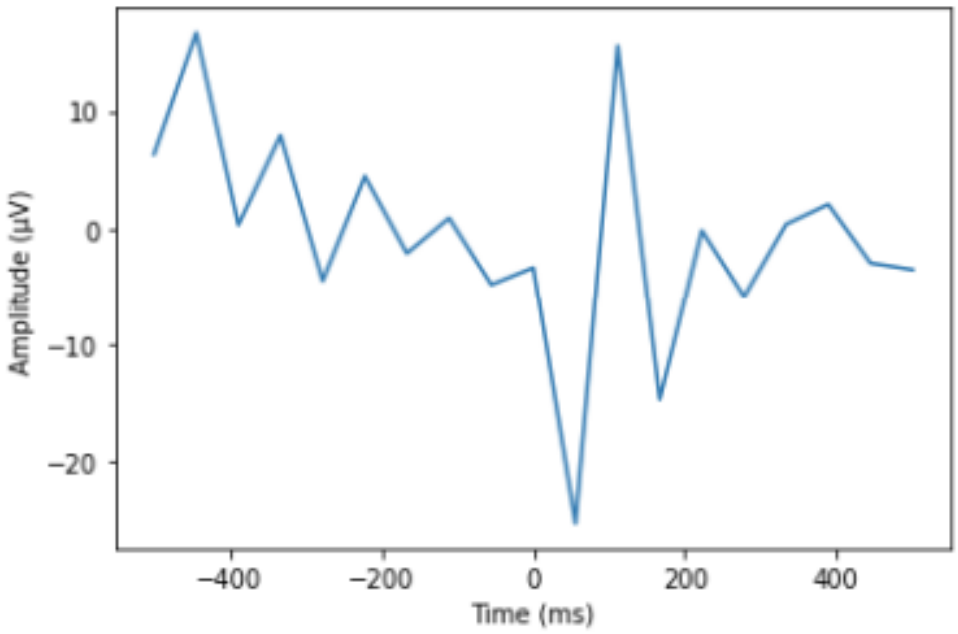


Figure 3.9: Average EEG signal for a window containing an event (labeled True)

generalized Partial Directed Coherence

After concluding the preprocessing and segmentation phase of the analysis, we moved on to the connectivity analysis. As has been explained in the previous Chapter of the present thesis, Generalized Partial Directed Coherence (gPDC)[49] is an important metric in the concept of effective connectivity analysis, especially in the context of epilepsy. It provides insights into the directional influences between different brain regions, revealing how different areas communicate during both normal and epileptic brain activity. When applied to epileptic patients, gPDC can highlight the patterns of neural interactions that may precede or accompany seizures. By examining these patterns, researchers can improve the understanding of epilepsy's underlying mechanisms, leading to better management and treatment strategies for patients. The precision and efficacy of gPDC, have made it become a powerful tool for the study of human neurological disorders like epilepsy [50].

gPDC : the spectral_connectivity package

To implement and calculate the gPDC metric in our code, we decided to exploit the spectral_connectivity package as it was produced by the Eden-Kramer-lab [51], which has also been recently partially incorporated by MNE-Python. For our study, we focused on the implemen-

tation of the generalized_partial_directed_coherence method [49]. This function, is designed to compute gPDC across the entire frequency spectrum of the multichannel time - series data. The process begins with the application of the Multitaper method, a spectral estimation technique known for its ability to reduce variance and bias in the spectral estimates. Upon several calculations concerning the multitaper spectral estimates, the function performs a Fast Fourier Transform (FFT) to obtain the frequency-domain representation of the data. Afterwards, a Connectivity object is instantiated using the Fourier coefficients derived from the FFT calculations, alongside the frequency and time attributes provided by Multitaper. Finally, the output of all the aforementioned process represents the connectivity matrix for each frequency, providing a comprehensive view of the directed coherence across the entire spectrum (Figure 4.1).

Thresholding : Surrogate data production

By now, it has become clear that gPDC as a measure, reveals the directional flow of information between two channels, meaning it reveals the inflow and outflow information from a channel to another. The problem that now arises, is that most parts of the brain are connected and neural information constantly flows from and to multiple channels, which can also be defined as the nodes of a graph [52]. The question that is subsequently asked, is how could it be determined which information/connection should be considered important? The answer, lies in the thresholding process, or as most recently called, the graph filtering.

To determine whether the information flow of a channel towards another is valuable, a robust statistical thresholding method should be applied, which will allow us to highlight all strong channel connections, identifying significant interactions between brain regions and preventing us from focusing on less informational ones. A plethora of studies have tackled this issue [53] [54]and a robust solution found, is the creation of surrogate data. In our code implementation, a function was devised to determine these thresholds by creating surrogate data.

By simulating the null hypothesis — that is, the assumption of no genuine connectivity — surrogate data allows for the establishment of thresholds that separate statistically significant interactions from those arising by chance. For the generation of the surrogate data, the iterative

Amplitude Adjusted Fourier Transform (iAAFT) was utilized. The iAAFT process [55] helps in creating statistically correct variations of our original dataset, maintaining the amplitude distribution and autocorrelation of the time series while randomizing the phase information. This technique is the basis on which we constructed a null model against which the actual data's connectivity patterns can be tested, ensuring that identified connections are not merely artifacts of the data's inherent structure but represent real neural interactions.

Upon the generation of surrogate data via iAAFT , our analysis advances to the critical stage of threshold determination and visualization. By calculating the gPDC values across a multitude of surrogate datasets and identifying the 96th percentile as our threshold, we establish a statistically robust criterion that keeps the significant directional interactions amidst the many possible channel connections. This percentile threshold, chosen based on prevailing standards in connectivity analysis literature [56], ensures that only the most important, consistent patterns of neural communication are highlighted, effectively filtering out spurious connections which do not need to be studied for this work. It is worth mentioning, that all the above, meaning surrogate data creation and gPDC as also surrogate data calculations, were applied on channel pairs and not single channels, in order to comprise the directionality of the gPDC measure. Finally, the thresholding process is presented through the generation of histograms for each channel pair, visualizing the distribution of surrogate gPDC values against the delineated threshold (Figure 3.10). These histograms not only serve as a graphical representation of our method, but also as a transparent and intuitive means for identifying significant neural pathways.

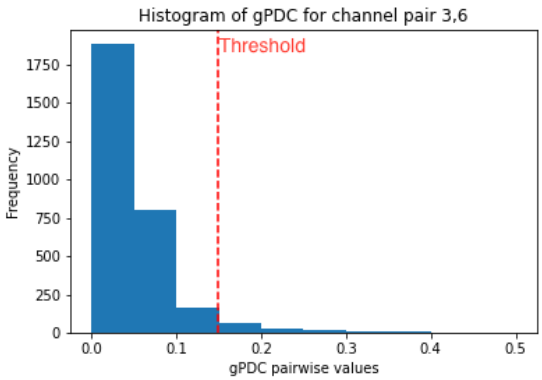


Figure 3.10: Histogram deriving from the thresholding process of the gPDC values with the calculation of pairwise thresholds. In this representation the pair studied is (3,6), meaning that this threshold value will be set for information flowing from channel 3 to channel 6, as they have been mapped. Axis x represents the gPDC pairwise values, and axis y the frequency with which the corresponding surrogate values lie in the same values as gPDC.

This way, the function embodies a comprehensive approach to connectivity analysis, from the generation of surrogate data with iAAFT to the statistical validation and visual exposition of significant neural interactions, providing us with a robust method for further calculations.

All graph representations deriving from the thresholding process and connectivity analysis in general will be explained in detail in 4, along with the rest of the analysis results.

3.2.3 Feature Extraction

After completing the connectivity analysis calculations, let's now change the concept of channels and translate it to nodes in a large graph, or otherwise, a vast network. In order to verify the quality of our effective connectivity analysis but also to move it a step forward by adding a ML classification afterwards, we decided to proceed to a feature extraction process. The analysis involved several steps, each designed to capture different aspects of the EEG data connectivity and dynamics. The strategy followed, was to exploit the connectivity properties calculated between EEG channels and then quantify these connections' characteristics.

After having already calculated the gPDC values for every window, as well as the pairwise thresholds for them, we considered these final matrices as graphs. Within these graphs, we computed global and local efficiency measures to understand the overall and local connectivity's effectiveness respectively. Additionally, betweenness centrality was calculated to identify the channels which are considered to serve as critical communication hubs in the network. To

capture the dynamic nature of connectivity, we analyzed transitions between different connectivity states, namely clusters [57], and quantified the flexibility of channels in changing their cluster memberships. This metric, was useful in helping us understand the variability in brain network configurations over time.

Besides network-based measures, we also extracted statistical features per channel, such as mean, standard deviation , skewness and kurtosis values. More specifically, the **mean** is calculated as the average value of the EEG signal over time, indicating the central tendency of the signal amplitude. **Standard Deviation (std)**, measures the amount of variation of the EEG signal from the mean. A higher standard deviation indicates greater variability in signal's amplitude. **Skewness**, measures the asymmetry of the probability distribution of the EEG signal. Positive skewness indicates a distribution with an asymmetric tail extending towards more positive values, while negative skewness indicates a tail extending towards more negative values. Finally **kurtosis**, measures the "tailedness" of the probability distribution of the EEG signal. High kurtosis means that the distribution has heavy tails or outliers, indicating extreme deviations from the mean. "Taildness," referring to the term of kurtosis in statistical language, is a measure of the shape of a probability distribution, specifically focusing on the "tails" of the distribution. As tails of a distribution are considered the extreme ends of the distribution curve, far from the mean or median. Kurtosis quantifies how heavy or light these tails are compared to a normal (Gaussian) distribution.

These statistical features provide a foundational understanding of the characteristics of EEG signals across different channels, highlighting variations and anomalies in brain activity. Finally, these diverse features were aggregated to form a multidimensional feature space for each time window in which we had split our EEG data. This features space, then served as the input for ML classifiers, aiming to verify our connectivity analysis' quality of results.

3.2.4 Machine Learning

After completing the feature extraction process, the next phase involved leveraging these features to train Machine Learning models. The goal was to assess the predictive capability of the extracted features in distinguishing between seizure and non-seizure states. This phase is

crucial for validating the effectiveness of the connectivity analysis performed earlier.

Data Preprocessing for Machine Learning

Before training the models, the extracted features underwent several preprocessing steps to ensure they were in a suitable format for machine learning algorithms. This included converting complex data representations into a structured format, handling missing values, and normalizing the data to ensure uniformity across all features. These preprocessing steps were vital in preparing the high-dimensional feature set for effective model training and validation. Another point of the process also worth mentioning, is that the initial EEG data were highly imbalanced. In most data recordings, from a total of approximately 1700 windows, on an average only about 200-250 windows contained events. Hence, in order to compute a robust analysis, while also reducing the complexity of our initial connectivity calculations, we decided to calculate connectivity and features only for an equal number of windows containing and not containing events. In order to also exploit fully the windows containing events, we decided to always sample a, random, but equal portion of False labeled windows as the True ones. This way, not only a lot of computation time was saved but also, due to the randomness of the False labeled windows chosen per file, the analysis was more robust and in the ML phase, the class imbalance problem had already been solved.

Model Selection and Training

To evaluate the predictive performance of the extracted features, we employed two different machine learning models: Support Vector Machines (SVM) and Random Forest classifiers. Each model was chosen for its unique ability to handle complex, high-dimensional data for binary classification, offering a comprehensive assessment of the features' predictive power.

SVM and Random Forest

Both the SVM and Random Forest classifier hyper-parameters were extensively fine-tuned using grid search techniques. The models were first trained, and then evaluated using ten-fold cross validation to ensure their robustness and generalisability across different subsets

of the data [58]. The performance of the fine-tuned models was finally tested on a separate test set (training set: 80%, test set:20%) to measure their generalization capability. Metrics such as accuracy, precision, recall, F1 score, and ROC-AUC were used for the evaluation of the final results. In Chapter 2, an extensive explanation of the concepts behind the SVM and Random Forest algorithms was provided.

