

**TECHNICAL UNIVERSITY OF CRETE, GREECE
SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING**

Compressed Sensing Methods for Analysis of Epileptic Seizures



Tsekos Nikolaos

Thesis Committee
Professor Michalis Zervakis (ECE)
Professor George Karystinos (ECE)
Collaborating Researcher at Foundation for Research and Technology
George Giannakakis (FORTH)

Chania, February 2020

Πολυτεχνείο Κρήτης, Ελλάδα
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών (Η.Μ.Μ.Υ.)

Μέθοδος Ανάλυσης Επιληπτικών Κρίσεων με Μεθόδους Μειωμένης Ευαισθησίας



Τσέκος Νικόλαος

Επιτροπή Διπλωματικής Εργασίας
Καθηγητής Μιχάλης Ζερβάκης (Η.Μ.Μ.Υ.)
Καθηγητής Γεώργιος Καρυστινός (Η.Μ.Μ.Υ.)
Συνεργαζόμενος ερευνητής στο Ίδρυμα Τεχνολογίας και Έρευνας
Γεώργιος Γιαννακάκης (Ι.Τ.Ε.)

Χανιά, Φεβρουάριος 2020

Abstract

This thesis investigates the preservation of Electroencephalogram's (EEG) focal epileptic seizure's information after compression and reconstruction of the signal through the Compressed Sensing (CS) technique and the effectiveness of Common Spatial Patterns (CSP) analysis of EEG signals on the automatic detection of focal epileptic seizures. Epilepsy is a neurological disorder characterized by an enduring predisposition to generate epileptic seizures with great neurological, cognitive, psychological and social consequences. According to the World Health Organization (WHO), it is estimated that in 2019, epilepsy affects around 50 million people worldwide which is quite common in childhood. In 2017, the prevalence and incidence of epilepsy are estimated to be 6.38 and 0.67 per 1000 persons respectively. In this thesis, we use the Discrete Cosine Transform (DCT) to have a sparse representation of the information, in order to be able to apply the CS technique. We further reduce the power of the EEG in order to have a sparser signal. After compressing the signals, we use the Basis Pursuit algorithm to reconstruct the sparse DCT signal and then the inverse Discrete Cosine transform to return to the time domain. Then we apply the Fourier transform and the Approximate Entropy to check the preservation of the original information of the seizure. CSP analysis has been frequently used in literature for multichannel EEG signal separation between -mainly- two states. In the present thesis, the EEG recordings from 10 subjects aged 6.8 ± 5.9 years, including 63 seizures, were analyzed with respect to seizure detection and discrimination between inter-ictal and ictal signal periods. Machine learning techniques of feature selection and classification were used in the analysis.

Περίληψη

Η διπλωματική εργασία διερευνά τη διατήρηση των δεδομένων των εστιακών επιληπτικών κρίσεων του Ηλεκτροεγκεφαλογράφηματος (ΗΕΓ) μετά από συμπίεση και ανακατασκευή του σήματος μέσω της τεχνικής συμπιεσμένης ανίχνευσης (Compressed Sensing) και την αποτελεσματικότητα της ανάλυσης κοινών χωρικών προτύπων (Common Spatial Patterns) των σημάτων ΗΕΓ στην αυτόματη ανίχνευση εστιακών επιληπτικών κρίσεων. Η επιληψία είναι μια νευρολογική διαταραχή που χαρακτηρίζεται από μια διαρκή προδιάθεση για τη δημιουργία επιληπτικών κρίσεων με μεγάλες νευρολογικές, γνωστικές, ψυχολογικές και κοινωνικές συνέπειες. Σύμφωνα με τον Παγκόσμιο Οργανισμό Υγείας (WHO), εκτιμάται ότι το 2019, η επιληψία επηρεάζει περίπου 50 εκατομμύρια ανθρώπους παγκοσμίως, κάτι που είναι συνηθισμένο στην παιδική ηλικία. Το 2017, η επικράτηση και η επίπτωση της επιληψίας εκτιμάται ότι είναι 6,38 και 0,67 ανά 1000 άτομα αντίστοιχα. Σε αυτή την εργασία χρησιμοποιούμε το Discrete Cosine Transform (DCT) για να έχουμε μια αραιή αναπαράσταση των πληροφοριών για να μπορέσουμε να εφαρμόσουμε την τεχνική CS. Μειώνουμε περαιτέρω την ενέργεια του ΗΕΓ προκειμένου να έχουμε ένα πιο αραιό σήμα. Μετά τη συμπίεση των σημάτων, χρησιμοποιούμε τον αλγόριθμο Pursuit Basis για την ανακατασκευή του αραιού DCT σήματος και κατόπιν τον αντίστροφο Discrete Cosine μετασχηματισμό για να επιστρέψουμε στο πεδίο χρόνου. Στη συνέχεια, εφαρμόζουμε το μετασχηματισμό Fourier και την Approximate Entropy για να ελέγξουμε τη διατήρηση των αρχικών πληροφοριών της κρίσης. Η ανάλυση CSP έχει χρησιμοποιηθεί συχνά στη βιβλιογραφία για διαχωρισμό σήματος πολυκαναλικού ΗΕΓ μεταξύ - κυρίως - δύο κατηγοριών. Στην παρούσα διπλωματική εργασία, οι καταγραφές ΗΕΓ από 10 άτομα ηλικίας $6,8 \pm 5,9$ ετών, περιλαμβανομένων 63 επιληπτικών κρίσεων, αναλύθηκαν σε σχέση με την ανίχνευση κρίσεων και τη διάκριση μεταξύ inter-ictal και ictal χρονικών διαστημάτων σήματος. Τεχνικές Machine learning για την επιλογή και την ταξινόμηση χαρακτηριστικών χρησιμοποιήθηκαν στην ανάλυση.

Acknowledgements

First, I would like to thank Professor Michalis Zervakis who as my advisor guided and supported me during the course of this thesis and for giving me the opportunity to do this research.

Next, I would like to thank George Giannakakis, for his support and for providing me with valuable knowledge about epilepsy.

Also, I would like to thank Katerina Giannakaki for her contribution in the successful conclusion of this thesis and Dr Pelagia Vorgia for her expertise as a neuropsychiatrician.

Last, but not least, I would like to thank my family for their love, support, and constant encouragement throughout these years.

Table of Context

Chapter One: Introduction	1
1.1 Epilepsy	1
1.2 Types, Syndromes and Characteristics	4
References	12
Chapter Two: Electroencephalogram and Diagnostic Epilepsy Techniques	14
2.1 Electroencephalogram	14
2.2 Acquisition systems	14
2.3 EEG Rhythms	19
2.4 Epilepsy diagnosis through EEG	21
References	23
Chapter Three: EEG compression using compressed sensing (CS)	24
3.1 Representation of EEG using compressed sensing (CS)	24
3.2 Sparsity	26
3.3 Conditions for sparse recovery	28
Restricted Isometry Property	29
3.4 Recovery algorithms	30
Convex Optimization Algorithms	30
Basis Pursuit Algorithm	31
3.5 Reconstruction of Compressed EEG	32
Frequency Domain Reconstruction (Fourier)	37
Reconstruction based on Approximate Entropy (ApEn)	41
References	45
Chapter Four: Analysis and Classification of epileptic seizures using Common Spatial Patterns	46
4.1 Common Spatial Patterns (CSP)	46
Mathematical Formulation	47
4.2 Time window determination	49

4.3 Frequency Analysis of EEG for Common Spatial Patterns filters (CSP)	50
4.4 Feature Extraction Using Common Spatial Patterns filters (CSP)	52
4.5 Feature Selection Methods.....	56
4.6 Classification Schemes	59
4.7 Cross Validation methods	61
Validation	61
K-Fold Cross Validation	61
References	63
Chapter Five: Proposed Approach and Results	65
5.1 Clinical Protocol and procedures	65
Inclusion criteria and ethics	65
Procedure	65
Dataset.....	66
EEG preprocessing and feature extraction	67
5.2 Proposed Pipeline and Use of Techniques	68
Feature Selection	68
Classification.....	68
5.3 Results	70
Refernces	88
Chapter Six: Conclusion	89
6.1 Conclusion	89
6.2 Future Work.....	90
References	91

List of Figures

Figure 1: Classification of Seizures	4
Figure 2: Classification of Focal Seizures	5
Figure 3: Classification of Generalized Seizures	5
Figure 4: Classification of Unknown Seizures	6
Figure 5: Classification of the Epilepsies	7
Figure 6: Position of Electrodes in 10-20 Model - source [6]	16
Figure 7: 10-20 Model (A) - 10-10 Model (B) - source [8]	16
Figure 8: EEG Electrode Recording – source [11]	18
Figure 9: Representation of Brain Rhythms – source [14]	19
Figure 10: Nyquist Theorem – source ([3])	24
Figure 11: Compressed Sensing – source([4])	25
Figure 12: Compressed Sensing Basis Transform - sources([5] [6])	25
Figure 13: Representing the sparsest l1-ball solution	27
Figure 14: Inter-Ictal EEG (Above) and DCT of EEG (Below).....	32
Figure 15: Ictal EEG (Above) and DCT of EEG (Bellow).....	33
Figure 16: 100%,99.5% Energy for inter-ictal and Ictal DCT.....	33
Figure 17: Inverse Transform 99.5% Energy, inter-Ictal(Above)-Ictal(Below).....	34
Figure 18: Inverse Transform 99% Energy, inter-Ictal(Above)-Ictal(Below).....	35
Figure 19: original EEG (green) - DCT (blue)	36
Figure 20: reduced DCT (red) - Compressed signal after CS (black)	36
Figure 21: Reconstructed signal (dark red) - inverse DCT signal (light green).....	37
Figure 22: inter-Ictal EEG (Above) and Ictal EEG (Below).....	38
Figure 23: 100%, 99% Comparison in Frequency Spectrum.....	38
Figure 24: 100%, 99.5% Comparison in Frequency Spectrum.....	39
Figure 25: 100%, 99.9% Comparison in Frequency Spectrum.....	39
Figure 26: 100%, 99% Comparison in Frequency Spectrum.....	40
Figure 27: 100%, 99.5% Comparison in Frequency Spectrum.....	40
Figure 28: 100%, 99.9% Comparison in Frequency Spectrum.....	41
Figure 29: Approximate Entropy (ApEn)	44
Figure 30: Data Organization for CSP	50
Figure 31: Spectrum of Each Rhythm	51
Figure 32: Time Domain of Whole Spectrum and Rhythms.....	52
Figure 33: Inter-ictal (left graph) and ictal (right graph) most significant spatial filters for the patient (PAT_13)	56
Figure 34:The graphs represent the mean CSP features map for inter-ictal (left graph) and ictal (right graph).	56
Figure 35: Feature Selection	58
Figure 36: 4-fold cross validation	62
Figure 37: Sensitivity – Specificity – source([10])	69
Figure 38: Classification between non-ictal (green color) and ictal period (red color) representing the first 2 best features (for visualization purposes) – source([11]).	72

List of Equations

Equation 1: Compressed Sensing	25
Equation 2: Sparsest Solution	26
Equation 3: Spark(A)	28
Equation 4: Null Space Property	28
Equation 5: Mutual Coherence $\mu(A)$	29
Equation 6: Restricted Isometry Property	29
Equation 7: Mean Square Error (Above) and Root Mean Square Error (Bellow)	37
Equation 8: Approximate Entropy	43

List of Tables

Table I: Brain Rhythms.....	20
Table II: Feature matrix with the extracted features, consists of the patient-subject segments after the balancing procedure	54
Table III: Final Feature Matrix for each Segment	55
Table IV: Study population Demographics and Seizure Types.....	67
Table V: Features and Methods used in the analysis of the EEG signals (γ rhythm out).....	71
Table VI: Features and Methods used in the analysis of the EEG signals	71
Table VII: Results of the analysis (γ rhythm out).....	81
Table VIII: Results of the analysis.....	87

Chapter One: Introduction

1.1 Epilepsy

Epilepsy is a brain disease characterized by abnormal brain activity provoking seizures or abnormal behavior, sensations and sometimes loss of awareness [1]. It results in neurological, cognitive, psychological and social consequences. Epilepsy is affecting around 50 million people worldwide and accounts for a considerable percentage of the world's burden of disease. In 2017, the point prevalence of active epilepsy was 6.38 per 1,000 persons, while the lifetime prevalence was 7.60 per 1,000 persons. The annual cumulative incidence and incidence rate of epilepsy were 67.77 and 61.44 per 100,000 person-years respectively. Age group, sex, or study quality did not affect the prevalence of epilepsy. Low to middle-income countries had higher rates of active annual period prevalence, lifetime prevalence, and incidence rate of epilepsy. The highest prevalence was on epilepsies of unknown etiology and those with generalized seizures [2].

The percentage of the people worldwide ailing with epilepsy is anticipated to increase further due to rising life expectancy worldwide and also because of the increasing proportion of people surviving insults which frequently lead to epilepsy, such as birth trauma, traumatic brain injury (TBI), brain infections and strokes. Practically just about 80% of people with epilepsy live in low- and middle-income countries (LMIC). In those countries, the treatment gaps transcend 75% in the majority of the low-income and 50% in the majority of middle-income countries. Although the effectiveness and low cost of antiseizure medicines, these problems persist. Epileptic people ailing from recurring seizures that often occur spontaneously and without any warning. With suitable diagnosis and usage of cost-effective, and commonly available, antiseizure medicines up to 70% of epileptic people could become seizure-free [1]. This can result in epileptic people to continue, or get back to a full, productive, normal life. The definition and the classification of seizures and epilepsy were recently revised by the ILAE to meet advances in scientific knowledge. Classification of seizure types is of great importance for the selection of suitable therapies and renders a common language for providing adequate quality care. The classification of seizures, originally based on the 1981 classification [3], was revised by the ILAE with modifications in terminology and recognition of new seizure types based on advances in scientific knowledge [4]. A fundamental distinction is made between seizure onset that is focal (seizures arise in one hemisphere of the brain); generalized (originates in both hemispheres simultaneously); and seizures of unknown onset.