Chapter 4

Results

In this chapter the findings derived from our complete pipeline's analysis will be presented, starting from the early preprocessing stages until the application of Machine Learning models for the evaluation of effective connectivity and the evaluation of the hypothesis that tDCS indeed reduces the IEDs while also changing the severity/strength of seizure symptoms. The aim of this chapter, is to provide a simple visual explanation of all the aforementioned concepts and processes followed in the present work.

4.1 Connectivity Analysis: Results

As it has been previously explained, one of the main purposes of this thesis is to evaluate the performance of effective connectivity on the EEG data recordings of a patient with refractory focal epilepsy. After the preprocessing phase was completed and the signals were segmented, we implemented the gPDC method. This method, produces, among others, a meshgrid shaped (n_channels,n_channels), which in our case was (19,19), where the generalized partial directed coherence values are displayed in the frequency domain. These values, were thresholded using surrogates and iAAFT, as it has already been described in Chapter 3. Here, we present a figure of a coherence meshgrid resulted after the thresholding process.

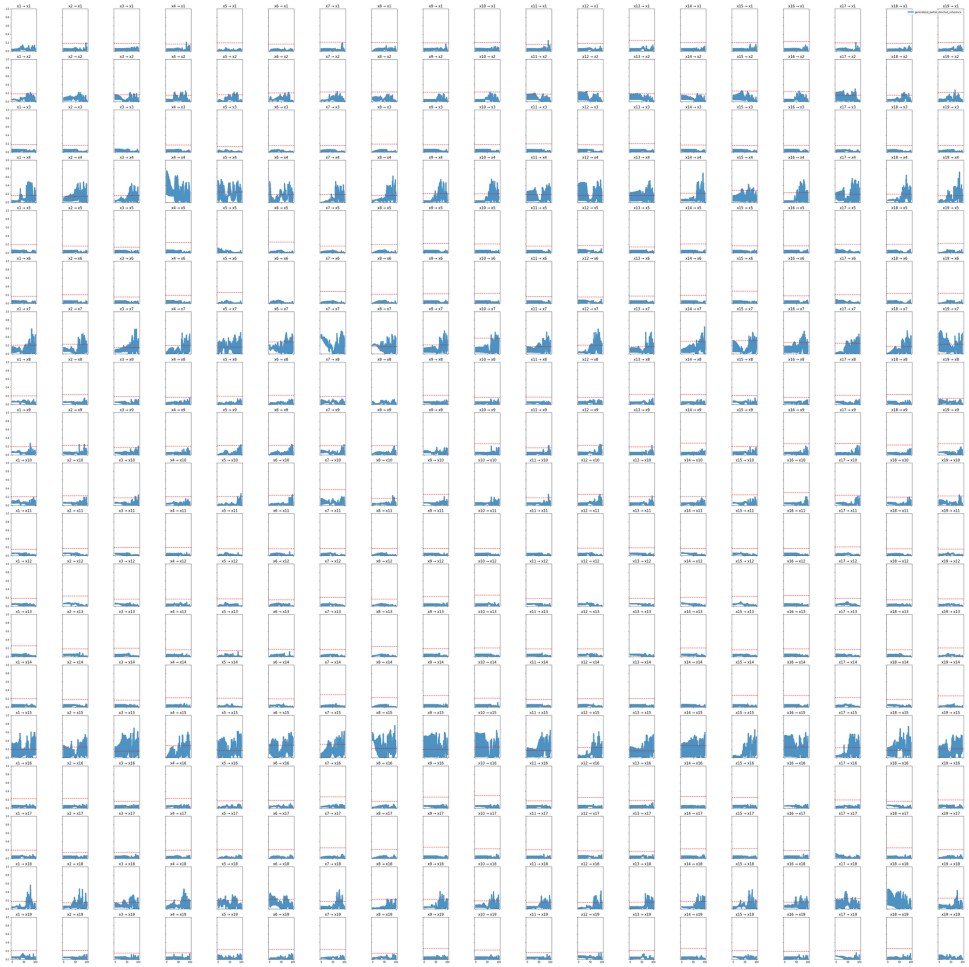


Figure 4.1: (19,19) meshgrid representation of the gPDC measure. The red dashed lines represent the threshold applied per pair of channels

Despite the fact that images such as the one depicted above are rich in information and provide a comprehensive overview of the necessary data, their readability and ease of interpretation leave much to be desired. Consequently, it was mandatory for us to address this issue, to facilitate the clear understanding and presentation of the effective connectivity analysis information. Hence, we produced the connectivity strength matrices, where gPDC values across all channel pairs were plotted in heatmaps, taking into account the pre-calculated pairwise thresholds, indicating the flow/strength of connectivity for each (source,destination) channel pair. The method devised for producing these plots, operated by iterating through each possible source and destination channel, excluding self-connections, to accumulate the sum of gPDC values that surpass the specified threshold, indicating the strength of connectivity from the source to the destination.

Another significant distinction, was also made between EEG recordings conducted

before and after stimulation, as well as those recorded before and after the sham procedure. As explained in Paragraph 2.5, tDCS is considered a promising treatment method especially for patients with refractory focal epilepsy, as is classified the subject of this study. Consequently, it's expected that the patient will present pronounced outflow connectivity patterns close to the seizure origin, which was found in the left frontal area (see Paragraph 3.1.1). Subsequently to the transcranial Direct Current Stimulation (tDCS) treatment though, it is further expected that there will be a modification in these connectivity patterns, likely demonstrating a more dispersed connectivity throughout the brain's regions. Following a similar trail of thought, the connectivity patterns before and after the Sham process are not expected to vary in connectivity and are in fact predicted to both showcase strong connectivity outflow patterns near the origin of the epileptic seizures, similar to the ones appearing before the stimulation. It should also be mentioned, that the IED markings of the epileptologists confirmed this hypothesis, as seen in the Figure below:

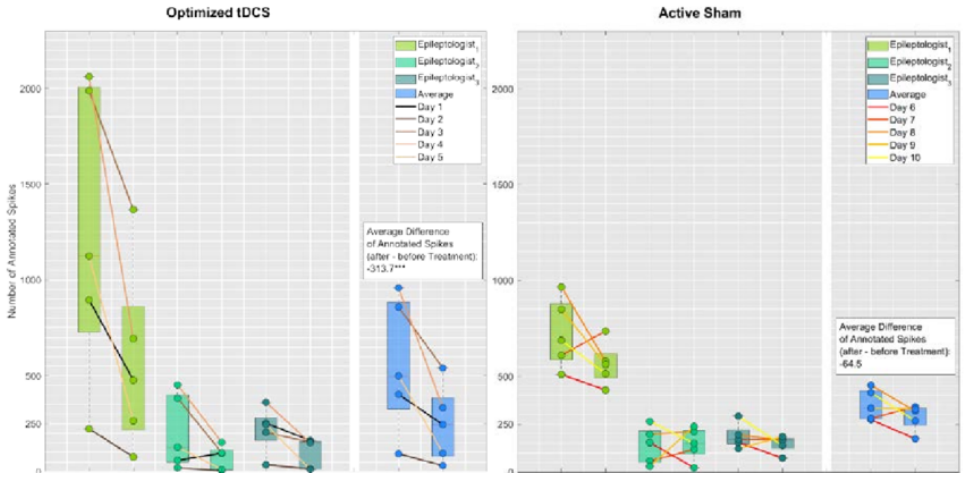


Figure 4.2: Interictal Epileptiform Discharges (IEDs) before and after Treatment. The points indicate the number of annotated spikes (marked IEDs) during 1 hour of EEG found by Epileptologist 1 (light green), 2 (medium green), 3 (dark green) and their average (blue) before (Pre) and after (Post) treatment. Boxplots show the median (central mark) and the 25th and 75th percentiles (box), and the whiskers extend to the most extreme data-points which are not considered to be outliers : This result was adapted directly from the trial [42]

This hypothesis, which was originally tested via EMEG [42], was also confirmed by our gPDC calculations on the simple EEG recordings. Below, follow some illustrations, showcasing this finding for all the different cases.

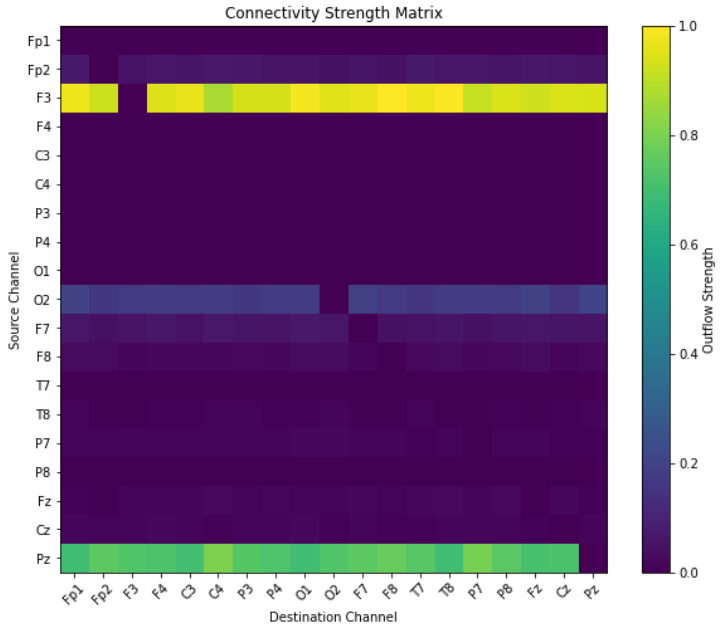


Figure 4.3: Outflow gPDC heatmap for an EEG recording before tDCS was applied, for a window containing an event. Both x-axis (destination channel) and y-axis (source channel) depict the channel names as they have been mapped by us, based on the head montage utilized in the trial. Each channel pairs' directed effective connectivity based on the gPDC measure is calculated in a range between 0 and 1. The strongest connections are highlighted in yellow and green and the weakest in blue.

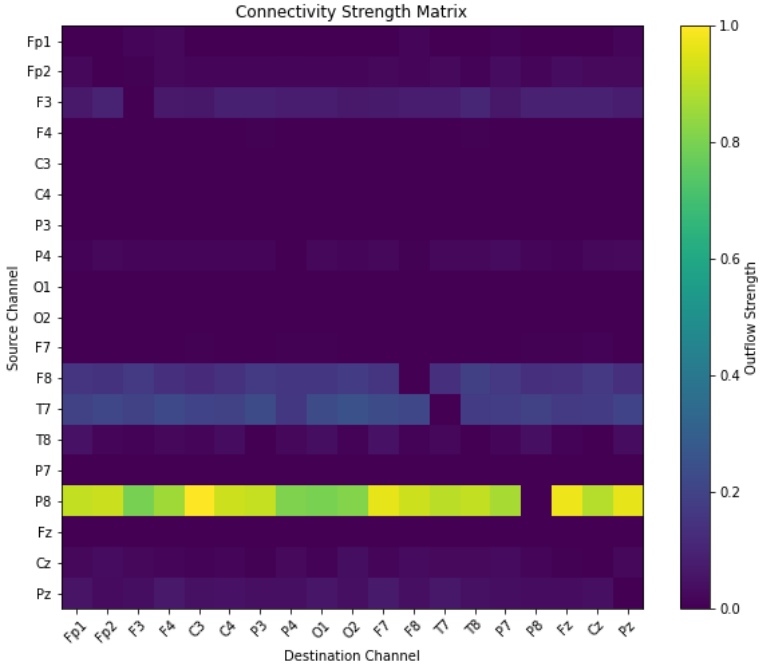


Figure 4.4: Outflow gPDC heatmap for an EEG recording after tDCS was applied, for a window containing an event.

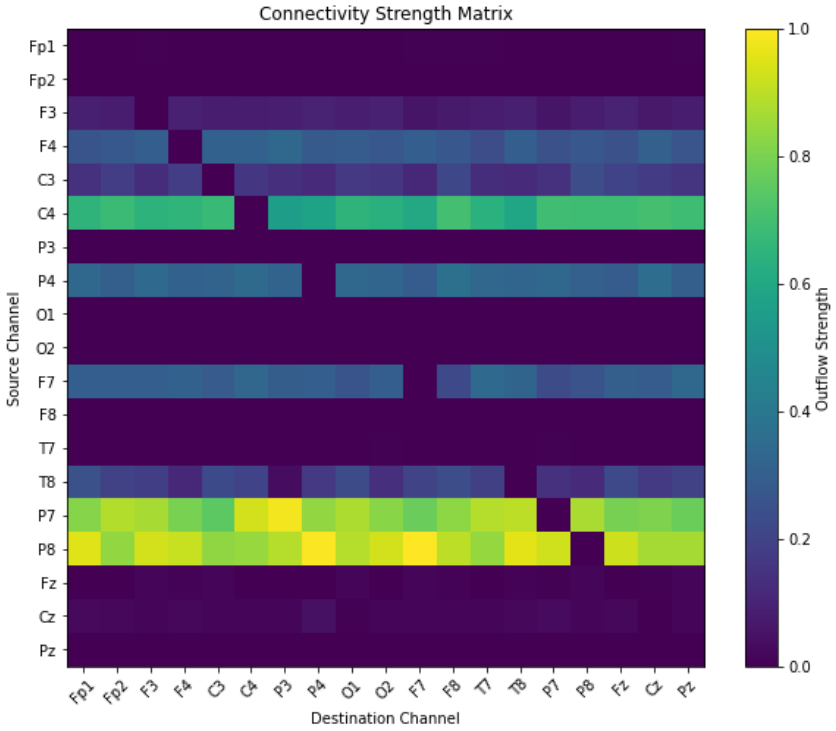


Figure 4.5: Outflow gPDC heatmap for an EEG recording before tDCS was applied, for a window not containing an event, meaning no seizure was marked in it.

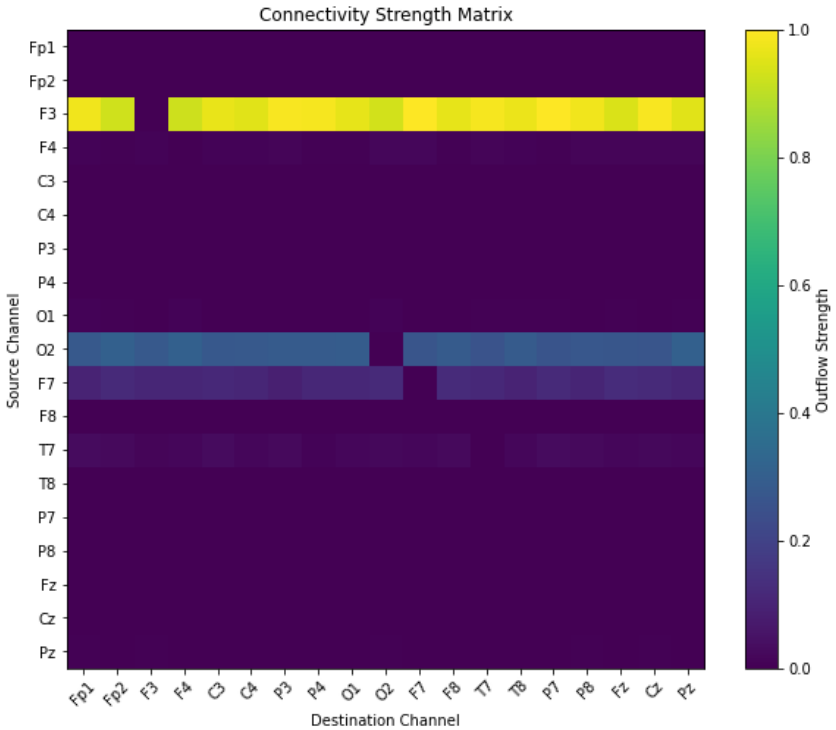


Figure 4.6: Outflow gPDC heatmap for an EEG recording before ActiSham was applied, for a window containing an event, meaning seizure was marked in it.

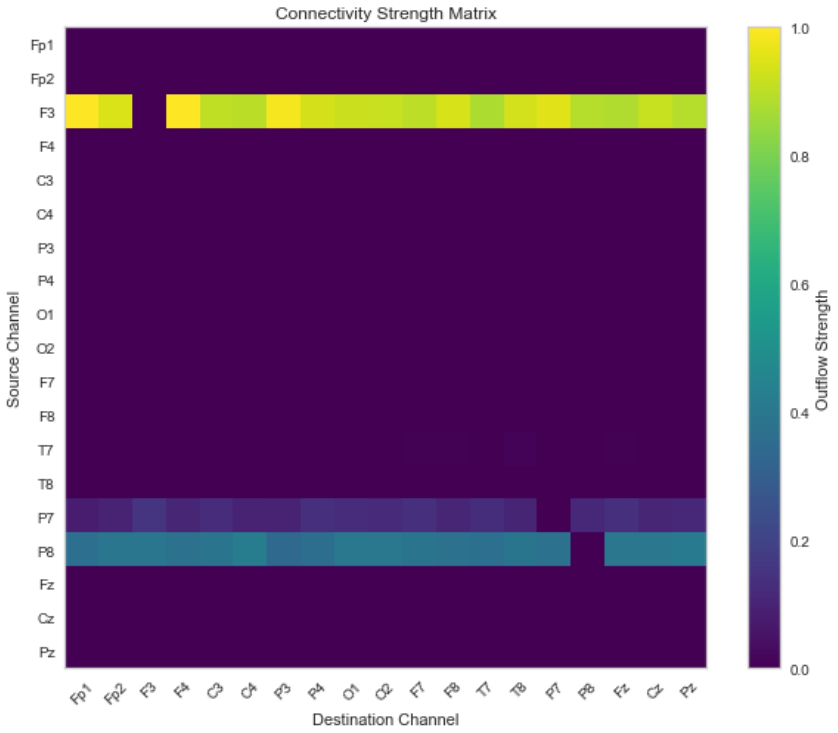


Figure 4.7: Outflow gPDC heatmap for an EEG recording after ActiSham was applied, for a window containing an event, meaning seizure was marked in it.

At this point, the difference between the gPDC outflow heatmaps depending on whether the patient was being recorded before or after stimulation or before or after sham, was clear. But to provide a more robust and comprehensive view of this analysis, containing information about all windows with events in all four cases (before stim, after stim, before sham, after sham), some histograms were calculated too, in order for the complete results per recording to be displayed. In these histograms, are depicted the frequencies with which a specific channel is highlighted as having the highest gPDC variance across its window's values. The windows included in this analysis, are only the ones that were characterized as windows containing seizures by the epileptologists. This comprehensive analysis was considered essential because, despite the fact that the channels which were located closely to the epileptogenic zone (e.g. F3) often exhibited the highest values of generalized partial directed coherence (gPDC), indicating strong outflow connectivity, adjacent channels also demonstrated significant connectivity values (e.g. T7,F7). Furthermore, the presence of artifacts within the signal, which are challenging to remove completely without also removing important signal information, necessitates a cautious interpretation of the data.

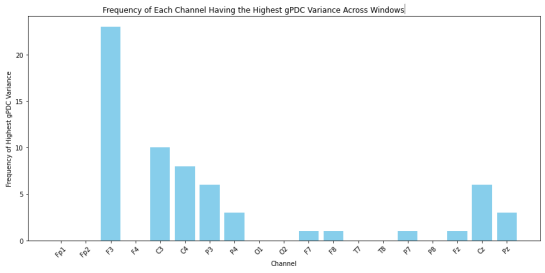


Figure 4.8: Best case scenario **pre-stimulation**: Strong connectivity outflow from channel F3, near the location where the epilepsy was localized.

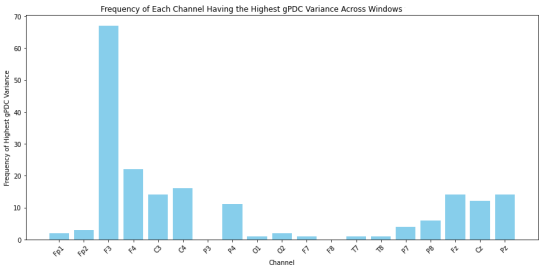


Figure 4.9: **Pre-stimulation** display of outflow connectivity for another positive result, on an EEG recording with a plethora of marked IEDs.

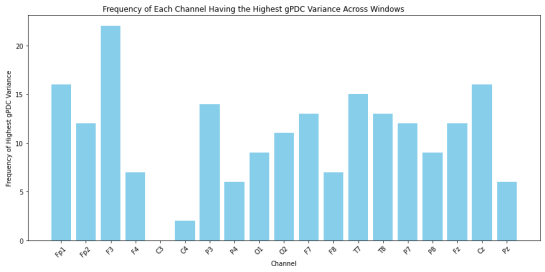


Figure 4.10: Another good case **pre-stimulation** display of outflow connectivity for a case where the hub node can still be identified as F3 but the neighbouring channels also showcase high gPDC values.

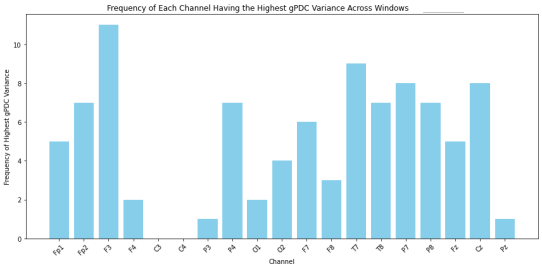


Figure 4.11: **Pre-sham** display of outflow connectivity for a case where the hub node can still be identified as F3 but the neighbouring channels also showcase high gPDC values.

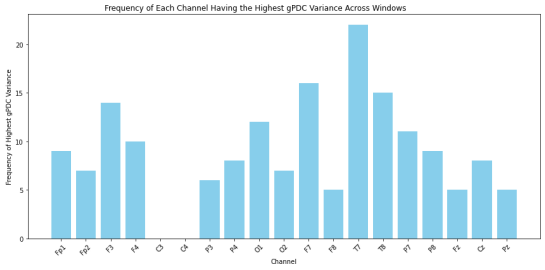


Figure 4.12: **After the stimulation**, the highest gPDC outflow connectivity is now sparse across other channels.

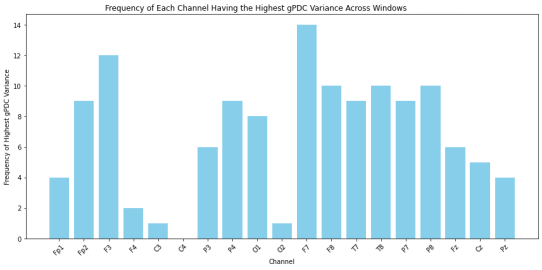


Figure 4.13: **After sham**: F3 still showcases high outflow connectivity, the plot shows high similarity to the pre-sham findings (Figure 4.11)

In conclusion, even though we did not isolate a single hub node in every scenario, the accuracy achieved in localizing its approximate position remains high. This outcome is encouraging and highlights the potential of our methodology. It suggests that with further refinement,

probably through even more enhanced filtering to more effectively remove noise and artifacts, though without causing severe damage to the signal, our approach could yield even more precise localization of the hub node. This promising result underscores the value of our data and the need for continued optimization of our preprocessing strategies in order for our pipeline to provide deeper insights into the connectivity patterns underlying epileptic seizures.

Finally, considering the fact that our goal was to showcase the results of effective connectivity on a brain, and Python's MNE library [44] provides us with scalp topographies for our EEG montage, the step that naturally followed was the creation of scalp topographies containing all the inflow and outflow gPDC information per window. These figures, display once again the results of gPDC in the frequency domain, showcasing the strength of connectivity resulted from the gPDC calculations.