Focal seizures are subclassified according to whether awareness (a marker for consciousness) is intact or impaired [4].

Epilepsy has a higher appearance in children and elder people, while teenagers and adults are more scarcely affected. Based on the International League Against Epilepsy (ILAE), epilepsy is characterized by the following [5]:

1. At least two unprovoked (or reflex) seizures occurring more than twenty-four hours apart
2. One unprovoked (or reflex) seizure and a probability of further seizures similar to the general recurrence risk (at least 60%) after two unprovoked seizures, occurring over the next 10 years.
3. Diagnosis of an epilepsy syndrome

Epilepsy is considered to be resolved for individuals who had age-dependent epilepsy syndrome but are now past the applicable age or those who have remained seizure-free for the last 10 years, with no seizure medicines for the last 5 years.

Epilepsy syndrome adverts to a complex of features incorporating types of seizures, EEG and imaging features, which incline to be shewed together. Typically, it has age-related features such as the age of onset and recession (possibly), seizure triggers, daytime fluctuation, and sometimes prognosis.

It may also have discreet co-morbidities, such as mental and psychiatric dysfunction, along with specific findings in EEG and imaging studies. It may have relevant etiology, prognosis, and effects on the treatment. It is important to note that an epilepsy syndrome does not have a one-to-one connection with an etiological diagnosis and subserves a different purpose, such as management guidance. There exist many well-known syndromes, such as childhood epilepsy, West syndrome, and Dravet syndrome.

Elements of Consciousness

- **Awareness** of ongoing activities
- **Memory** for time during the event
- **Responsiveness** to verbal or nonverbal stimuli
- **Sense of Self** as being distinct from others

Causes of epilepsy

The causes of epilepsy are divided into the following categories: structural, genetic, infectious, metabolic, immune and unknown [6]. Investigating the causes of epilepsy is crucial to ensure early detection and management of the various etiologies, but also to implement interventions aimed at decreasing the large portion of preventable epilepsies, for example, from traumatic brain injury (TBI), stroke, etc.

- Structural: Abnormalities visible on structural neuroimaging which may be acquired (e.g. epilepsy from stroke, trauma and infection), or may be genetic (e.g. epilepsy from malformation in development of the cerebral cortex).
- Genetic: A known or presumed genetic mutation in which seizures are a core symptom of the disorder.
- Infectious: A known infection in which seizures are a core symptom (such as meningitis or encephalitis). Common examples are neurocysticercosis, tuberculosis, HIV, cerebral malaria, and congenital infections such as Zika virus and cytomegalovirus.
- Metabolic: A known or presumed metabolic disorder in which seizures are a core symptom (e.g. porphyria, uraemia or pyridoxine-dependent seizures).
- Immune: An immune disorder in which seizures are a core symptom. Autoimmune disorders affect multiple organ systems and frequently involve CNS inflammation.
- Unknown: The cause of the epilepsy is not yet known

Effects of epilepsy

Epilepsy is a brain disorder and cause of that, it is able of affecting many different systems throughout the body. Seizures can cause arrhythmia by affecting the normal rhythm of the heart. This can be caused by slowing the beating rate of the heart or making it go faster or even erratically. Another effect caused by epilepsy is the change in the hormonal that can cause interference with the reproduction of both men and women. Of course, the majority of the epileptic people can reproduce normally. Furthermore, seizures can disrupt the autonomic nervous system that regulates body functions like breathing, causing it to briefly stop functioning correctly. Last but important mention is that certain types of epilepsy can affect the nervous system, causing a disturbance to the muscles that enable you to walk,

jump, and lift things. Some types of seizures can provoke involuntarily tightening, jerking and twitching to the muscles (e.g. tonic, atonic, clonic,...).

1.2 Types, Syndromes and Characteristics

The epilepsy classification is based on the location of the neuron discharge and the extent of the electrical activity spreading to the other cerebral neurons. Regarding the 3rd criterion, it concludes in 3 different types of seizure, the partial or focal, the generalized and the unknown. These segregations are a consequence of the seizures that the patient suffers. Seizures are divided into Focal (or partial) and Generalized. Furthermore, Focal seizures are subclassified according to whether awareness (a marker for consciousness) is intact or impaired. Next, focal seizures are further divided into motor or non-motor. A seizure that begins focally (in one area of the brain) and then spreads bilaterally is termed focal to bilateral tonic-clonic. Tonic refers to the stiffening and clonic to the rhythmical jerking. In the same way, we have the Generalized seizures categorized as motor and non-motor (absence).

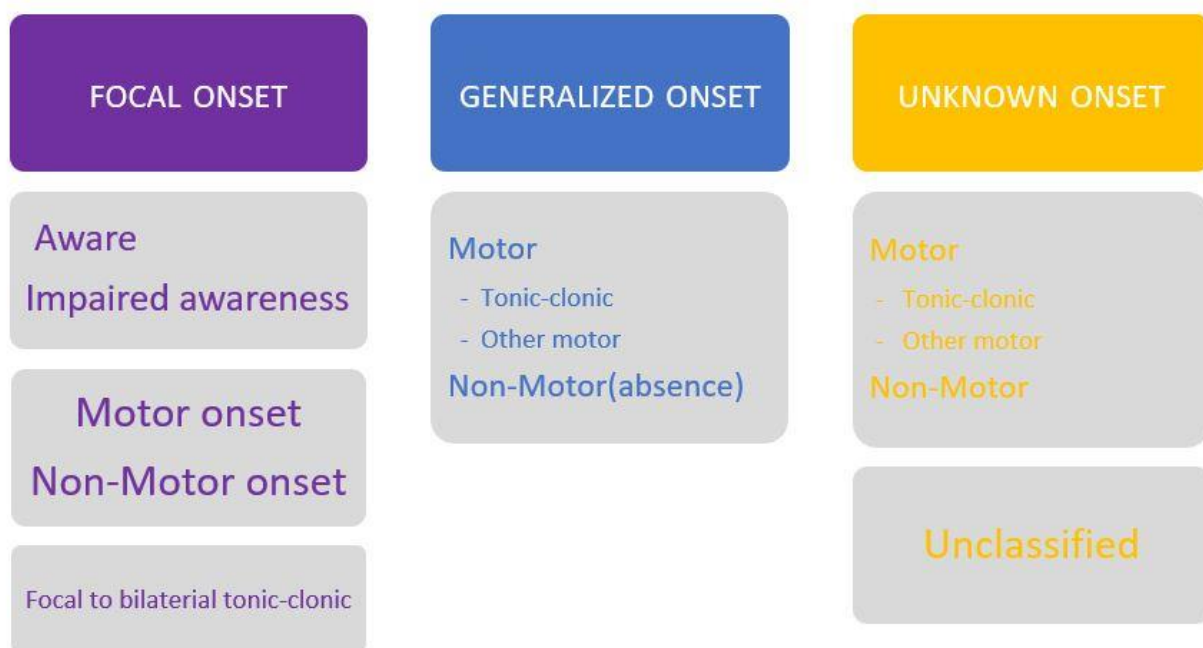


Figure 1: Classification of Seizures

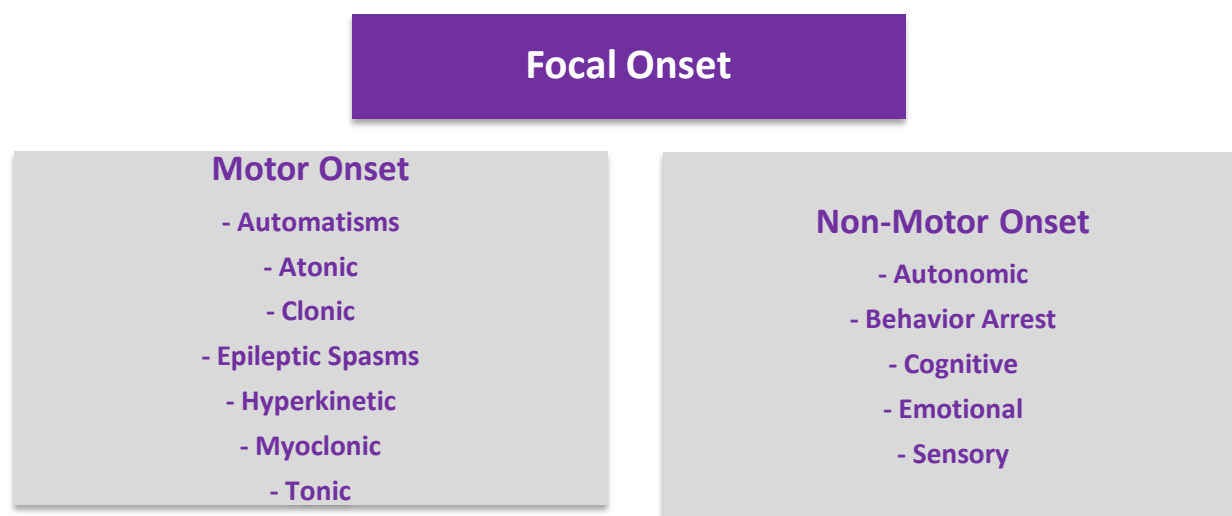


Figure 2: Classification of Focal Seizures

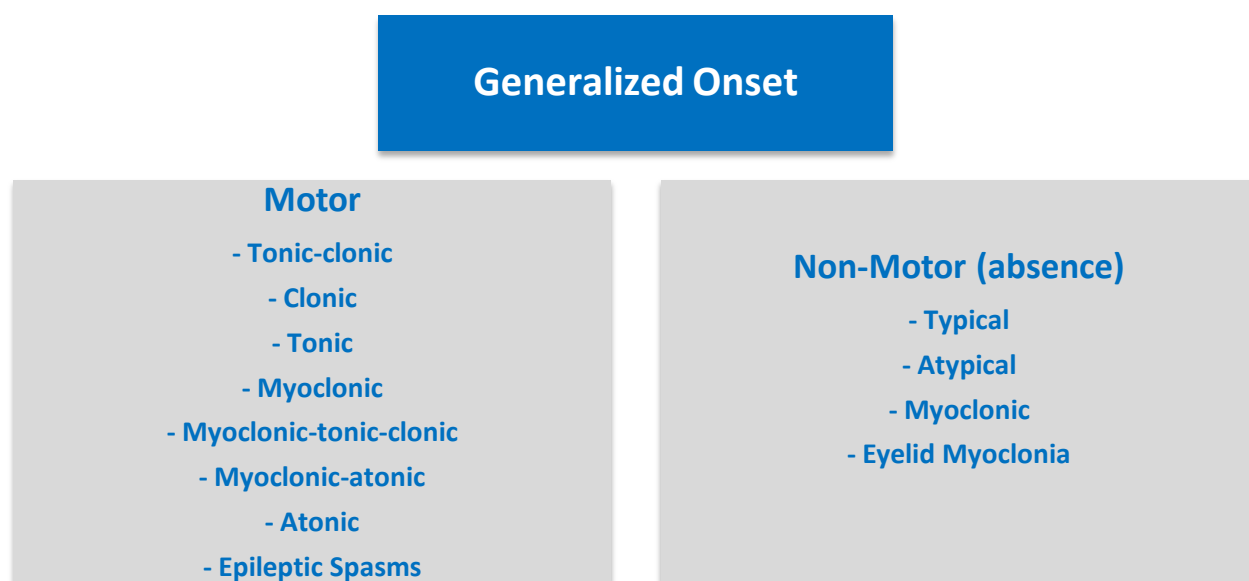


Figure 3: Classification of Generalized Seizures

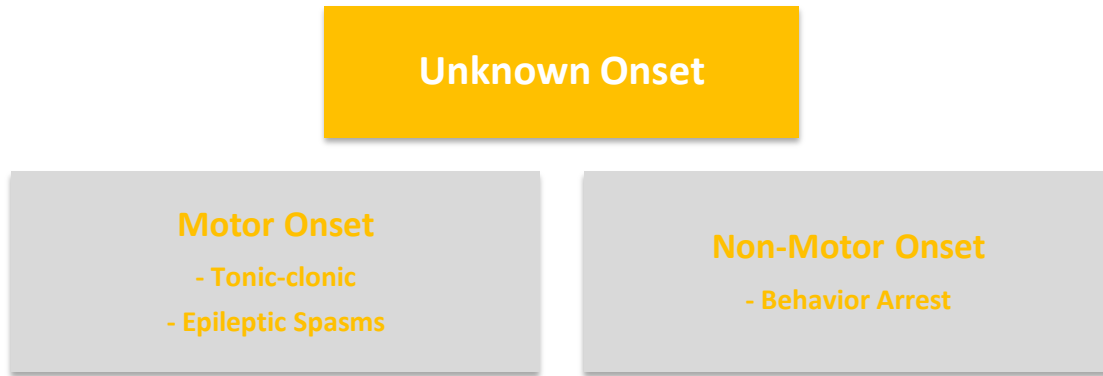


Figure 4: Classification of Unknown Seizures

Epilepsy classification aligns with the seizure classification. Levels of classification include seizure type, epilepsy type and epilepsy syndrome.