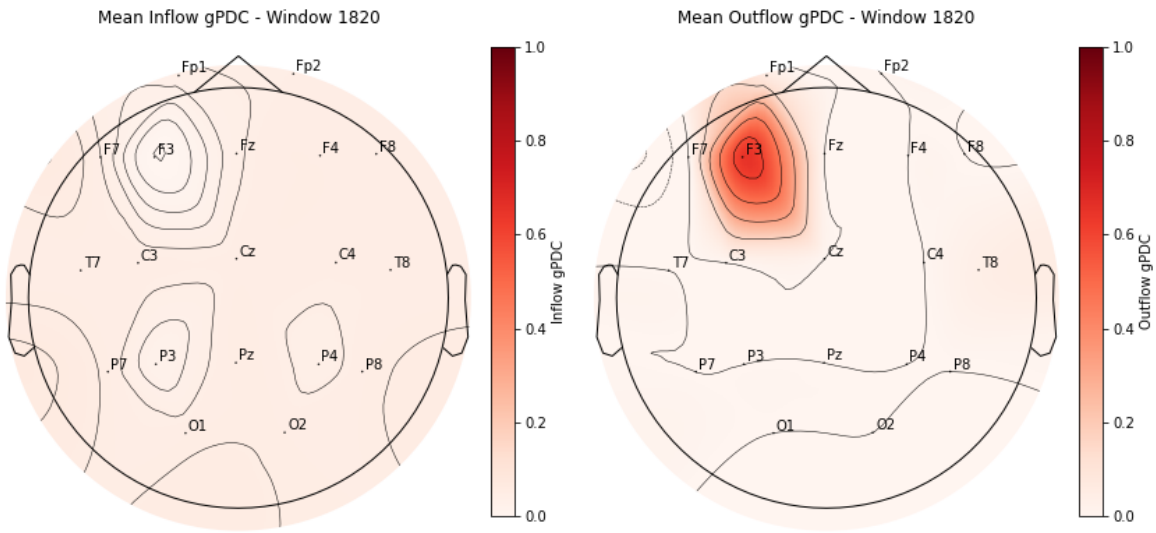


Figure 4.14: gPDC inflow and outflow connectivity in the frequency domain displayed on a scalp topography, highlighting where the highest frequency is found for a window containing an event

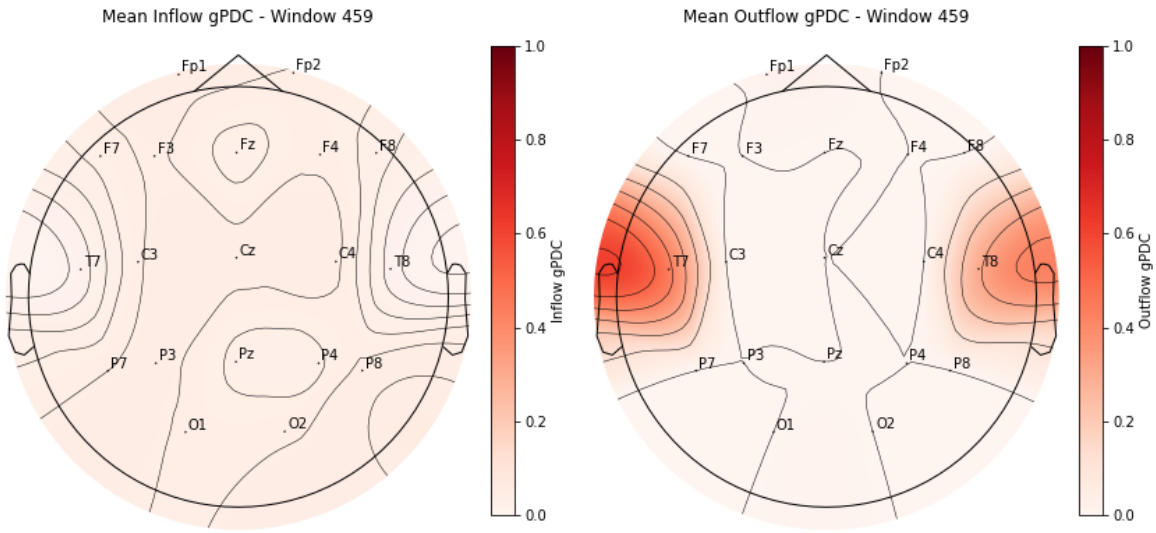


Figure 4.15: gPDC inflow and outflow connectivity in the frequency domain displayed on a scalp topography, highlighting where the highest frequency is found for a window not containing an event, we can clearly notice the spread of gPDC strength around more brain areas

After that, the effective connectivity analysis part of this work was considered completed. The question which then arose was how else we could further exploit all the connectivity information we had collected and how could we prove the impact of tDCS on the patient? Then, came the idea for the feature extraction and ML analysis. Although the results from feature extraction did not yield direct graphical outcomes, they represent an important milestone in the analytical process followed in this thesis, as they facilitated the transition towards the application of Machine Learning models.

4.2 Machine Learning: Results

Upon extracting the connectivity features and some statistical properties, such as skewness and kurtosis, from our dataset, this information was utilized as input for Support Vector Machines (SVMs) and Random Forest (RF) models. Initially, SVMs and RF models were selected due to their efficacy in binary classification tasks. The experimental design was once again segmented into four distinct phases: pre-stimulation, post-stimulation, pre-sham, and post-sham, mirroring the structure adopted in earlier stages of the research. The hypothesis based on which the addition of ML models was decided in order to strengthen the results of our analysis, is that

the pre-stimulation data are more easily separable - classifiable than the post-stimulation ones. Intuitively, due to the fact that the epileptic spikes marked before the stimulation process appear to showcase stronger connectivity, around the epileptogenic area, than the ones marked in the after-stimulation EEG recordings throughout the brain, due to the fact that tDCS was proven effective, the pre-stimulation seizures can be clearly classified as their correct class, 'spikes', but the post-stimulation data, can be easily confused with all kinds of normal neural activity appearing through connectivity patterns. Subsequently if this hypothesis holds, the patient's symptoms after the tDCS would have to be milder than the ones before the stimulation. One way to validate the aforementioned hypothesis with the help of ML classifiers, is to show that the accuracy of the pre-stimulation model is indeed higher than for the data after the stimulation process. On the same note, we would expect the classification of data before and after sham not to vary importantly in accuracy, due to the fact that the patient should not be subject to a true difference in the epileptic seizure's intensity. The outcomes of these models were notably promising, proving once more the robustness of the connectivity results and the true effect of tDCS on the patient's symptoms. The Random Forest model, in particular, demonstrated a very good performance, while the SVMs yielded satisfactory results, to a lesser extent. This disparity underscores the varying efficacies of these models in handling the given classification challenges.

The ML experiments and results were split once again in the four separate parts of the trial; pre-stimulation, post-stimulation, pre-sham and post-sham. The inputs to the ML models were the outputs of the feature extraction analysis which were also all normalized, which finally resulted in the following evaluation measures: global efficiency, local efficiency, outer degree, strength, flexibility index and also some statistical features per channel, namely skewness, kurtosis, mean and standard deviation.

Finally, both the Random Forest and SVM classifer were ran iteratively, for different seeds, 4 different times, one per trial part, and ten fold cross validation was performed. All the experiments were conducted on a similar number of data, in order for the results to be robust. The final average performance can be seen in Table 4.1

Models	Before Stimulation	After Stimulation	Before Sham	After Sham
RandomForest	0.76	**0.66**	0.78	0.74
SVM	0.68	**0.58**	0.71	0.70

Table 4.1: Accuracy for SVM and RandomForest classifiers during different stages of the trial

As shown in Table 4.1, the accuracy of both models after the stimulation is significantly smaller than the one pre stimulation, validating our initial hypothesis. Furthermore, the results before stimulation and before and after sham, lie in the same accuracy percentiles, as expected based on our initial intuition too. More detailed figures on the calculations of the accuracy results, as well as further details on the ten-fold cross validation performed, to prove the robustness of our results are displayed below.

4.3 Random Forest results - Before Stimulation

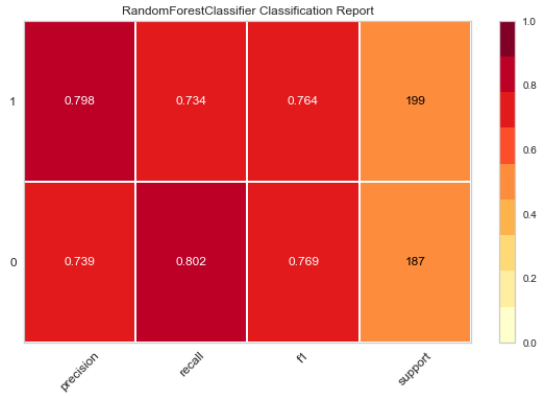


Figure 4.16: RandomForest's classification report for the connectivity features extracted from the EEG recordings before the stimulation

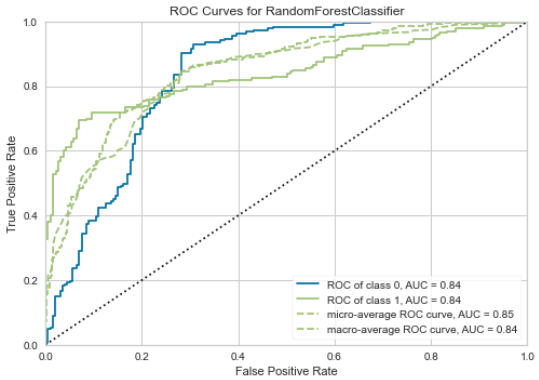


Figure 4.17: ROC curves for the RandomForest model before stimulation. These ROC curves feature true positive rate (TPR) on the Y axis, and false positive rate (FPR) on the X axis. This means that the top left corner of the plot is the "ideal" point - a FPR of zero, and a TPR of one. This is not very realistic, but it does mean that a larger Area Under the Curve (AUC) is usually better. The "steepness" of ROC curves is also important, since it is ideal to maximize the TPR while minimizing the FPR

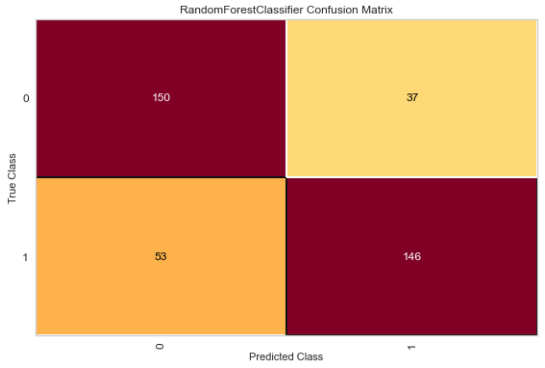


Figure 4.18: RandomForest's confusion matrix for the EEG recordings before the stimulation

```
Variance for each metric across all folds:
Accuracy       0.000626
Precision      0.000697
NPV            0.000680
Specificity    0.000791
F1 Score       0.000904
ROC AUC        0.000232
dtype: float64

Detailed results per fold:
   Accuracy  Precision       NPV  Specificity  F1 Score   ROC AUC
0  0.694805   0.716418  0.678161     0.756410  0.671329  0.775304
1  0.740260   0.757143  0.726190     0.782051  0.726027  0.794534
2  0.707792   0.724638  0.694118     0.756410  0.689655  0.783570
3  0.733766   0.761194  0.712644     0.794872  0.713287  0.791245
4  0.753247   0.771429  0.738095     0.794872  0.739726  0.773701
5  0.688312   0.705882  0.674419     0.743590  0.666667  0.792426
6  0.720779   0.736111  0.707317     0.753247  0.711409  0.807978
7  0.688312   0.730159  0.659341     0.779221  0.657143  0.759993
8  0.753247   0.782609  0.729412     0.805195  0.739726  0.809074
9  0.714286   0.714286  0.714286     0.714286  0.714286  0.788582
```

Figure 4.19: Figure representation of the ten-fold cross validation process followed to validate the robustness of our results.

4.4 SVM results - Before Stimulation

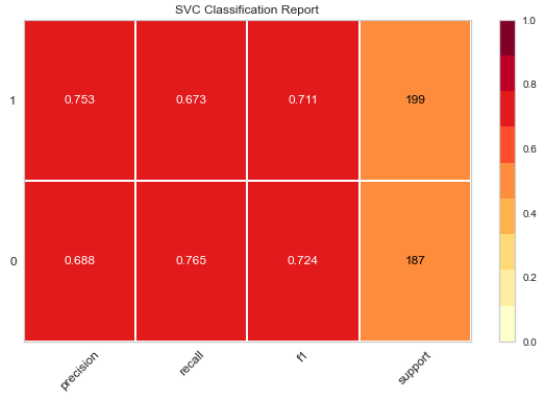


Figure 4.20: SVM's classification report for the EEG recordings before the stimulation

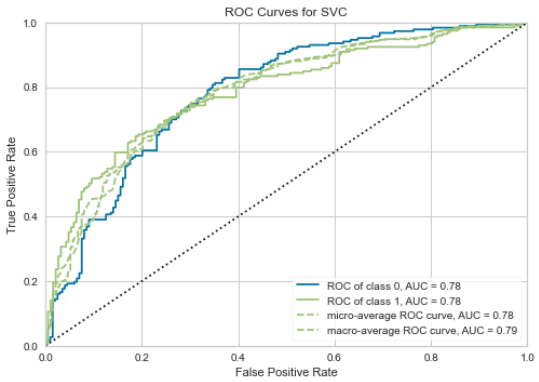


Figure 4.21: SVM's ROC before stimulation. It presents worse results than RandomForest

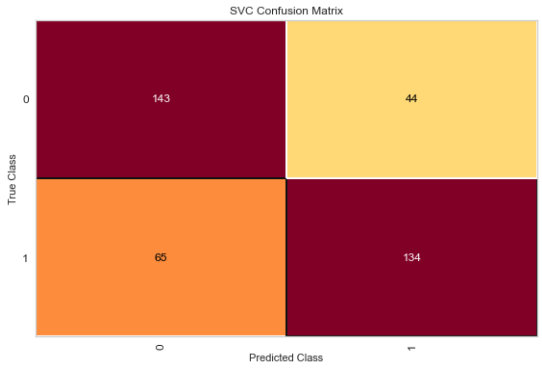


Figure 4.22: SVM's confusion matrix for the EEG recordings before the stimulation

```
Variance for each metric across all folds:
Accuracy       0.000542
Precision      0.000390
NPV            0.001331
Specificity    0.002532
F1 Score       0.000458
ROC AUC        0.000616
dtype: float64

Detailed results per fold:
   Accuracy  Precision       NPV  Specificity  F1 Score   ROC AUC
0  0.649351   0.619565  0.693548     0.551282  0.670571  0.701333
1  0.623377   0.581818  0.727273     0.410256  0.688172  0.675439
2  0.584416   0.561224  0.625000     0.448718  0.632184  0.659244
3  0.603896   0.582418  0.634921     0.512821  0.634731  0.634953
4  0.610390   0.578431  0.673077     0.448718  0.662921  0.688765
5  0.636364   0.604167  0.689655     0.512821  0.674419  0.673920
6  0.610390   0.582524  0.666667     0.441558  0.666667  0.688986
7  0.577922   0.562500  0.603448     0.454545  0.624277  0.624220
8  0.629870   0.606383  0.666667     0.519481  0.666667  0.650531
9  0.590909   0.566038  0.645833     0.402597  0.655738  0.677180
```

Figure 4.23: Figure representation of the ten-fold cross validation process followed to validate the robustness of our results.

4.5 Random Forest results - After Stimulation

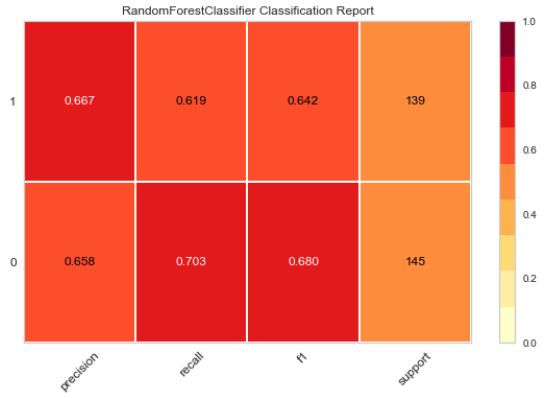


Figure 4.24: RandomForest's classification report for the connectivity features extracted from the EEG recordings after the stimulation

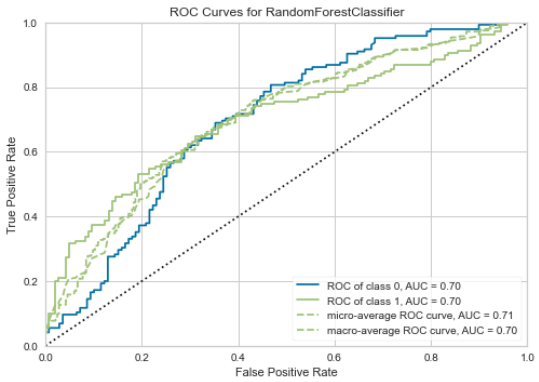


Figure 4.25: ROC curves for the RandomForest model after stimulation. In this figure it can be clearly seen that after the stimulation, the model's behavior yields much less accuerate results.

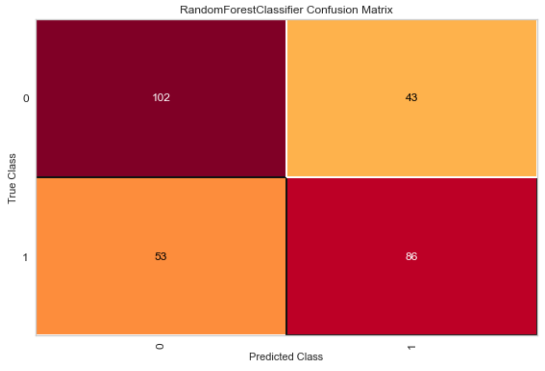


Figure 4.26: RandomForest's confusion matrix for the EEG recordings after the stimulation

```
Variance for each metric across all folds:
Accuracy        0.000970
Precision       0.001461
NPV             0.000793
Specificity     0.002625
F1 Score        0.001179
ROC AUC         0.001982
dtype: float64

Detailed results per fold:
   Accuracy  Precision       NPV  Specificity  F1 Score   ROC AUC
0  0.622807   0.640000  0.609375     0.684211  0.598131  0.679902
1  0.578947   0.595745  0.567164     0.666667  0.538462  0.636504
2  0.578947   0.568966  0.589286     0.568966  0.578947  0.628233
3  0.628319   0.645833  0.615385     0.701754  0.596154  0.685934
4  0.637168   0.653061  0.625000     0.701754  0.609524  0.712719
5  0.601770   0.603774  0.600000     0.631579  0.587156  0.633615
6  0.557522   0.560000  0.555556     0.614035  0.528302  0.555764
7  0.637168   0.631579  0.642857     0.631579  0.637168  0.655702
8  0.610619   0.630435  0.597015     0.701754  0.568627  0.612939
9  0.654867   0.680851  0.636364     0.736842  0.621359  0.673872
```

Figure 4.27: Figure representation of the ten-fold cross validation process followed to validate the robustness of our results.

4.6 SVM results - After Stimulation

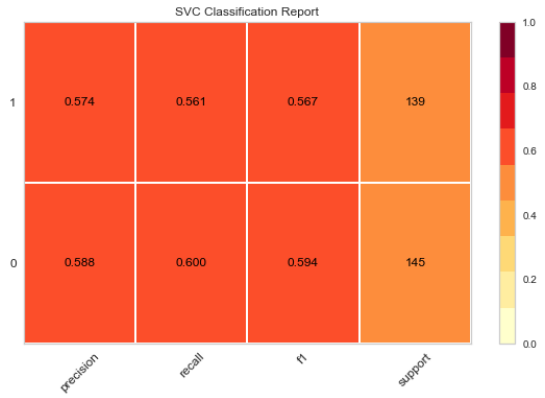


Figure 4.28: SVM's classification report for the EEG recordings after the stimulation

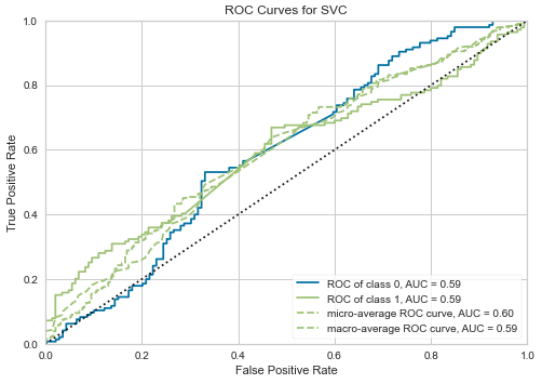


Figure 4.29: SVM's ROC after stimulation. The accuracy of the classification is clearly diminished and the data doesn't seem to be easily separable.

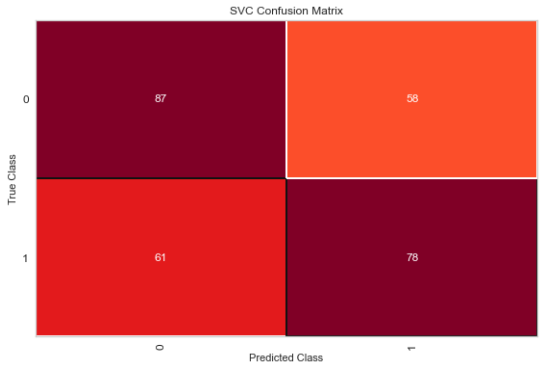
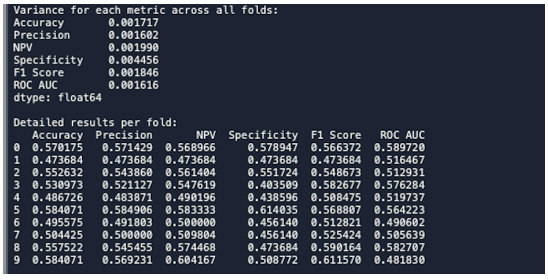


Figure 4.30: SVM's confusion matrix for the EEG recordings after the stimulation



```
Variance for each metric across all folds:
Accuracy       0.001717
Precision      0.001602
NPV            0.001990
Specificity    0.004456
F1 Score       0.001846
ROC AUC        0.001616
dtype: float64

Detailed results per fold:
   Accuracy  Precision       NPV  Specificity  F1 Score   ROC AUC
0  0.570175   0.571429  0.568966     0.578947  0.566372  0.589720
1  0.473684   0.473684  0.473684     0.473684  0.473684  0.516467
2  0.552632   0.543860  0.561404     0.551724  0.548673  0.512931
3  0.530973   0.521127  0.547619     0.403509  0.582677  0.576284
4  0.486726   0.483871  0.490196     0.438596  0.508475  0.519737
5  0.584071   0.584906  0.583333     0.614035  0.568807  0.564223
6  0.495575   0.491803  0.500000     0.456140  0.512821  0.490602
7  0.504425   0.500000  0.509804     0.456140  0.525424  0.505639
8  0.557522   0.545455  0.574468     0.473684  0.590164  0.582707
9  0.584071   0.569231  0.604167     0.508772  0.611570  0.481830
```

Figure 4.31: Figure representation of the ten-fold cross validation process followed to validate the robustness of our results.

4.7 Random Forest results - Before Sham

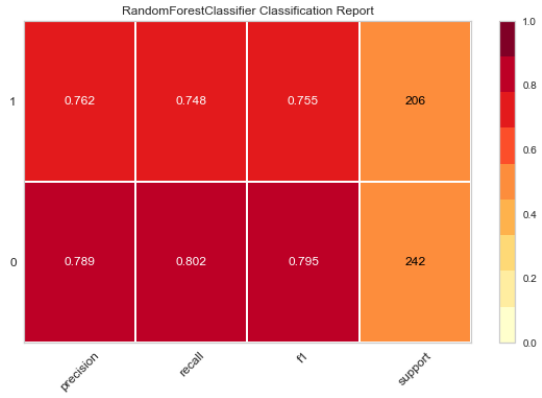


Figure 4.32: RandomForest's classification report for the connectivity features extracted from the EEG recordings before sham

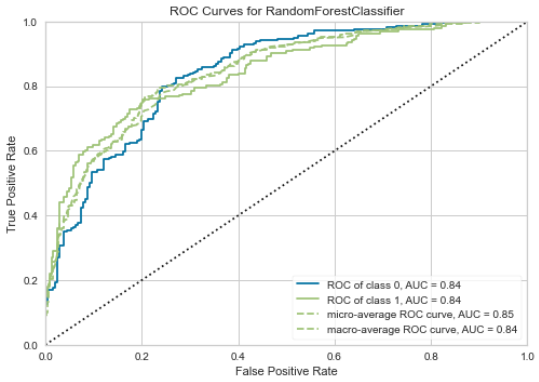


Figure 4.33: ROC curves for the RandomForest model before sham.