Seizure Types					
Focal Onset		Generalized Onset		Unknown Onset	
Motor Onset	Non-motor Onset	Motor Onset	Non-motor Onset	Motor Onset	Non-motor Onset
Automatisms	Autonomic	Tonic-clonic	Typical	Tonic-clonic	
Atonic	Behavior arrest	Clonic	Atypical	Epileptic spasms	Behavior arrest
Clonic	Cognitive	Tonic	Myoclonic		
Epileptic spasms	Emotional	Myoclonic	Eyelid myoclonia		
hyperkinetic	sensory	Myoclonic-tonic-clonic			
Myoclonic		Myoclonic-atomic			
tonic		atonic			
		Epileptic spasms			

Figure 5: Seizure Types

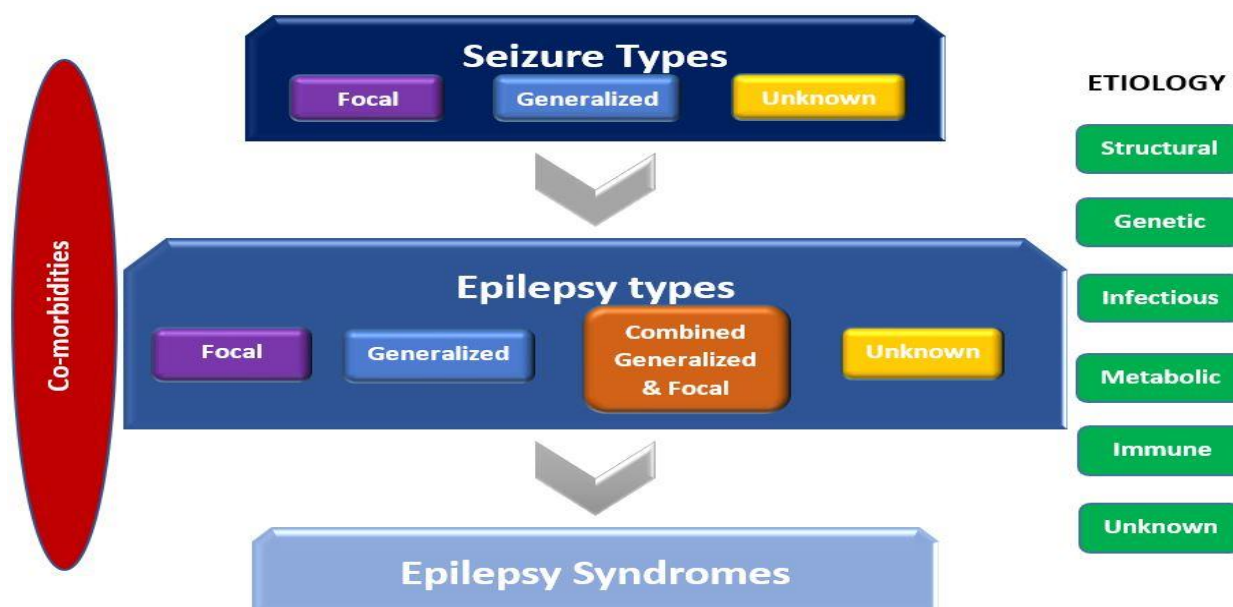


Figure 6: Classification of the Epilepsies

Indeed, epilepsy is an umbrella term that includes a sum of different types of epilepsies, some of which are recognized syndromes (e.g. childhood absence epilepsy) but a significant number for which an epilepsy syndrome has not been defined. When a person first presents with seizures, a physician or other health care provider starts by classifying the type of seizure. Then, the person's type of epilepsy needs to be classified and, in many cases, a specific epilepsy syndrome diagnosis can be made. There is augmenting awareness that many of the epilepsies are associated with co-morbidities such as intellectual and psychiatric impairment, and behavioral problems. The causes of epilepsy are divided into the following categories: structural, genetic, infectious, metabolic, immune and unknown. Investigating the causes of epilepsy is crucial to ensure premature detection and management of the various etiologies, but also to implement interventions aimed at decreasing the large portion of preventable epilepsies, for example, from traumatic brain injury (TBI), stroke, etc. [1].

Characteristics

Focal seizures

According to the previously referenced, if seizures are caused by abnormal activity in a certain area of the brain, they are termed focal seizures. These seizures fall into two main categories and carry the following characteristics:

Seizures without loss of consciousness.

These seizures do not cause loss of consciousness. They can cause a variation in feelings or affect eyesight, odor, sensation, taste or hearing. They can also lead to an involuntary jolt of a part of the body, such as an arm or leg, and spontaneous sensory symptoms such as tingling, dizziness, and flashes.

Seizures with Impaired Awareness.

These include impairment or loss of consciousness or awareness. During a complex partial seizure, the patient may be gazing in the vacuum and not responding normally to his surroundings or making repetitive movements, such as hand rubbing, chewing, swallowing or walking in circles. Symptoms of focal seizures may be confused with other neurological disorders, such as migraine, narcolepsy or mental illness. Extensive examination and control are required to distinguish epilepsy from other neurological disorders.

Generalized seizures

Generalized seizures are affecting both cerebral hemispheres (sides of the brain) from the onset of the seizure. They produce a loss of consciousness, either briefly or for a longer period of time. Generalized-onset seizures are categorized into motor and non-motor (absence) seizures. The level of awareness is not used as a classifier for generalized seizures since the large majority (although not all) of generalized seizures are associated with impaired awareness. By definition of the generalized branch of the classification, motor activity should be bilateral from the onset, but in the basic classification, the type of motor activity does not need to be specified. In cases where the bilateral onset of motor activity is asymmetrical, it may be difficult in practice to determine whether a seizure has focal or generalized onset.

The sudden suspension of activity and awareness are some of the characteristics of absence seizures. Their appearance is more frequent in younger age groups, the start, and termination of these seizures is briefer, and they commonly demonstrate less complex automatisms than focal seizures with impaired awareness. One of the main issues is that they are not easily distinguished from the focal seizures. For accurate classification, EEG information may need to be provided. Focal epileptiform activity is possible

to be seen with focal seizures. Also, bilaterally synchronous spike-waves can be perceived with absence seizures.

Unknown seizures

Seizures of unknown onset can be categorized as motor, non-motor, or unclassified. The term unclassified was used because it can accommodate both seizures with patterns that do not match into either of the two other categories or seizures displaying insufficient data to allow categorization to one. The most significant usage of this classification is for the tonic-clonic seizures for which the beginning was overclouded. Sufficient information might permit reclassification as a focal or generalized-onset seizure. Some other types of unknown onset seizures are epileptic spasms and behavior arrests. Detailed video-EEG monitoring is necessary for the clarification of the nature of the onset of the epileptic spasms, and also it can help with the treatment, because it may correspond to a focal pathology that may be treatable. This classification is used for any seizures with inadequate information to be classified to either focal or generalized and for seizures that present characteristics that do not match with the characteristics of known seizures, like an unknown-onset behavior arrest seizure could represent a focal impaired awareness behavior arrest seizure or an absence seizure.

Characteristics – Description

Here are the descriptions of some types and glossary terms [7]:

Absence, typical: A sudden onset, interruption of ongoing activities, a blank stare, possibly a brief upward deviation of the eyes. Usually the patient will be unresponsive when spoken to. Duration is a few seconds to half a minute with very rapid recovery.

Absence, atypical: An absence seizure with changes in tone that are more pronounced than in typical absence or the onset and/or cessation is not abrupt, often associated with slow, irregular, generalized spike-wave activity.

Atonic seizures: Atonic seizures, also known as falling seizures, are seizures with sudden loss or diminution of muscle tone without apparent preceding myoclonic or tonic event lasting ~ 1 to 2 seconds, involving head, trunk, jaw, or limb musculature.

Automatism: A more or less coordinated motor activity usually occurring when cognition is impaired and for which the subject is usually (but not always) amnesic afterwards. This often resembles a voluntary movement and may consist of an inappropriate continuation of pre-ictal motor activity.

Autonomic seizure: A distinct alteration of autonomic nervous system function involving cardiovascular, pupillary, gastrointestinal, sudomotor, vasomotor, and thermoregulatory functions.

Aura: A subjective ictal phenomenon that, in a given patient, may precede an observable seizure.

Behavior Arrest: A focal behavior arrest seizure shows arrest of behavior as the prominent feature of the entire seizure. Arrest (pause) of activities, freezing, immobilization.

Clonic seizures: Clonic seizures are associated with recurrent or rhythmic movements of the muscles. These seizures usually affect the neck, face and hands. Jerking, either symmetric or asymmetric, that is regularly repetitive and involves the same muscle groups.

Epileptic spasms: A sudden flexion, extension, or mixed extension–flexion of predominantly proximal and truncal muscles that is usually more sustained than a myoclonic movement but not as sustained as a tonic seizure. Limited forms may occur: Grimacing, head nodding, or subtle eye movements. Epileptic spasms frequently occur in clusters. Infantile spasms are the best-known form, but spasms can occur at all ages.

Eyelid myoclonia: Are myoclonic jerks of the eyelids and upward deviation of the eyes, often precipitated by closing the eyes or by light. Eyelid myoclonia can be associated with absences, but also can be motor seizures without a corresponding absence, making them difficult to categorize.

Focal onset bilateral tonic–clonic seizure: A seizure type with focal onset, with awareness or impaired awareness, either motor or nonmotor, progressing to bilateral tonic–clonic activity. The prior term was seizure with partial onset with secondary generalization.

Myoclonic seizures: Myoclonic seizures usually appear as sudden jerks or spasms of the hands and feet. Sudden, brief (<100 mseconds) involuntary single or multiple contraction(s) of muscle(s) or muscle groups of variable topography (axial, proximal limb, distal). Myoclonus is less regularly repetitive and less sustained than is clonus

Myoclonic–atonic: A generalized seizure type with a myoclonic jerk leading to an atonic motor component. This type was previously called myoclonic–astatic.

Myoclonic–tonic–clonic: One or a few jerks of limbs bilaterally, followed by a tonic–clonic seizure. The initial jerks can be considered to be either a brief period of clonus or myoclonus. Seizures with this characteristic are common in juvenile myoclonic epilepsy.

Tonic seizures: Tonic seizures are brief seizures (usually < 60 seconds) consisting of the sudden onset of increased tone in the extensor muscles, cause muscle stiffness. These seizures usually affect the muscles of the patient's back, hands and feet and may cause a fall on the ground. These seizures are invariably longer than myoclonic seizures. Occasionally tonic seizures terminate with a clonic phase.

Tonic-clonic seizures: Tonic-clonic seizures, also known as grand mal seizures, are the most dramatic type of epileptic seizure and can cause a sudden loss of consciousness, blood pressure and trembling, and sometimes the loss of control of the bladder or tongue. During a tonic– clonic seizure, awareness is lost before or contemporaneously with the stiffening and jerking movements. Some tonic–clonic seizures may invoke a nonspecific feeling of an impending seizure or a brief period of head or limb version.

Unclassified: 75-year-old man known to have epilepsy reports an internal sense of body trembling and a sense of confusion. No other information is available. EEG and MRI are normal. This event is unclassified.

References

- [1] "Epilepsy: a public health imperative: summary," World Health Organization report, 2019.
- [2] K. M. Fiest, K. M. Sauro, S. Wiebe, S. B. Patten, C.-S. Kwon and J. Dykeman, "Prevalence and incidence of epilepsy A systematic review and meta-analysis of international studies," *Neurology*, vol. 88, pp. 296-303, 2017.
- [3] ILAE, "Proposal for revised clinical and electroencephalographic classification of epileptic seizures. From the Commission on Classification and Terminology of the International League against Epilepsy," *Epilepsia*, vol. 22, pp. 489-501, 1981.
- [4] R. S. Fisher, "The new classification of seizures by the International League Against Epilepsy 2017," *Current neurology and neuroscience reports*, vol. 17, p. 48, 2017.
- [5] R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. E. Jr, L. Forsgren, J. A. French, M. Glynn, D. C. Hesdorffer, B. Lee, G. W. Mathern, S. L. Moshe and E. Perucca, "A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475-482, 2014.
- [6] I. E. Scheffer, S. Berkovic, G. Capovilla, M. B. Connolly, J. French and L. Guilhoto, "ILAE classification of the epilepsies. Position paper of the ILAE Commission for Classification and Terminology," *Epilepsia*, vol. 58, no. 4, pp. 512-521, 2017.
- [7] R. S. Fisher, J. H. Cross, C. D'Souza, J. A. French, S. R. Haut, N. Higurashi, E. Hirsch, F. E. Jansen, L. Lagae, S. L. Moshe and J. Peltova, "Instruction manual for the ILAE 2017 operational classification of seizure types," *Epilepsia*, vol. 58, no. 4, pp. 531-542, 2017.
- [8] "World Health Organization (WHO) report, fact sheet No 999," February 2018.
- [9] R. S. Fisher, W. v. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee and e. al., "Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia*, vol. 46, pp. 470-472, 2005.
- [10] G. Giannakakis, M. Tsiknakis and P. Vorgia, "Focal Epileptic Seizures Anticipation based on Patterns of Heart Rate Variability Parameters," *Computer Methods and Programs in Biomedicine*, vol. 178, pp. 123-133, 2019.
- [11] G. Giannakakis, V. Sakkalis, M. Pediaditis, C. Farmaki, P. Vorgia and M. Tsiknakis, "An approach to absence epileptic seizures detection using Approximate Entropy," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013.
- [12] G. Giannakakis, V. Sakkalis, M. Pediaditis and M. Tsiknakis, "Methods for seizure detection and prediction: an overview," *Modern Electroencephalographic Assessment Techniques, Neuromethods*, vol. 91, pp. 131-157, 2015.

- [13] V. Sakkalis, G. Giannakakis, C. Farmaki, A. Mousas, M. Pediaditis and P. Vorgia, "Absence seizure epilepsy detection using linear and nonlinear EEG analysis methods," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013.

Chapter Two: Electroencephalogram and Diagnostic Epilepsy

Techniques

2.1 Electroencephalogram

The Electroencephalogram (EEG) was discovered in 1924 and it's a method of recording the electrophysiological activity of the brain. EEG is to date the only and irreplaceable method of diagnosis, classification, assessment of epileptic seizures. Previously, its diagnostic use was wider (as in brain tumors, hematomas, etc.), but with the discovery of imaging examinations (mainly Axial and Magnetic Resonance imaging), its diagnostic value in these conditions was almost nullified. The EEG remains potentially useful outside of epilepsy and in various other conditions of functional brain disorders such as headaches, disturbances of the level of consciousness.

The first recording of the electrical field of the human brain was made by the German psychologist H. Berger in 1924, who gave the recording the name Electroencephalogram (EEG) [1]. The EEG is the recording of electrophysiological activity of the brain utilizing electrodes attached to the scalp [2].

Differences in electrical potentials are caused by aggregated postsynaptic graded potentials from pyramidal cells that generate electrical dipoles between soma (cell of the neuron) and apical dendrites (neural branches). Brain electrical current consists predominantly of the fluctuation of the concentration of K^+ , Ca^{++} , Na^+ , and Cl^- ions, that are pumped through different channels in neuron membranes in the direction governed by membrane potential. The measurement of these signals requires special attention and accuracy since the measured electrical signals are weak and range from about $10\mu V$ to $200\mu V$ [3]. Thus, the need to amplify these signals (for optimum imaging) and the denser head cover with lead electrodes (for greater precision and superintendence of brain function) becomes apparent.

2.2 Acquisition systems

Many patterns have been invented for the selection of the positions of each electrode on the scalp, but what has preponderated is the International System 10-20 [4]. The name of the system is due to selecting 20% of the interval between the two ears as the interval between two electrodes and selecting 10% of the interval between the two ears as the interval from the ear to the nearest electrode. Additionally, abductions are positioned on the ear lobes as well as in positions near the eyes.

Abductions in the ear lobes can be used as ground electrodes – reference electrodes. Note that in actuality, the ground electrode can be situated anywhere on the scalp or ear. The reason being that in order to acquire the Probe Dynamics at a point on the dermal surface of the skull, we measure it as the difference in potential between this point and the ground electrode, which is mutual for all the electrodes of the same hemisphere. Cause to the fact that the auricles are perforated by a small number of nerves and have low perfusion, they have a particularly constant and low potential, making them optimal for attaching the ground electrodes. The widespread usage of the 10-20 system is because it can be adjusted to different scalp sizes, from infants to grown adults. Dr. Jasper established some specific guidelines at his work [4], which would be established in recommending a specific system to the federation. The specific guidelines were:

1. The position of electrodes placed should be determined by specific measurements of standard skull landmarks. The measurements should be proportional to the size and shape of the skull, insofar as possible.
2. Adequate coverage of all parts of the head should be provided with standard electrode placement, even if some would not be used in a given examination.
3. Designation of positions would be expressed in terms of brain areas covered (Frontal, Parietal, etc.) rather than only in numbers, in order to make communications more meaningful to the non-specialist, as well as workers in other laboratories.
4. Finally, anatomical studies would be carried out, which would provide additional documentation determining the cortical areas beneath each standard electrode position in the average subject.

The 10-20 electrode system was adopted for trial and the meetings of the General Assembly of the International Federation in Paris in 1949. The use of this system helps to make the results obtained in each examination more comparable with the result from various examinations. Also, it facilitates the communication between doctors and physicians who become familiar with the localization of EEG abnormalities in terms of the standard skull landmarks [5]. It should be pointed that each laboratory can use any system deemed appropriate.

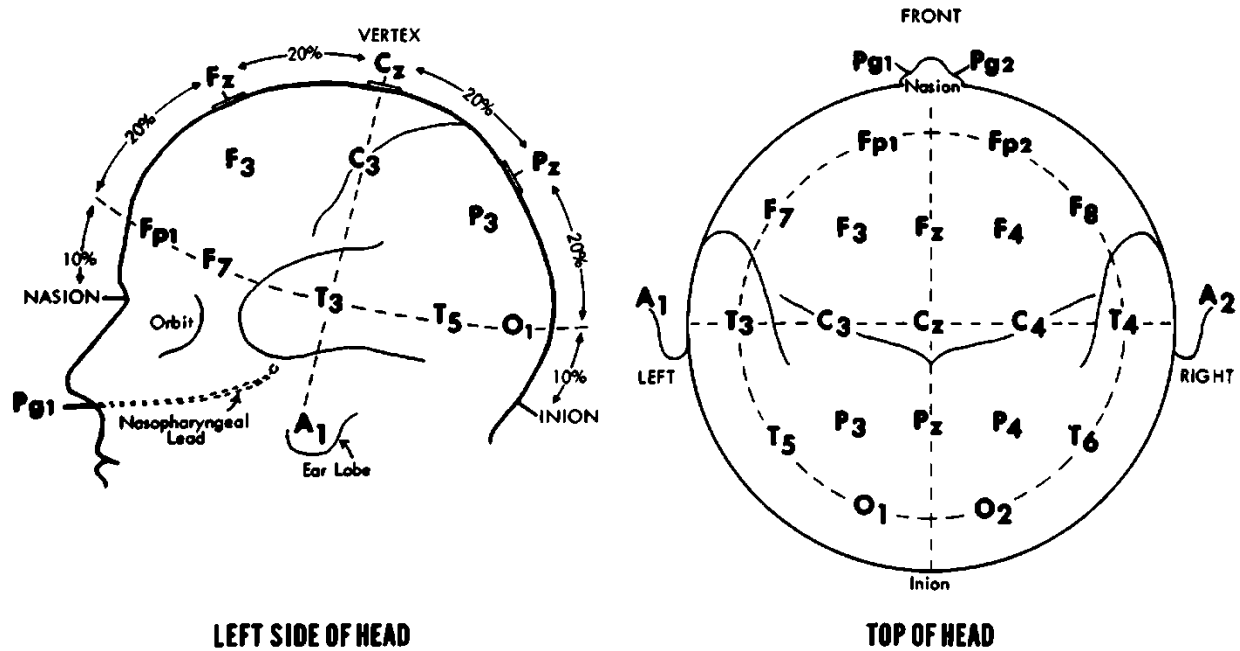


Figure 7: Position of Electrodes in 10-20 Model - source [6]

In these names the letter indicates the region of the brain (e.g. F - Frontal lobe). The even numbers correspond to electrodes located on the right side of the head and the redundant numbers on the left side of the head. To date, various extensions of model 10-20 (A), such as model 10-10 (B) [7] and model 10-5, in which electrodes are inserted between the positions of the 10-20 system, have been proposed.

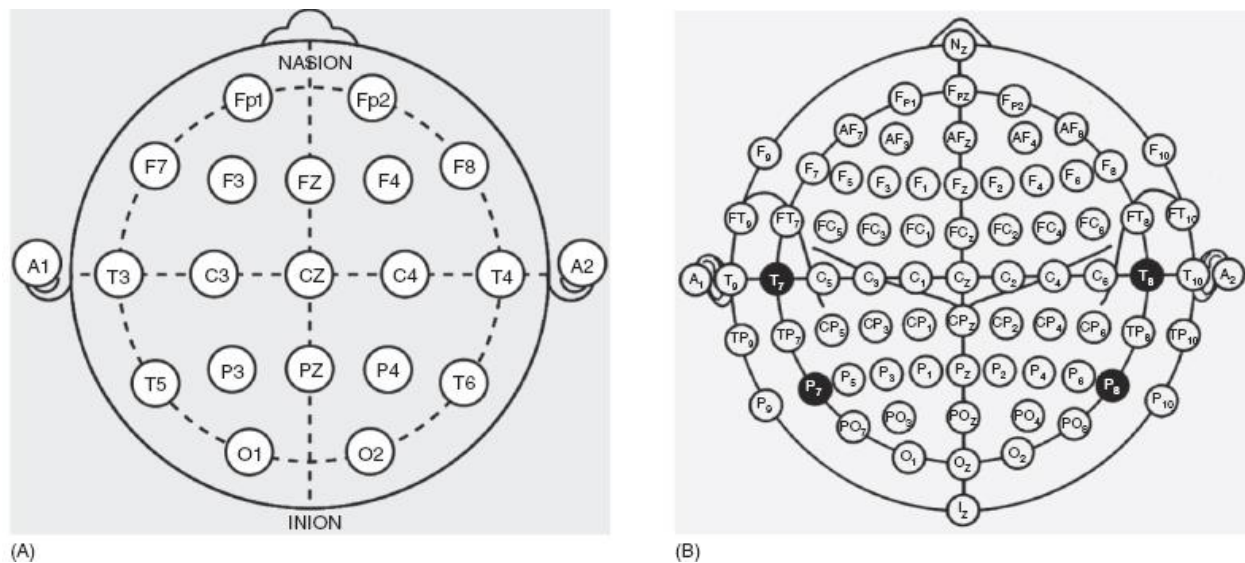


Figure 8: 10-20 Model (A) - 10-10 Model (B) - source [8]

These extensions are made in order to increase the number of channels provided. The model that has been accepted and documented by the American Electroencephalographic Society is 10-10 which is widely used today [16]. Model 10-5, although promising, is still in experimental stages. It is used to diagnose diseases of motor neurons and brain muscles, brain damage and epilepsy.

With this method, the neurologist can diagnose epilepsy and determine the type of epileptic seizures. It is also used to study sleep disorders (e.g. Narcolepsy) or when there are symptoms such as loss of consciousness, convulsions, disorientation, headaches and diffuse intense dizziness.

The electrodes are specific sensors that convert the ion current into the head of the electron beam. The received current, which is also the original electrical signal for the system, is advanced to the next processing steps. The contact of the electrodes with the surface of the skin is through an adhesive substance or through a small ring, which on one side adheres to the skin and on the other side to the main electrode. At the positions where the electrodes will be instated, the skin should be thoroughly cleaned with alcohol to achieve a low contact resistance of less than $5k\Omega$. In the same position, a special electrolyte-containing liquid is used. The electrode thus comes in direct contact with the underlying electrolyte used. Thus, it is possible to move the ions through the electrode-electrolyte border until equilibrium is achieved. This equilibrium is the outcome of the ionic concentration on both sides of the border. Finally, two charged layers are created on both sides of the border, one onto the metal surface and one on the liquid around the electrode, thereby generating a potential difference that obstructs the ion from continuing but is also sensitive to changes in ion concentration. When there exists electrical activity in the brain, i.e. ion flux, a change in ionic concentration is induced and a change in the potential difference of the layers and thus an electron flow from the conductive electrode side. It is desirable that the trend at the "boundary" be influenced only by ionic currents of the human brain and not by temperature changes or mechanical movements of the electrodes [10].

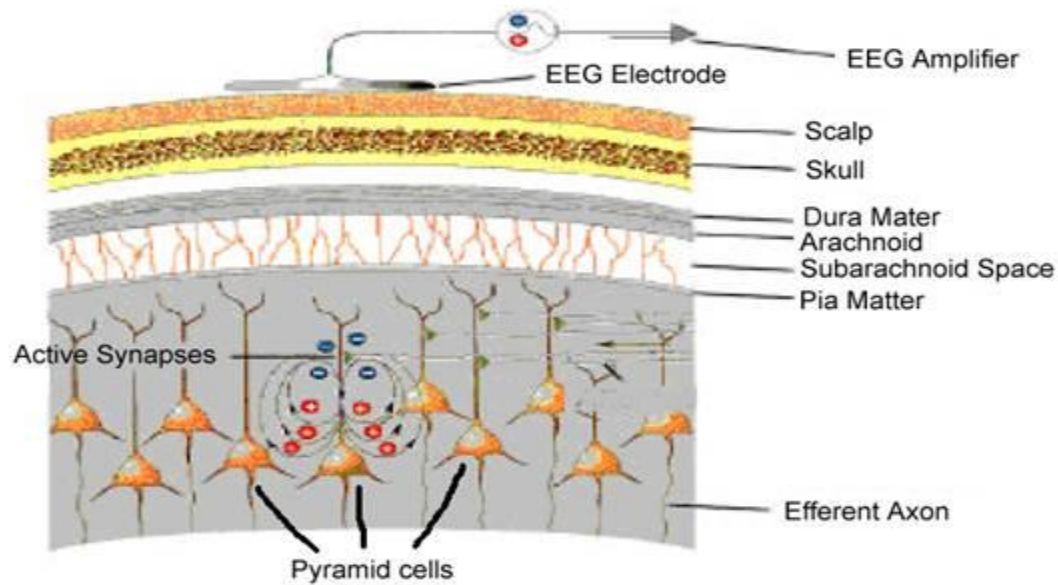


Figure 9: EEG Electrode Recording – source [11]

The amplifier differs from system to system depending on the signal to be amplified. The amplifier is used to amplify the signal, especially in the case of EEG, which is very weak. Electrodes that lie above brain regions that are active are said to correspond to active points. Instead, electrodes located above areas considered unrelated to cerebral function are said to correspond to inactive points. When the measured signal arises as a potential difference of two electrodes of active regions, then according to the terminology of EEG we have a bipolar measurement.

This method offers the advantage of rejecting any artifacts common to the two electrodes [2]. In the case of psychophysiological research, however, the measured signal results as a potential difference of an active region electrode and an inactive region electrode, so we have unipolar measurement.

Montages are logical and orderly arrangements of electrodes, that display EEG activity over the entire scalp, allow comparison of activity on both sides of the brain and aid in the localization of recorded activity to a specific brain region. From a clinical and practical standpoint, only a limited number of montages are needed during a recording session. The most mainly used montages are longitudinal bipolar "Double Banana", coronal bipolar, circumferential bipolar, laplacian, common average referential and ear (A1, A2) referential [12]. The use of bipolar montages during recording sessions provides better localized focal or regional features, while the use of referential montages is better for the detection of the morphology of widespread phenomena [13]. Each laboratory may use a variety of montages, but a significant amount of the used montages occasionally fails to display the EEG adequately or are inordinately complex. In order

to have an adequate EEG examination or for finding the solution to specific problems, additional EEG montages may be requisite.