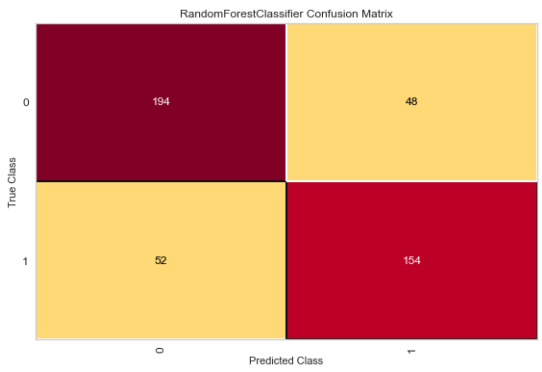


Figure 4.34: RandomForest's confusion matrix for the EEG recordings before sham

```
Variance for each metric across all folds:
Accuracy       0.000539
Precision      0.001460
NPV            0.001178
Specificity    0.003973
F1 Score       0.000694
ROC AUC        0.000354
dtype: float64

Detailed results per fold:
   Accuracy  Precision       NPV  Specificity  F1 Score   ROC AUC
0  0.770950   0.780220  0.761364     0.770115  0.775956  0.862944
1  0.748603   0.805195  0.705882     0.827586  0.733728  0.839830
2  0.726257   0.741573  0.711111     0.735632  0.729282  0.827899
3  0.737430   0.768293  0.711340     0.784091  0.728324  0.807942
4  0.715084   0.692308  0.746667     0.636364  0.738462  0.829733
5  0.765363   0.795181  0.739583     0.806818  0.758621  0.826611
6  0.715084   0.750000  0.686869     0.772727  0.701754  0.795392
7  0.748603   0.802632  0.708738     0.829545  0.730539  0.829233
8  0.770950   0.820513  0.732673     0.840909  0.757396  0.838911
9  0.775281   0.752475  0.805195     0.712644  0.791667  0.844954
```

Figure 4.35: Figure representation of the ten-fold cross validation process followed to validate the robustness of our results.

4.8 SVM results - Before Sham

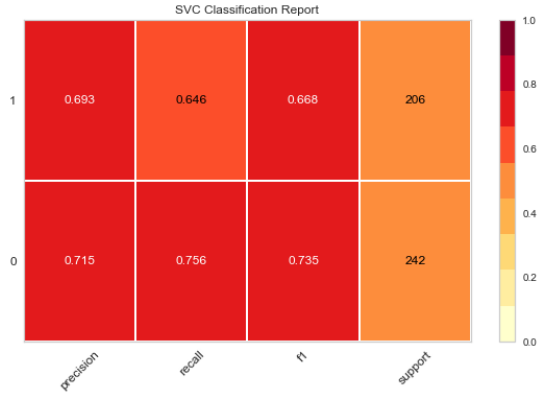


Figure 4.36: SVM's classification report for the EEG recordings before the stimulation

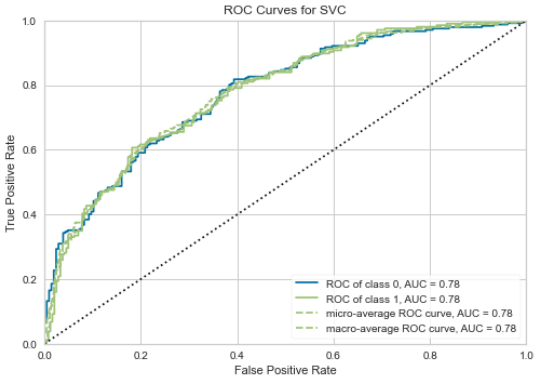


Figure 4.37: SVM's ROC before sham.

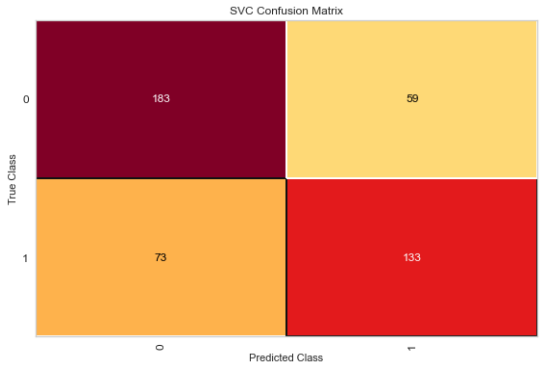
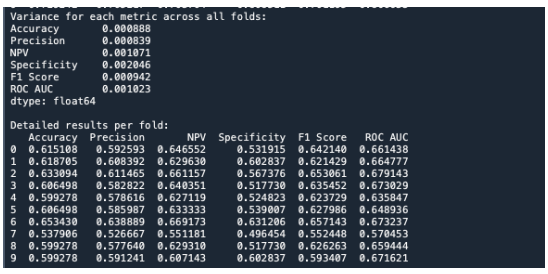


Figure 4.38: SVM's confusion matrix for the EEG recordings before sham



```
Variance for each metric across all folds:
Accuracy       0.000888
Precision      0.000839
NPV            0.001071
Specificity    0.002046
F1 Score       0.000942
ROC AUC        0.001023
dtype: float64

Detailed results per fold:
   Accuracy  Precision       NPV  Specificity  F1 Score   ROC AUC
0  0.615108   0.592593  0.646552     0.531915  0.642140  0.661438
1  0.618705   0.608392  0.629630     0.602837  0.621429  0.664777
2  0.633094   0.611465  0.661157     0.567376  0.653061  0.679143
3  0.606498   0.582822  0.640351     0.517730  0.635452  0.673029
4  0.599278   0.578616  0.627119     0.524823  0.623729  0.635847
5  0.606498   0.585987  0.633333     0.539007  0.627986  0.648936
6  0.653430   0.638889  0.669173     0.631206  0.657143  0.673237
7  0.537906   0.526667  0.551181     0.496454  0.552448  0.570453
8  0.599278   0.577640  0.629310     0.517730  0.626263  0.659444
9  0.599278   0.591241  0.607143     0.602837  0.593407  0.671621
```

Figure 4.39: Figure representation of the ten-fold cross validation process followed to validate the robustness of our results.

4.9 Random Forest results - After Sham

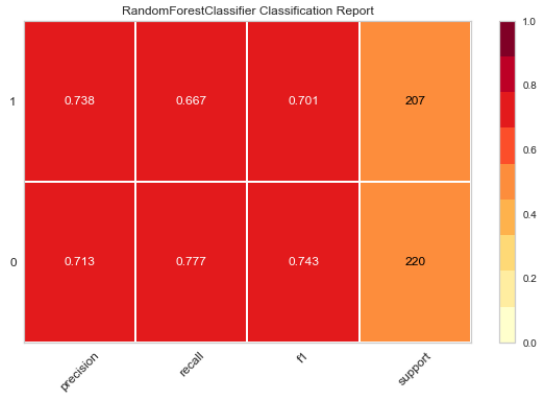


Figure 4.40: RandomForest's classification report for the connectivity features extracted from the EEG recordings after sham

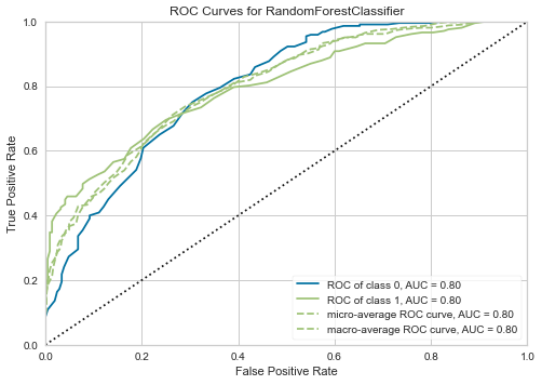


Figure 4.41: ROC curves for the RandomForest model after sham.

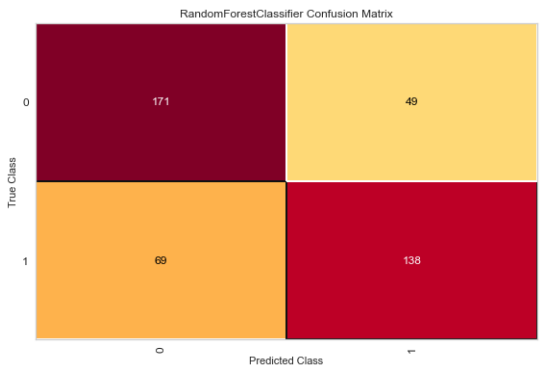


Figure 4.42: RandomForest's confusion matrix for the EEG recordings after sham

```
Variance for each metric across all folds:
Accuracy       0.002086
Precision      0.002513
NPV            0.002157
Specificity    0.002132
F1 Score       0.002791
ROC AUC        0.002276
dtype: float64

Detailed results per fold:
   Accuracy  Precision       NPV  Specificity  F1 Score   ROC AUC
0  0.719298   0.743590  0.698925     0.764706  0.707317  0.816347
1  0.748538   0.779221  0.723404     0.800000  0.736196  0.845349
2  0.684211   0.716216  0.659794     0.752941  0.662500  0.740971
3  0.736842   0.825397  0.685185     0.870588  0.697987  0.796785
4  0.777778   0.815789  0.747368     0.835294  0.765432  0.838167
5  0.766082   0.794872  0.741935     0.811765  0.756098  0.842066
6  0.652941   0.671053  0.638298     0.705882  0.633540  0.720415
7  0.747059   0.762500  0.733333     0.776471  0.739394  0.804637
8  0.664706   0.710145  0.633663     0.761905  0.632258  0.740864
9  0.776471   0.792683  0.761364     0.797619  0.773810  0.840047
```

Figure 4.43: Figure representation of the ten-fold cross validation process followed to validate the robustness of our results.

4.10 SVM results - After Sham

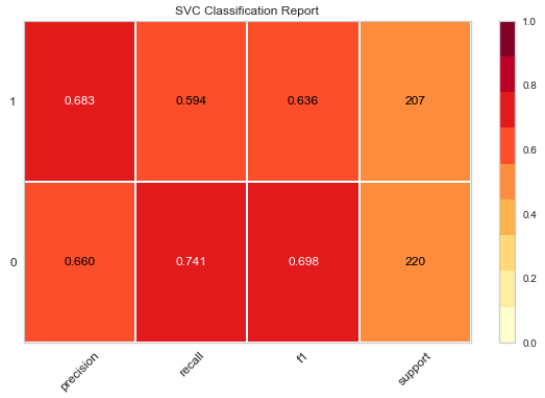


Figure 4.44: SVM's classification report for the EEG recordings after the stimulation

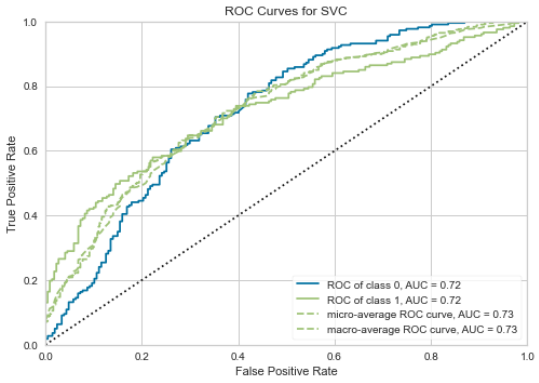


Figure 4.45: SVM's ROC after sham.