2.3 EEG Rhythms

An approach to the study of EEG is based on the existence or not of specific waveforms called rhythms which are specific sub-bands of spectral content. The main EEG rhythms are delta, theta, alpha, beta and gamma. Their spectral content, location commonly appearing and the activity mostly associated with are shown in the image and table:

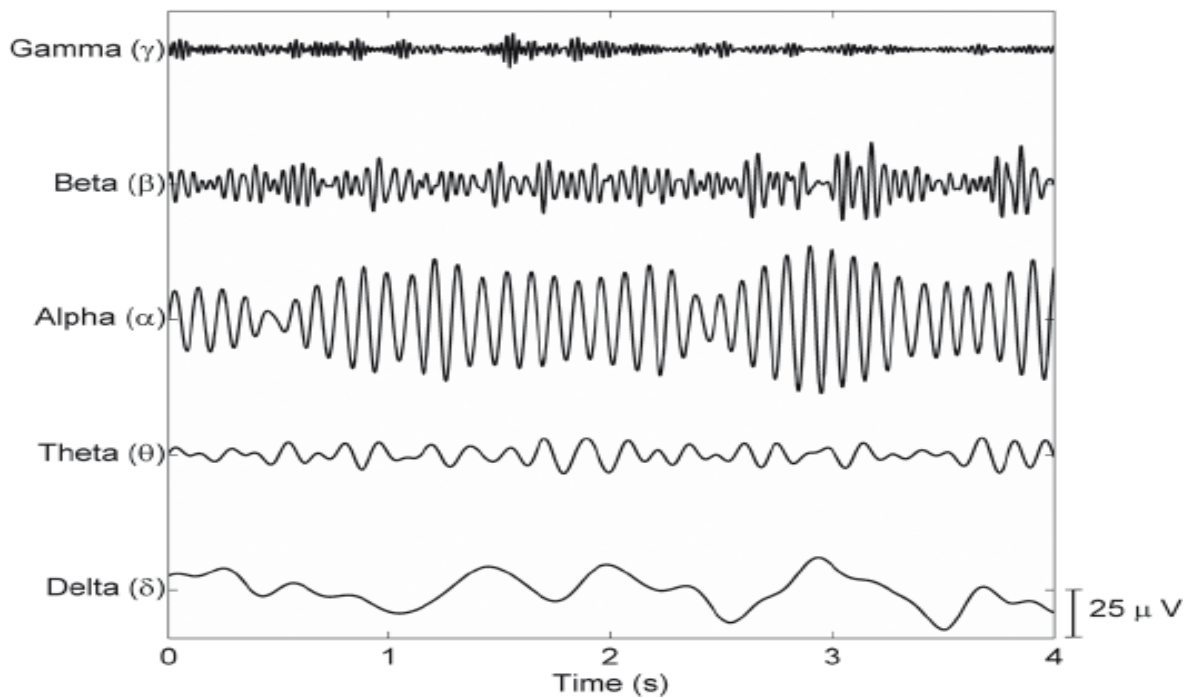


Figure 10: Representation of Brain Rhythms – source [14]

Sub-band	Freq (Hz)	Location	AMP (μV)	Activity
Delta	0.5 – 4	Thalamus	20 – 200	Deep sleep
Theta	4 – 7	Hippocampus, Cortex	20 – 100	Deep Relaxation Creativity
Alpha	8 – 13	Posterior regions, occipital lobe, Cortex	20 -60	Relaxation
Beta	13 – 30	Cortex	2 – 20	Memory, problem- solving
Gamma	30 - 100	Cortex	20 - 70	Cognition, information processing, learning

Table 1: Brain Rhythms

In about 75% of adults when they close their eye, the alpha rhythm is rising and the opening of their eyes cause reduction. It was named alpha because it was the first rhythm studied. The beta rhythm has been associated with active thinking, focus, high alert, anxiety. The delta rhythm has been associated with deep sleep in the adults and in infants. Theta rhythm has been associated with suppressive mechanisms like at the relaxation phase entry and drowsiness in adults. It is higher in young children. As a result, the EEG signal is also highly correlated with the level of alertness of the subject. When human activity is increased, the prevailing frequency of the EEG also increases but the amplitude lowers. During eye closure, alpha rhythm is dominant. When the subject is asleep, the EEG has lower frequency. Sleep is generally divided into 2 broad types: non-rapid eye movement (NREM) sleep and REM sleep. In deep sleep, EEG has long and slow deflections called delta rhythms. No brain activity can be detected by a patient in complete stroke death [15]. EEG signals are small enough, measured in microvolt (μV), making them very sensitive to noise. The rhythms of human brain are:

Delta: has a frequency of 0.5 to 4 Hz. The amplitude is between 20-200 μV and has the highest peak among the rhythms and the slowest waves. It is the predominant rhythm in infants up to one year and in stages 3 and 4 of sleep. It may occur out of sleep and may indicate the existence of structural or functional impairment. It is usually more prominent in adults (e.g. FIRDA - Frontal Intermittent Rhythmic Delta [16]) and posterior to children, e.g. OIRDA - Occipital Intermittent Rhythmic Delta [17]).

Theta: It has a frequency of 4 to 8 Hz, width up to 100 μV and is characterized as "slow" activity. It is perfectly normal in children up to 13 years of age and sleep, but abnormal in sleeping adults (non-sleeping). It can be considered as a manifestation of focal subcortical lesions.

Alpha: It has a frequency of between 8 to 13 Hz and a width of 20-60 μ V. It usually looks better at the rear areas of the head on each side and is higher in width on the predominant side. Observed in people who have eyes closed and relaxed, but not sleepy. With the onset of mental or physical activity, a de-synchronization is observed, as the rate is replaced by asynchronous rates of higher frequency and lower amplitude. The same phenomenon occurs as soon as the subject opens his eyes (blocking reaction). It is the main rhythm observed in relaxed adults, for the biggest part of their life, especially after the thirteenth year.

Beta: It is low in width about 20 μ V, mainly from parietal and frontal area. It has a frequency of 13 to 30Hz. It usually appears on both sides of the brain in a symmetrical distribution and is more apparent frontal. It is enhanced by sedative-hypnotic drugs, especially benzodiazepines and barbiturates. It may be absent or reduced in areas of cortical wear. It is generally considered to be a normal rhythm. It is the dominant rhythm in patients who are awake or anxious or with their eyes open.

Gamma: has a frequency above 30Hz. What is interesting is that rhythm is associated with consciousness, a function that is still incomprehensible, and which stimulates the interest of many scientists from many disciplines. Some researchers, however, do not distinguish rhythm c as a separate class, but place it on rhythm b.

2.4 Epilepsy diagnosis through EEG

Because of their nature, EEG signals can be used effectively to study the mental states and ailments related to the brain. Electroencephalogram (EEG) signals are a direct reflection of the electrophysiological junctures of the brain at a given timestamp. There are inherent issues with the EEG signals, because of the highly non-linear nature and the visual interpretations are subjective prone to inter-observer variations. In order to aid researchers and doctors to better analyze EEG signals, the use of various signal analysis techniques such as linear, non-linear, frequency domain and time-frequency methods are significant. Depending on the patient's behavior and current psychological condition the EEG differs, but even though the patient may be nervous or blinking his eyes, the EEG can still be rightly categorized as normal. The doctor can interpret the EEG and categorize it as normal or abnormal. If the EEG is abnormal, then there is a possibility of the patient being epileptic. With the help of EEG, the seizure type and epilepsy syndrome in patients with epilepsy can be defined, and thereby the selection of antiepileptic medication and the prediction of prognosis. The multi-axial diagnosis of epilepsy is being aided by the findings in the

EEG recordings, in terms of whether the seizure disorder is focal or generalized, idiopathic or symptomatic, or part of a specific epilepsy syndrome [14]. A significant number of epilepsy syndromes associated with specific EEG features is shewed in infancy or childhood. Not all the epileptic syndromes are accepted, others are questionable or they cannot be included in the current ILAE classification systems due to inadequate data on this particular type of epilepsy. The clinical diagnosis of epilepsy requires many matters to be evaluated. Some of these are neurological tests along with blood tests, sometimes cerebrospinal fluid tests to check for the probability of related causes and detailed patient history. Imaging techniques like CT (Computerized Tomography) or MRI (Magnetic Resonance Imaging) scans may be used to check for any structural abnormalities, such as tumors, abnormal blood vessels, and ischemia in the brain, which may be causing seizures that may not be due to epilepsy. However, the most common and effective diagnostic method for the detection of epilepsy is the EEG test. Seizure disorders, like focal or generalized seizures, display a percentage of overlap of both clinical and electrographic manifestations. The existence of unihemispheric epilepsies makes the differentiation even more difficult. All the same, the conceptual segregation of partial and generalized seizures or epilepsy types is up to the present time valid and clinically helpful. In practice, ground on the information provided by the patient and/or eye witness, the clinicians will be reasonably confident about the seizure type. On the other hand, when the history is indecipherable, EEG can assist in the differentiation betwixt a complex partial seizure with focal IED and an absence type seizure with generalized IED inter-ictal epileptiform discharge.

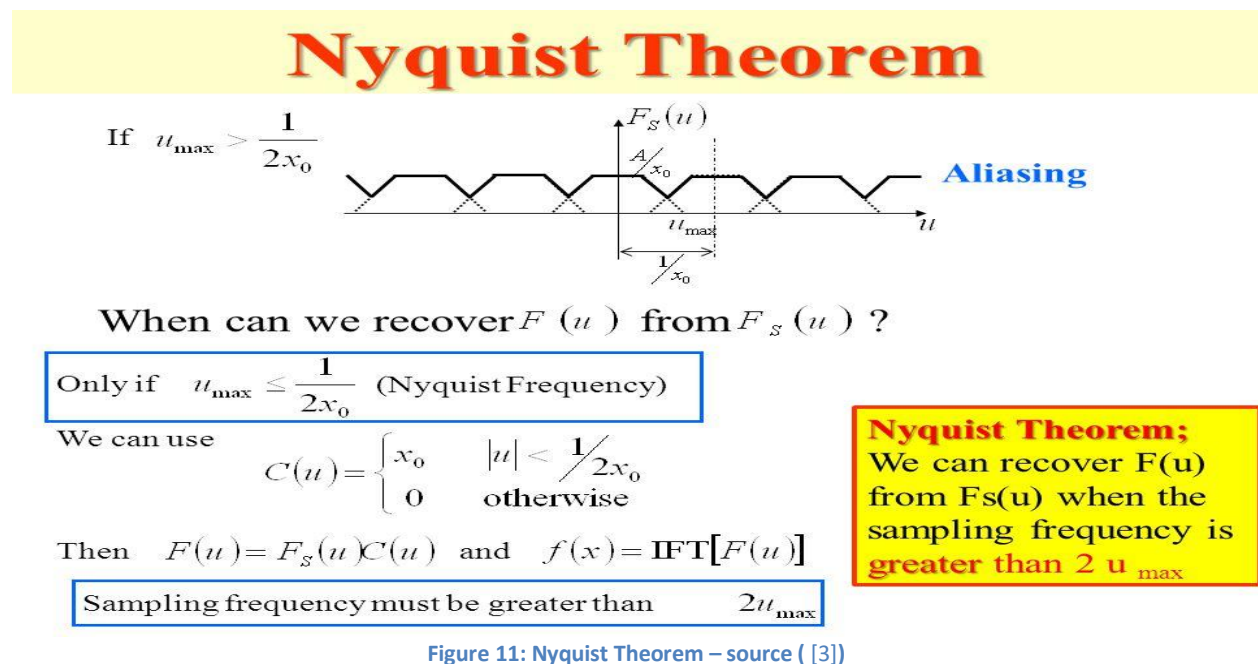
References

- [1] B. E. Swartz, "Timeline of the history of EEG and associated fields," *Electroencephalography and clinical Neurophysiology*, vol. 106, pp. 173-176, 1998.
- [2] P. Gloor, Hans Berger on the Electroencephalogram of Man. The Fourteen Original Reports on the Human Electroencephalogram, vol. Electroencephalography and Clinical Neurophysiology, New York: Elsevier, 1969.
- [3] J. Malmivuo and R. Plonsey, Bioelectromagnetism - Principles and Applications of Bioelectric and Biomagnetis, New York: Oxford University Pres, 1995.
- [4] H. H. Jasper, "The 10–20 electrode system of the International Federation," *Electroencephalography and clinical Neurophysiology*, vol. 10, pp. 370-375, 1958.
- [5] G. H. Klem, H. O. Luders, H. H. Jasper and C. Elger, "The ten±twenty electrode system of the International Federation," *Recommendations for the Practice of Clinical Neurophysiology: Guidelines of the International Federation of Clinical Physiolog*, no. 52, pp. 3-6, 1999.
- [6] "Instrumentation Forum," [Online]. Available: <https://instrumentationforum.com/t/electrode-10-20-system/5486>.
- [7] G. E. Chatrian, E. Lettich and P. L. Nelson, "Ten percent electrode system for topographic studies of spontaneous and evoked EEG activity," *American Journal EEG Technology*, vol. 25, pp. 83-92, 1985.
- [8] "Neupsy KeyFastest Neupsy Insight Engine," [Online]. Available: <https://neupsykey.com/eeg-instrumentation/>.
- [9] "American Electroencephalographic Society, Guideline thirteen: Guidelines for standard electrode position nomenclature," *J Clin Neurophysiol*, vol. 1994, pp. 111-113, 1994.
- [10] L. G. Tassinary, T. H. Geen, J. T. Cacioppo and R. Edelberg, "Issues in biometrics: Offset potentials and the electrical stability of Ag/AgCl electrodes," *Psychophysiology*, vol. 27, pp. 236-242, 1990.
- [11] "Epilepsy Foundation," Basic Science of EEG, [Online]. Available: <https://www.epilepsy.com/learn/professionals/diagnosis-treatment/basic-science-eeeg>.
- [12] "Guideline 6: A Proposal for Standard Montages to Be Used in Clinical EEG," American Clinical Neurophysiology Society, 2006.
- [13] M. M. Kabiraj and N. M. Biary, "Basic principles and interpretations of electroencephalography," *Neurosciences*, vol. 2, no. 8, pp. 131-143, 2003.
- [14] "IEEE Computer Society, Digital Library," [Online]. Available: <https://www.computer.org/csdl/magazine/co/2012/07/mco2012070087/13rUYgT7Bk>.
- [15] "American Clinical Neurophysiology Society Guideline 6: Minimum Technical Standards for EEG Recording in Suspected Cerebral Death," *J Clin Neurophysiol*, vol. 4, no. 33, pp. 324-327, 2016.
- [16] J. V. d. Drift and o. Magnus, "The value of the EEG in the differential diagnosis of cases with cerebral lesion," *Electroencephalogr Clin Neurophysiol.*, no. 11, pp. 733-746, 1959.
- [17] W. A. Cobb, "Rhythmic slow discharges in the electroencephalogram," *J Neurol Neurosurg Psychiatry*, no. 8, pp. 65-78, 1945.
- [18] S. J. M. Smith, "EEG in the diagnosis, classification and management of patients with epilepsy," *J Neurol Neurosurg Psychiatry*, no. 76, 2005.

Chapter Three: EEG compression using compressed sensing (CS)

3.1 Representation of EEG using compressed sensing (CS)

Compressed sensing belongs to the field of signal processing. Its purpose is typically to measure, filter or compress continuous analogue physical signals. Prior to compressed detection, the main method for signal sampling without loss of information was Shannon-Nyquist. This had the undesirable consequence of the very high rate in wireless applications, the excessive waste of time, energy, and resource use. With the theory of compressed detection, we are given the ability to construct sparse signals from several samples significantly smaller than Nyquist's [1] [2]. It has been proven that reconstruction is possible when the signal or even some transformation is sparse, that is, it contains few non-zero elements in relation to its total length.



To succeed in rebuilding the signal, CS is based on 2 principles: sparsity and in-coherence. This means that the test signal is sparse or it has a sparse representation in a certain base/domain, and the information can be encoded by a small number of inconsistent measurements. Therefore, Compression Analysis

Our main assumption is that the signal \mathbf{x} , that we are attempting to recover, is itself sparse or has a sparse representation \mathbf{s} on an orthonormal basis, such that $\mathbf{x} = \Psi^* \mathbf{s}$. The compressed Sensing problem is how to recover \mathbf{x} if the information we have comes from $\mathbf{y} = \Phi \times \mathbf{x}$, or in the case where \mathbf{x} is not sparse then to recover its sparse representation \mathbf{s} from $\mathbf{y} = \Phi \times \Psi \times \mathbf{s}$. In both situations, we have an underdetermined linear system of equations and only the fact that \mathbf{x} or \mathbf{s} are sparse. Φ is our sensing matrix and it possesses no zero columns. Usually, it is a random matrix, like an independent and identically distributed random Gaussian matrix, but we will analyze in a later part the attributes our sensing matrix must-have so that we can recover \mathbf{x} or \mathbf{s} .

Questions arising:

Which are the right signals and sparse models?

Which are suitable sensing matrices?

How to reconstruct the original signal?

3.2 Sparsity

Sparsity is the prerequisite information about the vector we seek to detect efficiently or whose dimension we intend to reduce. Many signals, if expressed on a different basis, can be described in a sparse representation. If the sparse signal \mathbf{x} or \mathbf{s} is k -sparse, meaning it has k non-zero data, then we can determine our solution.

The signal is measured by l_0 "norm" and is defined as follows:

$$\|\mathbf{x}\|_0 := \#\{i : x_i \neq 0\} \leq k$$

The total of all the k -sparse signals will be denoted by Σ_k . The issue in the problem $\mathbf{y} = \Phi \times \mathbf{x}$ $\|\mathbf{y} = \Phi \times \Psi \times \mathbf{s}$, is to find the as sparsest as possible solution, if it exists. This solution comes from solving the following problem:

$$(S_0) \min \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \Phi \times \mathbf{x}$$

Equation 2: Sparsest Solution

The specific problem to solve is characterized as NP-hard, so the l_0 -minimization problem usually reduces it to l_1 - and l_2 - minimization problems. Candes and Tao have discovered that this problem can be solved using the l_1 norm. With this method we can achieve accurate results. Thus the conclusion that the reconstruction is good enough. This problem is known as Basis Pursuit and is proposed as a cost-effective method for finding sparse solutions.

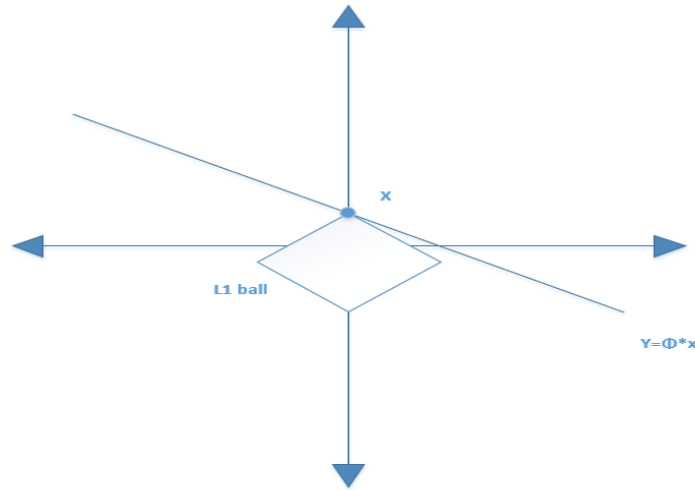


Figure 14: Representing the sparsest l_1 -ball solution

As a result, our new problem transforms to the following minimization problem $(S_1) \min \|x\|_1$ subject to $y = \Phi \times x$. However, in order to be able to implement it, we need both necessary and sufficient conditions, which are not only based on the sparsity of the original vector x but also on the incoherence of the sensing matrix Φ .

If the signal shows little or no sparsity then we can transform it to different representation systems, ie dictionary, until we get the desired sparsity. Various of these systems are wavelets, shearlets, Fourier base and so on. The existence or not of certain characteristics leads us to choose bases such as those mentioned above. In this case, however, the sparsity we will achieve may not be ideal. To improve sparsity, we need to select a dictionary-adaptive label, and to get it done we need a set of tests to learn it. This method is called dictionary learning, and the most well-known and used algorithm is the K-SVD algorithm. Certainly, from a mathematical point of view there are also some problems with the specific dictionaries.

3.3 Conditions for sparse recovery

After introducing sparsity, in the sense of signal models, we next need to find which conditions are needed to impose on the sparsity of the original vector and on the sensing matrix for a perfect recovery. The so-called spark is the condition we need on the sensing matrix in order to have an exact recovery. The concept of spark verbally fuses the notions of ‘sparse’ and ‘rank’. If A is a $m \times n$ matrix, then the spark of A , declared by $\text{spark}(A)$, is the minimal number of linearly dependent columns of A . We now need to reformulate the concept of spark in terms of the null space of A , until the end to be denoted by $N(A)$, and state its range.

$$\text{spark}(A) = \min \left\{ k : N(A) \cap \Sigma_k \neq \{0\} \right\}$$

$$\text{and } \text{spark}(A) \in [2, m+1].$$

Equation 3: Spark(A)

With Σ_k being the set of all k -sparse signals, as mentioned before. Through this concept, we can extract an equivalent condition on the unique solvability of (S_0) .

For A a $m \times n$ matrix and $k \in \mathbb{N}$, the two conditions are equivalent.

- (i) If x is a solution of (S_0) and satisfies $\|x\|_0 \leq k$, then x is the unique solution.
- (ii) $k < \text{spark}(A) / 2$.

Because of the underdeterminedness of A and therefore the ill-posedness of the recovery problem, the $N(A)$ has a special role in the analysis of the unique solution of the minimization problem (S_1) and is through the null space property that is defined as follows. A has the null space property (NSP) of order k , if, for all $h \in N(A) \setminus \{0\}$ and for all index sets $|\Lambda| \leq k$,

$$\|1_\Lambda h\|_1 < \frac{1}{2} \|h\|_1.$$

Equation 4: Null Space Property

In terms of the null space property, we can state an equivalent condition for the unique and sparse solution of (S_1) . For A a $m \times n$ matrix and $k \in \mathbb{N}$, the two conditions are equivalent.

- (i) If x is a solution of (S_1) and satisfies $\|x\|_0 \leq k$, then x is the unique solution.

- (ii) A satisfies the null space property of order k.

The main idea of compressed sensing is by determining when $\ell_o = \ell_1$, meaning when the unique solutions of (S_0) and (S_1) concur. The mutual coherence and the restricted isometry property are the most well-known conditions for when the two solutions coincide. The mutual coherence of a matrix measures the smallest angle between each pair of its columns. If $A = (a_i)_{i=1}^n$ is a $m \times n$ matrix, its mutual coherence $\mu(A)$ is stated as

$$\mu(A) = \max_{i \neq j} \frac{|\langle a_i, a_j \rangle|}{\|a_i\|_2 \|a_j\|_2}$$

Equation 5: Mutual Coherence $\mu(A)$

When two separate columns coincide then mutual coherence is 1, and it is the maximal value. The Welch bound is the lower bound displayed in the next equation. If A is a $m \times n$ matrix, then

$$\mu(A) \in \left[\sqrt{\frac{n-m}{m(n-1)}}, 1 \right]$$

There exist different variants of mutual coherence, in particular, the *Babel function*, the *cumulative coherence function*, the *structured p-Babel function*, the *fusion coherence* and *cluster coherence*.

Restricted Isometry Property

The restricted isometry property measures the degree to which each subset of k column vectors of A is close to being an isometry. If A is a $m \times n$ matrix, then A has the Restricted Isometry Property (RIP) of order k, if there exists a $\delta_k \in (0, 1)$ such that

$$(1 - \delta_k) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_k) \|x\|_2^2 \text{ for all } x \in \Sigma_k$$

Equation 6: Restricted Isometry Property

Matrix A satisfies the k-class RIP if the constant δ_k does not have a value very close to one. When this property holds true, all k-subsets of the columns of table A are almost rectangular. If this property is not valid then it may be impossible to reconstruct these vectors. Evaluating the RIP property is in itself an NP-

Hard problem and thus, while serving as a fundamental theorem for Compressed Sensing, it still does not give a perspective on how to construct a compressed sensing matrix [7].

As already mentioned, sensing matrices are required to satisfy certain incoherence conditions such as, for instance, a small so-called mutual coherence. If we are allowed to choose the sensing matrix freely, the best choice are random matrices such as Gaussian iid matrices, uniform random ortho-projectors, or Bernoulli matrices [8].

3.4 Recovery algorithms

Having completed the above procedures, we will need to proceed to a sparse recovery process. There are 3 types of algorithms that we can use for this process. Convex Optimization, Combinatorial and Greedy. The measurements required for Curved Convex Optimization algorithms are few in number but have higher complexity from a computational view. On the other hand, are the Combinatorial algorithms, which are very fast, often sub-linear, require a lot of measurements that are not always possible to obtain. The best compromise between the two aforementioned algorithms are Greedy algorithms in terms of computational complexity and the required number of measurements.

Convex Optimization Algorithms

We've analyzed the Convex Optimization problem above

$$\min \|x\|_1 \text{ subject to } y = \Phi \times x$$

In case that the measurements were affected by noise, there is a requirement for a conic constraint and the minimization problem needs to be altered to

$$\min \|x\|_1 \text{ subject to } \|y - \Phi \times x\|_2^2 \leq \varepsilon \text{ for } \varepsilon > 0.$$

For this purpose, cumulative optimization algorithms specifically developed for CS have been developed. These algorithms include the interior-point methods, the projected gradient, and the iterative thresholding.

Combinatorial Algorithms

These algorithms are applying a group testing to highly structured samples of the original signal. Of the various types of this category of algorithms, we refer the HHS pursuit and the sub-linear Fourier transform.

Greedy Algorithms

The greedy algorithms repeatedly approach the coefficients and support of the original signal. Their main advantages are that they are very fast and easy to implement. Frequently their theoretical performance guarantees are very akin to the ℓ_1 minimization results. The best-known greedy approach is Orthogonal Matching Pursuit [9]. Known greedy algorithms are stagewise OMP (StOMP), regularized OMP (ROMP), and compressive sampling MP (CoSamp).

Basis Pursuit Algorithm

Basis Pursuit (BP) is the convex optimization problem $\min \|x\|_1$ subject to $y = \Phi \times x$, where the optimization variable is $x \in R^n$, $\|x\|_1 = |x_1| + |x_2| + \dots + |x_n|$ is the ℓ_1 -norm of the vector x , and $A \in R^{m \times n}$ is a matrix with more columns than rows. BP seeks the “smallest” (in the ℓ_1 -norm sense) solution of the underdetermined linear system $y = \Phi \times x$ [10]. To make sure that $y = \Phi \times x$ has at least one solution, Φ is required to be full rank. BP has recently attracted attention due to its ability to find the sparsest solution of a linear system under certain conditions. In particular, BP is a convex relaxation of the combinatorial and nonconvex problem obtained by replacing the ℓ_1 -norm in (BP) by the ℓ_0 -norm $\|x\|_0$, which counts the number of nonzero elements of x . Note that the linear system $y = \Phi \times x$, has a unique k -sparse solution, i.e., a solution whose ℓ_0 -norm is k , if every set of $2k$ columns of Φ is linearly independent. Notice that A is full rank with probability one if the entries of A are independent and

identically distributed (i.i.d.) and drawn from some (nondegenerate) probability distribution, as commonly seen in compressed sensing.

3.5 Reconstruction of Compressed EEG

For the use of the Compressed Sensing technique, we need a sparse signal. However, electroencephalogram is not a sparse signal due to its increased complexity. In order to be able to use the CS technique we need to bring it in a sparse form. To achieve this, we convert the signal using Discrete Cosine Transform (DCT).