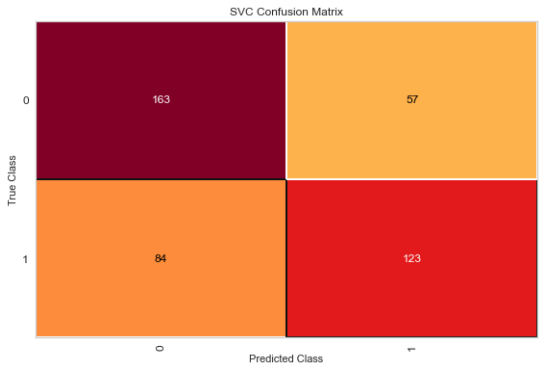


Figure 4.46: SVM's confusion matrix for the EEG recordings after sham

```
Variance for each metric across all folds:
Accuracy       0.001976
Precision      0.001743
NPV            0.002372
Specificity    0.004322
F1 Score       0.001749
ROC AUC        0.002495
dtype: float64

Detailed results per fold:
   Accuracy  Precision       NPV  Specificity  F1 Score   ROC AUC
0  0.608187   0.600000  0.618421     0.552941  0.629834  0.659918
1  0.596491   0.587629  0.608108     0.529412  0.622951  0.601505
2  0.543860   0.547619  0.540230     0.552941  0.541176  0.552668
3  0.532164   0.532609  0.531646     0.494118  0.550562  0.552257
4  0.578947   0.583333  0.574713     0.588235  0.576471  0.618741
5  0.602339   0.595745  0.610390     0.552941  0.622222  0.625718
6  0.617647   0.621951  0.613636     0.635294  0.610778  0.626298
7  0.500000   0.500000  0.500000     0.411765  0.540541  0.530242
8  0.535294   0.535354  0.535211     0.452381  0.572973  0.534192
9  0.635294   0.625000  0.648649     0.571429  0.659341  0.659884
```

Figure 4.47: Figure representation of the ten-fold cross validation process followed to validate the robustness of our results.

Chapter 5

Discussion

In the present work, the efficacy of dCMI-tDCS in reducing the seizure activity in patients with refractory focal epilepsy were studied and validated, utilizing exclusively EEG recordings. The main object of our research, was the development and implementation of a comprehensive, Python-based analytical pipeline which could provide interesting insights into neural activity during seizures and on the effect of tDCS. This pipeline, integrated the preprocessing of the data, connectivity analysis, feature extraction and the application of Machine Learning algorithms. The findings from our connectivity analysis showcased the changes in the brains' neural networks post-tDCS treatment, affirming the potential of non-invasive neurostimulation techniques in the management of epilepsy. These outcomes, were also validated through ML models, which once again highlighted the difference in the classification accuracy results before and after real stimulation and before and after sham, proving our hypothesis that tDCS reduced both the frequency and the strength of IEDs.

Furthermore, this research contributes to gaining a deeper understanding of the neural mechanisms underlying epilepsy, utilizing a non-invasive modality, as is EEG. This approach not only ensures patient comfort and safety but also gives us the opportunity to conduct extended hours of trials, without the potential hazards associated with other, invasive techniques. The ability to gather and analyze more extensive EEG datasets over longer periods, enhances the depth and reliability of our findings, offering a richer understanding of the effects of Transcranial Direct Current Stimulation (tDCS) on epilepsy. This non-invasive methodology, if further refined, could provide us with valuable insights into the long-term efficacy of tDCS but also the dynamic nature of epileptic seizures.

5.1 Limitations, Outlook and Future Work

Despite the strengths of our methodological approach, we acknowledge certain limitations that were encountered. Firstly, this study was performed on a single patient. Even though this allowed us to perform a very specific and deep study of the case, aiming for a truly personalized treatment approach, the evaluation of our results could benefit from being tested to more data, from similar cases of refractory focal epilepsy. Related work on the subject could offer new be interesting insights if applied to this pipeline, many studies have followed similar approaches to try and shed more light to the intricate connectivity brain patterns associated with epilepsy and other disorders of the human brain [59] [60] [61] [62]. Firstly, it would be beneficial to experiment with different multivariate connectivity methods [63], and then evaluate them with the ML models.

Additionally, exploring other machine learning algorithms or deep learning approaches could offer even greater insights into seizure prediction. If more data were to be gathered, the application of Neural Networks could be beneficial. Studies have already partially tackled this issue and yielded promising results [64] [65]. This could be particularly useful for understanding the complex dynamics of brain connectivity, where the influence of one region on another can be non-linear and dependent on the state of other regions. Furthermore, increasing the dataset could also help enhance the simple ML classifiers' levels of accuracy results, however, it's important to acknowledge that increasing the size of the dataset will also introduce challenges in terms of computational resources and processing time.

Finally, given that one of the significant advantages of this research is being based on the use of EEG, another interesting path that could be followed would be the integration of other non-invasive methodologies (e.g. electrocardiography (ECG)) to the same pipeline. This combined approach, could possibly provide valuable insights into the connectivity patterns of the human brain, as it could also potentially be utilized for patient monitoring and analysis in a broader way, opening new avenues for multidimensional research that could help perform a comprehensive exploration of the complex interactions between multiple organ systems, combining electroencephalogram (EEG) recordings, which capture the electrical activity of the brain, with maybe electrocardiograms (ECG) or other physiological signals, such as elec-

tromyograms (EMG).

Bibliography

[1] M. Antonakakis, F. Kaiser, S. Rampp, *et al.*, "Targeted and optimized multi-channel transcranial direct current stimulation for focal epilepsy: An n-of-1 trial," English, *Brain Stimulation*, vol. Ahead of print, S1935-861X(24)00025–1, 2024, ISSN: 1876-4754 (Electronic), 1876-4754 (Linking). DOI: 10.1016/j.brs.2024.02.010. [Online]. Available: https://doi.org/10.1016/j.brs.2024.02.010.

[2] L. Thau, V. Reddy, and P. Singh, *Anatomy, central nervous system*, https://www.ncbi.nlm.nih.gov/books/NBK542179/, Updated 2022 Oct 10. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-., Treasure Island (FL), 2022.

[3] K. A. Maldonado and K. Alsayouri, *Physiology, Brain*, English. Treasure Island (FL): StatPearls Publishing, Jan. 2024, PMID: 31869182.

[4] Johns Hopkins Medicine, *Anatomy of the brain*, https://www.hopkinsmedicine.org/health/conditions-and-diseases/anatomy-of-the-brain, Accessed: 2024-02-28, 2024.

[5] Khan Academy, *Overview of neuron structure and function*, https://www.khanacademy.org/science/biology/human-biology/neuron-nervous-system/a/overview-of-neuron-structure-and-function, Accessed: 2024-02-28, 2024.

[6] Johns Hopkins Medicine, *Generalized seizures*, https://www.hopkinsmedicine.org/health/conditions-and-diseases/epilepsy/generalized-seizures, Accessed: February, 2024.

[7] M. Bear, B. Connors, and M. A. Paradiso, *Neuroscience: exploring the brain, enhanced edition: exploring the brain*. Jones & Bartlett Learning, 2020.

[8] D. T. Stuss and R. T. Knight, *Principles of frontal lobe function*. Oxford University Press, USA, 2013.

[9] V. Rajmohan and E. Mohandas, "The limbic system," English, *Indian Journal of Psychiatry*, vol. 49, no. 2, pp. 132–139, Apr. 2007, Electronic ISSN: 1998-3794; Print ISSN: 0019-5545; Linking ISSN: 0019-5545, ISSN: 1998-3794. DOI: 10.4103/0019-5545.33264. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2917081.

[10] S. Al-Aqeel, R. Alotaiwi, and B. Albugami, "Patient preferences for epilepsy treatment: A systematic review of discrete choice experimental studies," *Health Economics Review*, vol. 13, no. 1, p. 17, 2023, ISSN: 2191-1991. DOI: 10.1186/s13561-023-00431-0. [Online]. Available: https://doi.org/10.1186/s13561-023-00431-0.

[11] R. N. L. Lamptey, B. Chaulagain, R. Trivedi, A. Gothwal, B. Layek, and J. Singh, "A review of the common neurodegenerative disorders: Current therapeutic approaches and the potential role of nanotherapeutics," English, *International Journal of Molecular Sciences*, vol. 23, no. 3, p. 1851, Feb. 6, 2022, ISSN: 1422-0067. DOI: 10.3390/ijms23031851. [Online]. Available: https://www.mdpi.com/1422-0067/23/3/1851.

[12] E. H. Reynolds, "History of epilepsy," in *Introduction to Epilepsy*, G. Alarcón and A. Valentín, Eds. Cambridge University Press, 2012, pp. 1–5.

[13] R. S. Fisher, "Redefining epilepsy," English, *Current Opinion in Neurology*, vol. 28, no. 2, pp. 130–135, 2015, ISSN: 1473-6551 (Electronic), 1350-7540 (Linking). DOI: 10.1097/WCO.0000000000000174. [Online]. Available: https://doi.org/10.1097/WCO.0000000000000174.

[14] O. Temkin, *The Falling Sickness: A History of Epilepsy from the Greeks to the Beginnings of Modern Neurology* (Publications : 1st ser., Monographs). Johns Hopkins University Press, 1971, ISBN: 9780801812118. [Online]. Available: https://books.google.gr/books?id=SoFtZg0DhI8C.

[15] H. Berger, "Über das elektrenkephalogramm des menschen," *Archiv für Psychiatrie und Nervenkrankheiten*, vol. 87, pp. 527–570, 1 1929, ISSN: 1433-8491. DOI: 10.1007/BF01797193. [Online]. Available: https://doi.org/10.1007/BF01797193.

[16] J. Engel Jr, "Report of the ilae classification core group," *Epilepsia*, vol. 47, no. 9, pp. 1558–1568, 2006.

[17] R. S. Fisher, J. H. Cross, C. D'Souza, *et al.*, "Instruction manual for the ilae 2017 operational classification of seizure types," *Epilepsia*, vol. 58, no. 4, pp. 531–542, 2017. DOI: https://doi.org/10.1111/epi.13671. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/epi.13671. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/epi.13671.

[18] I. E. Scheffer, S. Berkovic, G. Capovilla, *et al.*, "Ilae classification of the epilepsies: Position paper of the ilae commission for classification and terminology," English, *Epilepsia*, vol. 58, no. 4, pp. 512–521, 2017. DOI: 10.1111/epi.13709.

[19] M. A. Kramer and S. S. Cash, "Epilepsy as a disorder of cortical network organization," English, *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, vol. 18, no. 4, pp. 360–372, 2012, ISSN: 1089-4098. DOI: 10.1177/1073858411422754.

[20] S. Lagarde, C.-G. Bénar, F. Wendling, and F. Bartolomei, "Interictal functional connectivity in focal refractory epilepsies investigated by intracranial eeg," English, *Brain Connectivity*, vol. 12, no. 10, pp. 850–869, 2022, ISSN: 2158-0022. DOI: 10.1089/brain.2021.0190.

[21] K. M. Gunnarsdottir, A. Li, R. J. Smith, *et al.*, "Source-sink connectivity: A novel interictal eeg marker for seizure localization," English, *Brain : a journal of neurology*, vol. 145, no. 11, pp. 3901–3915, 2022, ISSN: 1460-2156. DOI: 10.1093/brain/awac300.

[22] Institute for Quality and Efficiency in Health Care (IQWiG), *What happens during an electroencephalogram (eeg)?* https://www.ncbi.nlm.nih.gov/books/NBK563104/, Available from: National Center for Biotechnology Information, Cologne, Germany: In-

stitute for Quality and Efficiency in Health Care (IQWiG), 2018. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK563104/.

[23] C. H. Wolters, M. Antonakakis, F. Kaiser, *et al.*, "Combined eeg/meg and optimized transcranial direct current stimulation for non-invasive diagnosis and therapy of focal epilepsy," in *IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI'22) jointly organized with the 17th IEEE-EMBS International Conference on Wearable and Implantable Body Sensor Networks (BSN'22)*, Ioannina, Greece, Sep. 27, 2022.