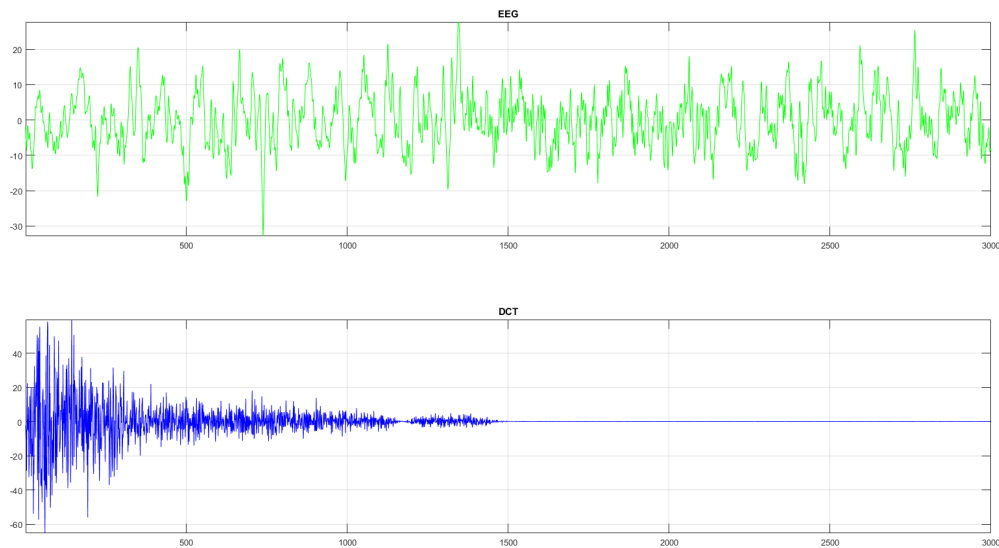


Figure 15: Inter-Ictal EEG (Above) and DCT of EEG (Below)

The green signal is the recording of the original channel, in which no epileptic event is recorded, and blue is its modification with DCT

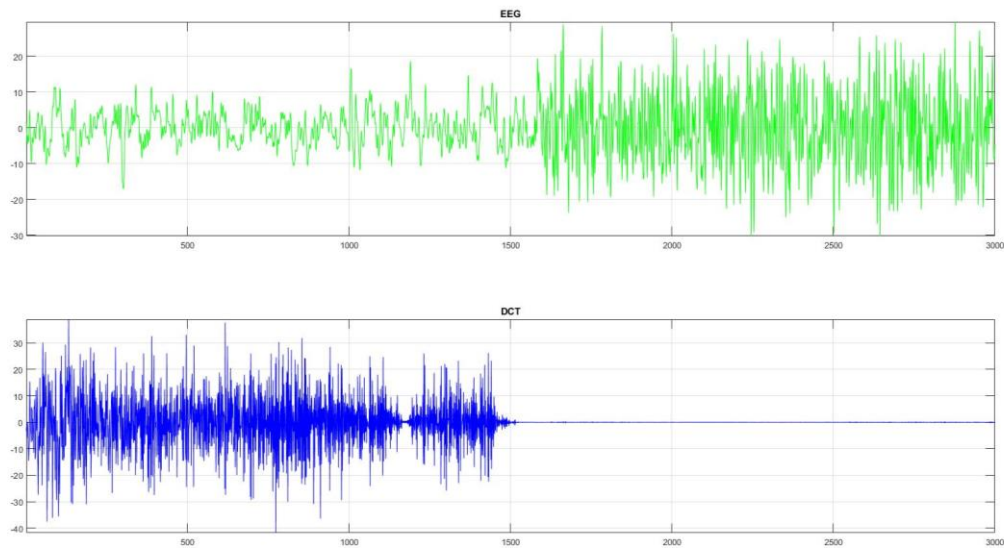


Figure 16: Ictal EEG (Above) and DCT of EEG (Bellow)

The green signal is the recording of the original channel, in which an epileptic event is recorded, and in blue is the modification of DCT

We notice that our new signal has several values around zero. If, for example, we have a signal segment of approximately 12 minutes (N) (190.000 samples), then to maintain 99.5% of the total energy and reset the rest of the information, we have about fifty four thousand (N99.5) (54,000) non-zero values, about the 28% of the original information.

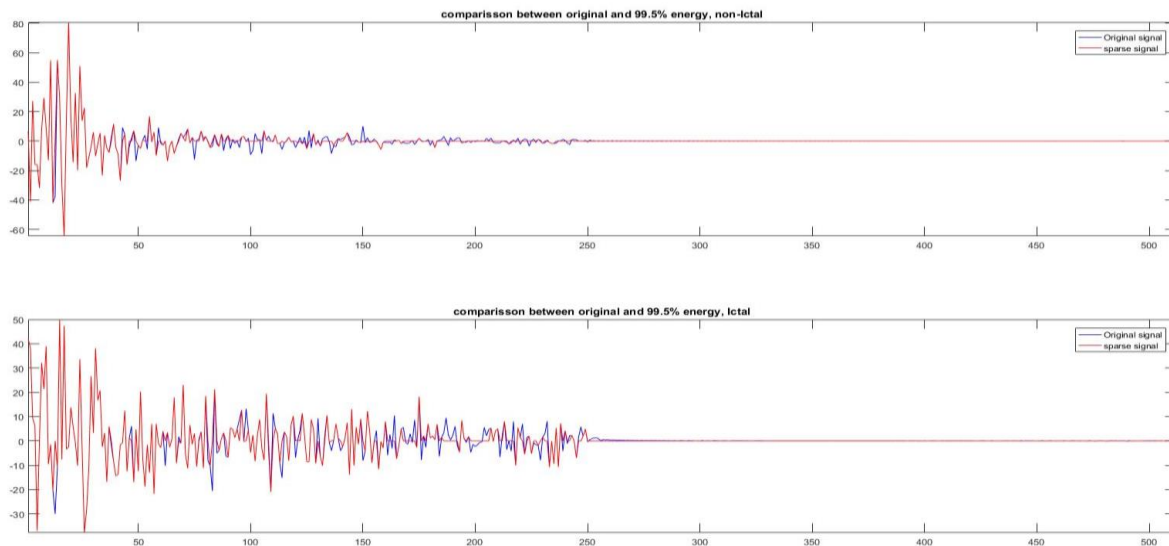


Figure 17: 100%,99.5% Energy for inter-Ictal and Ictal DCT

With blue is represented the 100% of the energy, while in red is represented 99.5%

Depending on the energy we want to keep, the number of non-zero elements is fluctuating, for example. If we want to have 99.9%, the number is increased to sixty-eight thousand (68.000) and if we want 99% to forty-five thousand (N99)(45.000) non-zeros data. The important question that arises is: What is the acceptable loss of information to continue to have meaningful information about brain activity?

For this purpose, we have been testing enough to be able to decide on this percentage.

To find the right percentage, we compared the energy-reduced signal to the original to see what we're missing out on - what we earn. At first, we tried 99.5% of the initial energy. The results were as follows:

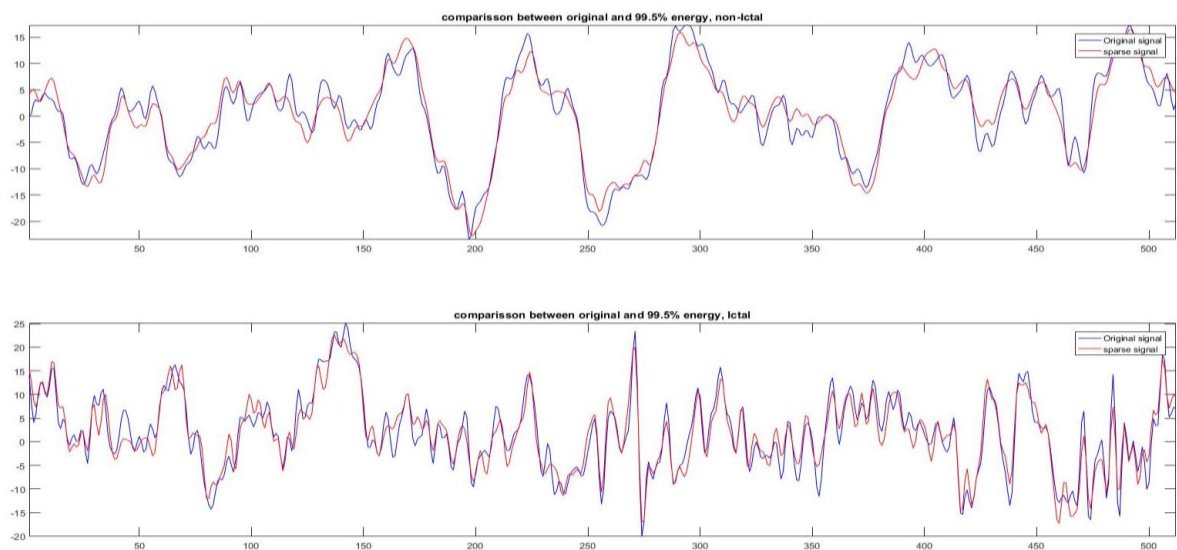


Figure 18: Inverse Transform 99.5% Energy, inter-Ictal(Above)-Ictal(Below)

In which, initially, there is no significant reduction, but if we focus on the point of epilepsy, we can see that we have a loss of intensity in the vertices, and the phenomenon intensifies in constant fluctuations.

Then we reduced energy to 99% of the original. The number of non-zero data was reduced to 24% of non-zero data but the lost information is now visible and pronounced at the point of epilepsy.

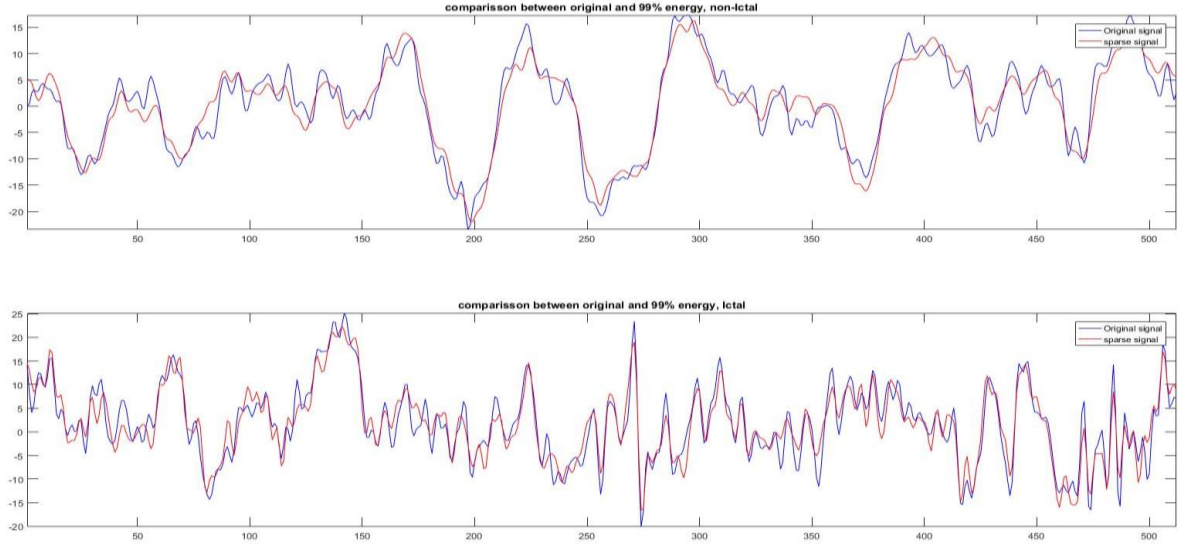


Figure 19: Inverse Transform 99% Energy, inter-Ictal(Above)-Ictal(Below)

Therefore, this price was also rejected. Other tests were performed with lower values to see the consequences, but the loss of information was too great, and epilepsy was almost normalized. Finally, we checked to maintain 99.9% of the total energy that raised the number of non-zero data to 36% of the original data and the loss of information was now very low. In channels with no intense seizure, the aforementioned numbers of data are valid, on the contrary, in the case of record-seizure channels, the figures number of non-zero elements rise by a little.

These tests were performed over the whole range of the signal, i.e. in N . But for the application of compressed detection we need to create some tables, as we have seen from the theory of the method.

In our case and due to the fact that our initial signal is not sparse, we have the second case, where for Sensing matrix we have an independent and identically distributed Gaussian matrix (iid Gaussian) and an Inverse Discrete Cosine Transform Matrix (IDCT) .

So that Φ^*c gives our original signal. From the theory of linear algebra it results that the dimensions of the matrices will be: Φ [length c , length c] and A [length y , length c]. This, however, results in a signal of one hundred and ninety thousand (190,000) elements requiring a matrix of Φ thirty six billion one hundred million (36,100,000,000) elements which will require excessive computer memory, which makes it impossible to apply to even a supercomputer. This led us to the implementation of smaller matrices. The N -size signal is now broken into parts of thirty thousand (P) (30,000) elements. In this way we ensure that the total size of the sensing matrix and the frame is limited to editable Giga Bytes. It is noted that the

ratio of non-zero elements in relation to the totality of the data remains approximately the same, with some fluctuations depending on the information contained.