[24] D. Yang, Q. Wang, C. Xu, *et al.*, "Transcranial direct current stimulation reduces seizure frequency in patients with refractory focal epilepsy: A randomized, double-blind, sham-controlled, and three-arm parallel multicenter study," *Brain Stimulation*, vol. 12, no. 5, pp. 1297–1305, 2019. DOI: 10.1016/j.brs.2019.09.006. [Online]. Available: https://doi.org/10.1016/j.brs.2019.09.006.

[25] E. Kaufmann, M. Hordt, M. Lauseker, U. Palm, and S. Noachtar, "Acute effects of spaced cathodal transcranial direct current stimulation in drug resistant focal epilepsies," *Clinical Neurophysiology*, vol. 132, pp. 1444–1451, 2021. DOI: 10.1016/j.clinph.2021.03.048. [Online]. Available: https://doi.org/10.1016/j.clinph.2021.03.048.

[26] P. Sudbrack-Oliveira, M. Z. Barbosa, S. Thome-Souza, *et al.*, "Transcranial direct current stimulation (tdcs) in the management of epilepsy: A systematic review," *Seizure*, Feb. 2021. DOI: 10.1016/j.seizure.2021.01.020. [Online]. Available: https://doi.org/10.1016/j.seizure.2021.01.020.

[27] E. Bullmore and O. Sporns, "Complex brain networks: Graph theoretical analysis of structural and functional systems," English, *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009, ISSN: 1471-0048 (Electronic), 1471-003X (Linking). DOI: 10.1038/nrn2575. [Online]. Available: https://doi.org/10.1038/nrn2575.

[28] J. Cao, Y. Zhao, X. Shan, *et al.*, "Brain functional and effective connectivity based on electroencephalography recordings: A review," English, *Human Brain Mapping*, vol. 43, no. 2, pp. 860–879, 2022, ISSN: 1097-0193 (Electronic), 1065-9471 (Print), 1065-9471

(Linking). DOI: 10.1002/hbm.25683. [Online]. Available: https://doi.org/10.1002/hbm.25683.

[29] V. Sakkalis, "Review of advanced techniques for the estimation of brain connectivity measured with eeg/meg," English, *Computers in Biology and Medicine*, vol. 41, no. 12, pp. 1110–1117, 2011, ISSN: 1879-0534 (Electronic), 0010-4825 (Linking). DOI: 10.1016/j.compbiomed.2011.06.020. [Online]. Available: https://doi.org/10.1016/j.compbiomed.2011.06.020.

[30] L. Baccala and K. Sameshima, "Partial directed coherence: A new concept in neural structure determination," *Biological Cybernetics*, vol. 84, pp. 463–474, May 2001. DOI: 10.1007/PL00007990.

[31] *Support vector machines (svm)*, https://towardsdatascience.com/support-vector-machines-svm-c9ef22815589, 2023.

[32] R. Boswell, *Introduction to support vector machines*, https://home.work.caltech.edu/~boswell/IntroToSVM.pdf.

[33] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[34] F. Kaiser, M. Antonakakis, S. Rampp, *et al.*, "Combined eeg/meg and optimized transcranial direct current stimulation,"

[35] F. Kaiser, M. Antonakakis, S. Rampp, *et al.*, *Combined eeg/meg and optimized transcranial direct current stimulation for non-invasive diagnosis and therapy of focal epilepsy*, Available online at https://www.sci.utah.edu/~wolters/PaperWolters/2022/PosterKaiser.pdf, 2022.

[36] M. Langdon-Down and W. R. Brain, "Time of day in relation to convulsions in epilepsy," *The Lancet*, vol. 213, no. 5516, pp. 1029–1032, 1929. DOI: 10.1016/S0140-6736(00)79288-9. [Online]. Available: https://doi.org/10.1016/S0140-6736(00)79288-9.

[37] D. C. Spencer, F. T. Sun, S. N. Brown, *et al.*, "Circadian and ultradian patterns of epileptiform discharges differ by seizure-onset location during long-term ambulatory intracranial monitoring," English, *Epilepsia*, vol. 57, no. 9, pp. 1495–1502, 2016, ISSN: 1528-1167 (Electronic), 0013-9580 (Linking). DOI: 10.1111/epi.13455. [Online]. Available: https://doi.org/10.1111/epi.13455.

[38] Neuroelectrics, *Starstim*, https://www.neuroelectrics.com/solutions/starstim, Accessed: 2024-02-28, 2024.

[39] M. Antonakakis, "The effect of experimental and modeling parameters on combined eeg/meg source analysis and transcranial electric stimulation optimization of somatosensory and epilepsy activity," Ph.D. dissertation, Technische Universität Ilmenau, 2021.

[40] A. Khan, M. Antonakakis, N. Vogenauer, *et al.*, "Constrained maximum intensity optimized multi-electrode tdcs targeting of human somatosensory network," in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 5894–5897. DOI: 10.1109/EMBC.2019.8857253.

[41] A. Khan, M. Antonakakis, N. Vogenauer, J. Haueisen, and C. H. Wolters, "Individually optimized multi-channel tdcs for targeting somatosensory cortex," *Clinical Neurophysiology*, vol. 134, pp. 9–26, 2022, ISSN: 1388-2457. DOI: https://doi.org/10.1016/j.clinph.2021.10.016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1388245721007938.

[42] M. Antonakakis, F. Kaiser, S. Rampp, *et al.*, "Targeted and optimized multi-channel transcranial direct current stimulation for focal epilepsy: An n-of-1 trial," Sep. 2023. DOI: 10.1101/2023.09.05.23295060.

[43] M. Scherg and P. Berg, *BESA®| Brain Electrical Source Analysis*, https://www.besa.de, Accessed: date-of-access, 2018.

[44] A. Gramfort, M. Luessi, E. Larson, *et al.*, "MEG and EEG data analysis with MNE-Python," *Frontiers in Neuroscience*, vol. 7, no. 267, pp. 1–13, 2013. DOI: 10.3389/fnins.2013.00267.

[45] *Psychology*. New York: Worth Publishers, 1995, p. 115.

[46] Neural Data Science. "Erp filtering." Accessed: 2024-02-20. (2023), [Online]. Available: https://neuraldatascience.io/7-eeg/erp_filtering.html.

[47] P. Abhang, B. Gawali, and S. Mehrotra, "Technological basics of eeg recording and operation of apparatus," in Dec. 2016, pp. 19–50, ISBN: 9780128044902. DOI: 10.1016/B978-0-12-804490-2.00002-6.

[48] M. Chaumon, D. V. Bishop, and N. A. Busch, "A practical guide to the selection of independent components of the electroencephalogram for artifact correction," *Journal of Neuroscience Methods*, vol. 250, pp. 47–63, 2015, Cutting-edge EEG Methods, ISSN: 0165-0270. DOI: https://doi.org/10.1016/j.jneumeth.2015.02.025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0165027015000928.

[49] L. A. Baccala, K. Sameshima, and D. Takahashi, "Generalized partial directed coherence," in *2007 15th International Conference on Digital Signal Processing*, 2007, pp. 163–166. DOI: 10.1109/ICDSP.2007.4288544.

[50] G. Wang, Z. Sun, R. Tao, K. Li, G. Bao, and X. Yan, "Epileptic seizure detection based on partial directed coherence analysis," English, *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 3, pp. 873–879, May 2016, Epub 2015 Apr 17, ISSN: 2168-2208. DOI: 10.1109/JBHI.2015.2424074.

[51] E. Denovellis, M. Myroshnychenko, M. Sarmashghi, and E. Stephen, "Spectral connectivity: A python package for computing multitaper spectral estimates and frequency-domain brain connectivity measures on the cpu and gpu," *Journal of Open Source Software (JOSS)*, vol. 7, p. 4840, 2022. DOI: 10.21105/joss.04840.

[52] F. Vecchio, F. Miraglia, and P. Maria Rossini, "Connectome: Graph theory application in functional brain network architecture," *Clinical Neurophysiology Practice*, vol. 2, pp. 206–213, 2017, ISSN: 2467-981X. DOI: https://doi.org/10.1016/j.cnp.2017.09.003. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2467981X17300276.

[53] M. Kaminski, M. Ding, W. A. Truccolo, and S. L. Bressler, "Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance," *Biological Cybernetics*, vol. 85, pp. 145–157, 2001.

[54] E. Siggiridou, C. Koutlis, A. Tsimpiris, and D. Kugiumtzis, "Evaluation of granger causality measures for constructing networks from multivariate time series," *Entropy*, vol. 21, no. 11, 2019, ISSN: 1099-4300. DOI: 10.3390/e21111080. [Online]. Available: https://www.mdpi.com/1099-4300/21/11/1080.

[55] T. Schreiber and A. Schmitz, "Improved surrogate data for nonlinearity tests," *Phys. Rev. Lett.*, vol. 77, pp. 635–638, 4 Jul. 1996. DOI: 10.1103/PhysRevLett.77.635. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.77.635.

[56] T. B. Team, *Connectivity: Tutorial*, Accessed: 2024-02-03. [Online]. Available: https://neuroimage.usc.edu/brainstorm/Tutorials/Connectivity/#Thresholding_of_connectivity_estimates.

[57] A. D. Marimpis, S. I. Dimitriadis, and R. Goebel, "Dyconnmap: Dynamic connectome mapping-a neuroimaging python module," *Human Brain Mapping*, vol. 42, no. 15, pp. 4909–4939, 2021, PMID: 34250674. DOI: 10.1002/hbm.25589. [Online]. Available: https://doi.org/10.1002/hbm.25589.

[58] K. M. Almustafa, "Classification of epileptic seizure dataset using different machine learning algorithms," *Informatics in Medicine Unlocked*, vol. 21, p. 100 444, 2020, ISSN: 2352-9148. DOI: https://doi.org/10.1016/j.imu.2020.100444. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352914820305943.

[59] S. Rijal, L. Corona, M. S. Perry, *et al.*, "Functional connectivity discriminates epileptogenic states and predicts surgical outcome in children with drug resistant epilepsy," English, *Scientific reports*, vol. 13, no. 1, p. 9622, 2023. DOI: 10.1038/s41598-023-36551-0.

[60] D. J. Englot, L. B. Hinkley, N. S. Kort, *et al.*, "Global and regional functional connectivity maps of neural oscillations in focal epilepsy," *Brain*, vol. 138, no. 8, pp. 2249–2262, May 2015, ISSN: 0006-8950. DOI: 10.1093/brain/awv130. eprint: https://academic.

oup.com/brain/article-pdf/138/8/2249/11141611/awv130.pdf. [Online]. Available: https://doi.org/10.1093/brain/awv130.

[61] A. Karasmanoglou, M. Antonakakis, and M. Zervakis, "Unsupervised detection of seizure-related dynamic alterations with autoencoder-derived deep features," in *2023 IEEE 23rd International Conference on Bioinformatics and Bioengineering (BIBE)*, Los Alamitos, CA, USA: IEEE Computer Society, Dec. 2023, pp. 195–200. DOI: 10.1109/BIBE60311.2023.00038. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/BIBE60311.2023.00038.

[62] M. Vázquez, A. Maghsoudi, and I. Mariño, "An interpretable machine learning method for the detection of schizophrenia using eeg signals," *Frontiers in Systems Neuroscience*, vol. 15, Apr. 2021. DOI: 10.3389/fnsys.2021.652662.

[63] K. Sameshima and L. A. Baccala, *Methods in brain connectivity inference through multivariate time series analysis*. CRC press, 2014.

[64] H. Daoud and M. A. Bayoumi, "Efficient epileptic seizure prediction based on deep learning," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 5, pp. 804–813, 2019. DOI: 10.1109/TBCAS.2019.2929053.

[65] Z. Li, K. Hwang, K. Li, J. Wu, and T. Ji, "Graph-generative neural network for eeg-based epileptic seizure detection via discovery of dynamic brain functional connectivity," *Scientific Reports*, vol. 12, no. 1, p. 18 998, 2022, ISSN: 2045-2322. DOI: 10.1038/s41598-022-23656-1. [Online]. Available: https://doi.org/10.1038/s41598-022-23656-1.