Here is the whole process that was implemented in this thesis

First is the original EEG signal (**green**), then we have the DC transform (**blue**) the signal

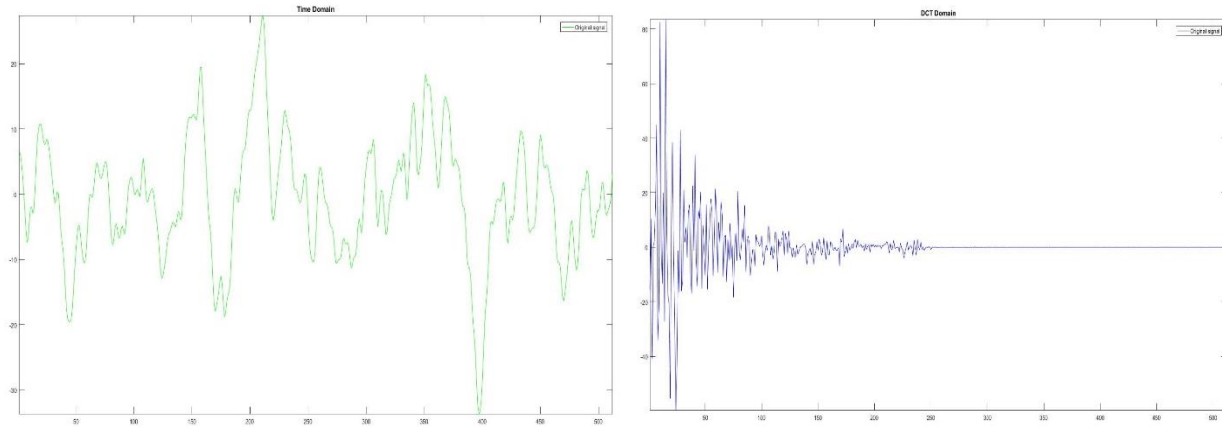


Figure 20: original EEG (green) - DCT (blue)

Next is the reduced DCT signal (**red**)(99.5% of the original DCT (**blue**)) and the compressed signal (**black**) after applying the compressed sensing method ($y = \Phi \times \Psi \times x = \Theta \times x$, where Φ the i.i.d. random gaussian, Ψ the DCT basis and x the reduced DCT signal (**red**))

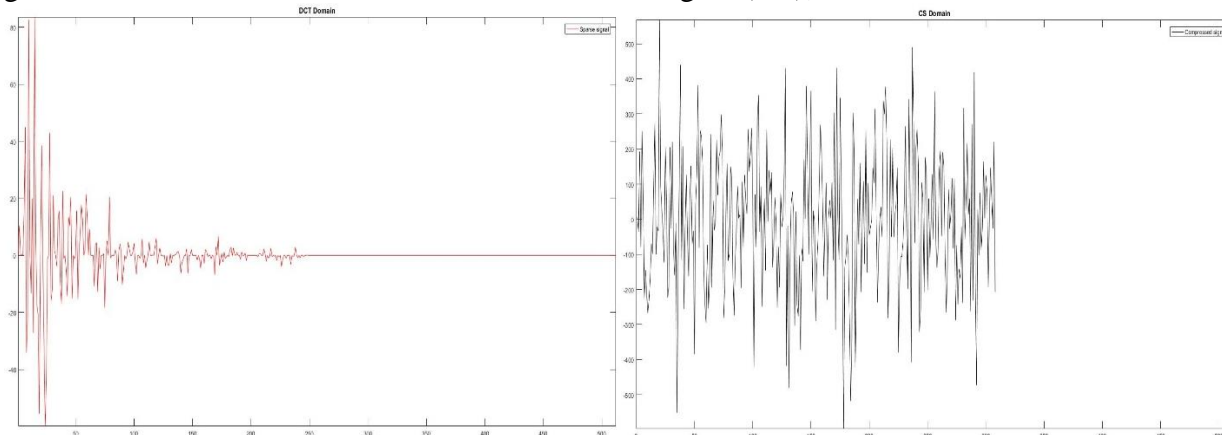


Figure 21: reduced DCT (red) - Compressed signal after CS (black)

Finally, we have the algorithmic reconstruction of the signal (**dark red**) using the BP algorithm and the inverse DC transform (**time domain**). If the reconstruction was without further loss of information then the reconstructed signal (**dark red**) and the reduced DCT signal (**red**) will be the same. We can not recover the original DCT signal, because after the energy reduction the constants that were removed are lost.

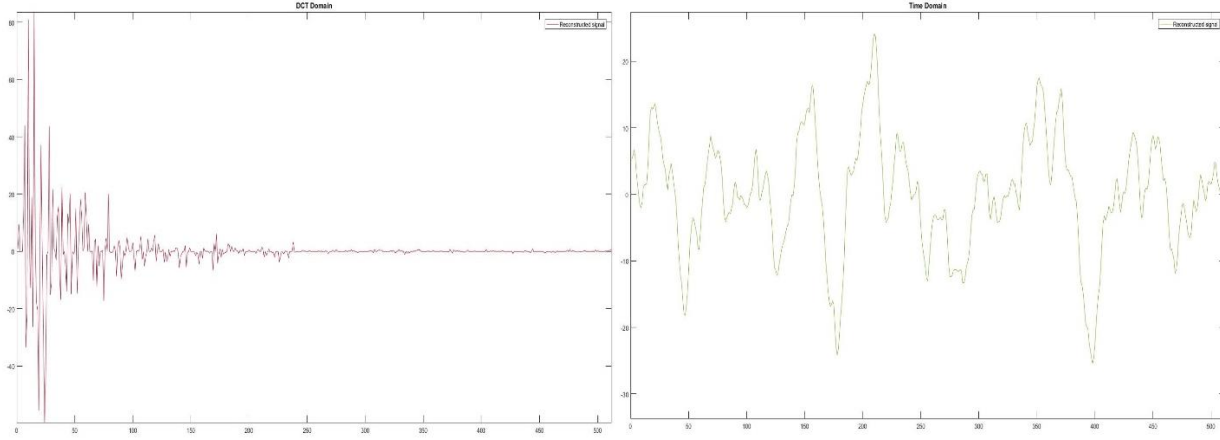


Figure 22: Reconstructed signal (dark red) - inverse DCT signal (light green)

Frequency Domain Reconstruction (Fourier)

We then checked the frequency of the compressed signals to see the frequency changes. The test was performed in 2 steps: first across the data range and then at selected points where the intensity of the modified signal was observed. In the first case (across the range) there was no significant difference but only a reduction in energy in the frequency domain.

Mean Square Error (MSE) is calculated from the formula

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

Equation 7: Mean Square Error (Above) and Root Mean Square Error (Below)

Where Y_i is our original signal and \hat{Y}_i is the signal after the energy deduction.

Mean square error averaged 8.8 for crisis recording intervals and 6.1 for non-crisis intervals, and RMSE averaged 2.7 and 2.4 respectively.

In the second case, the points we checked were selective and related to high frequencies and oscillations on one hand and points with normal frequencies and oscillations on the other. The test was

performed in a 4-second window, i.e. 1024 samples (256Hz EEG sampling frequency * 4), for all three data reduction cases, 99%, 99.5% and 99.9% of the total initial energy. Below are illustrative graphs:

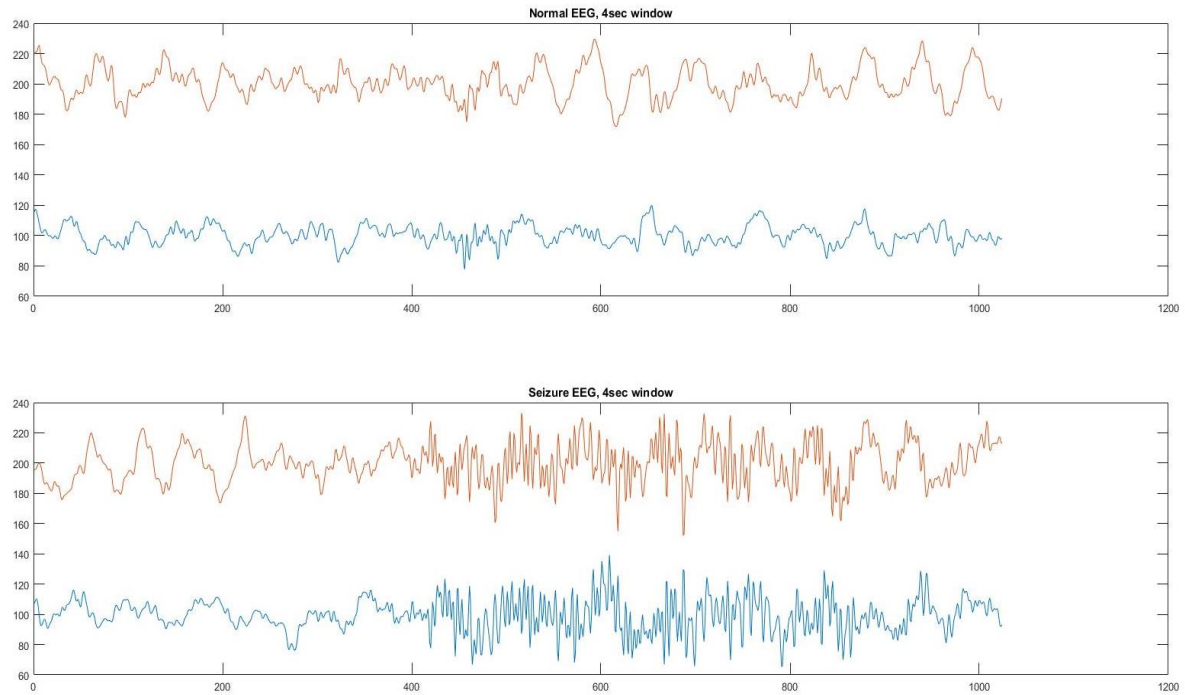


Figure 23: inter-Ictal EEG (Above) and Ictal EEG (Below)

The graph shows the 4sec windows for segments with ictal and inter-ictal recordings.

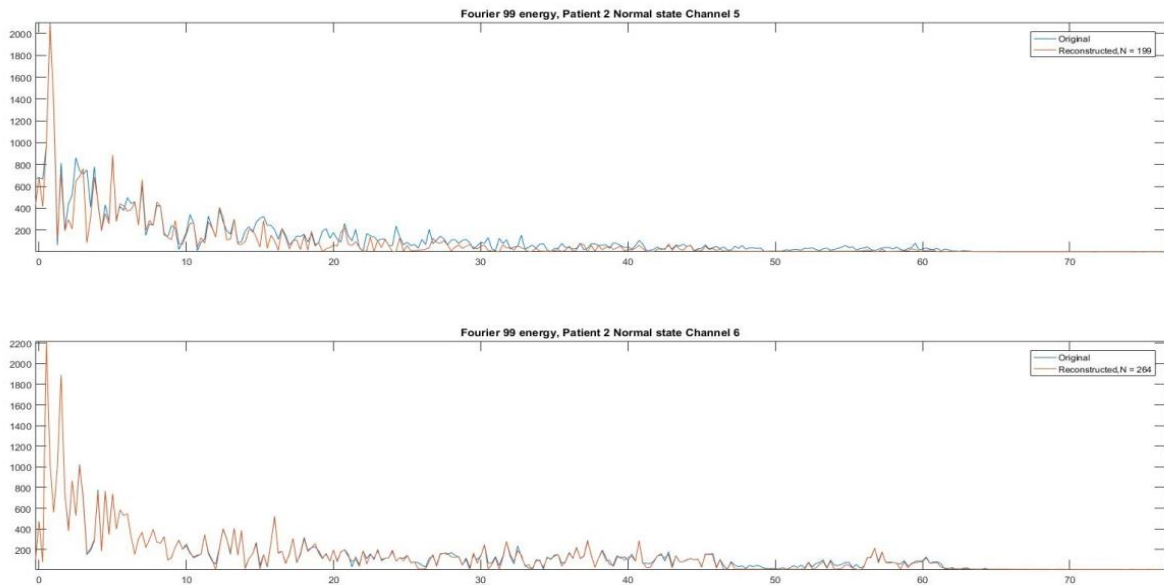


Figure 24: 100%, 99% Comparison in Frequency Spectrum

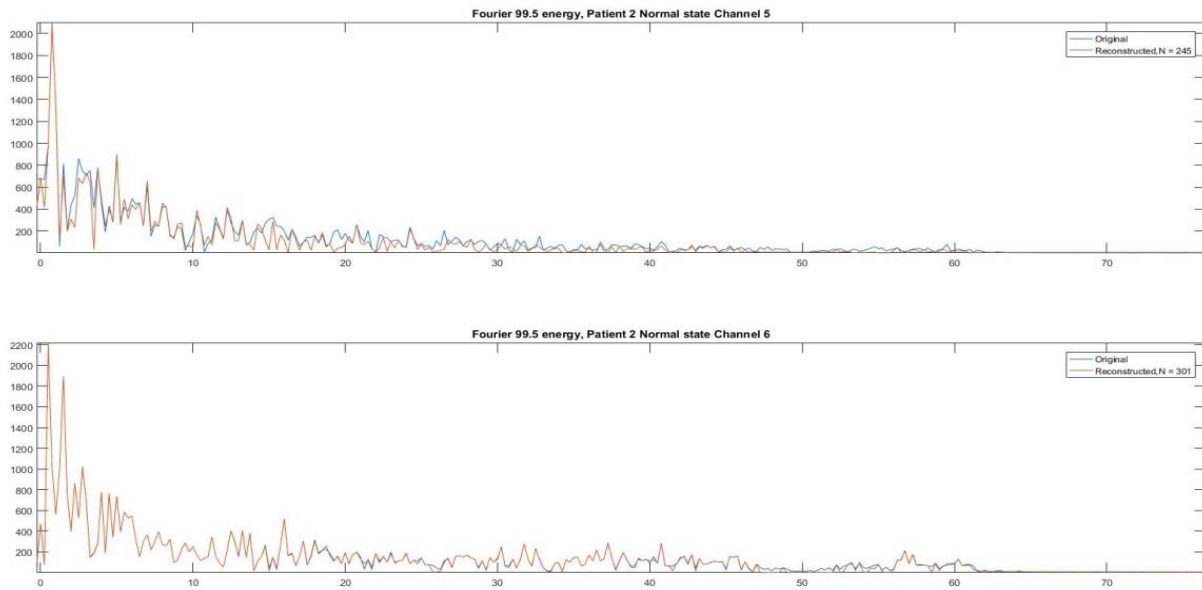


Figure 25: 100%, 99.5% Comparison in Frequency Spectrum

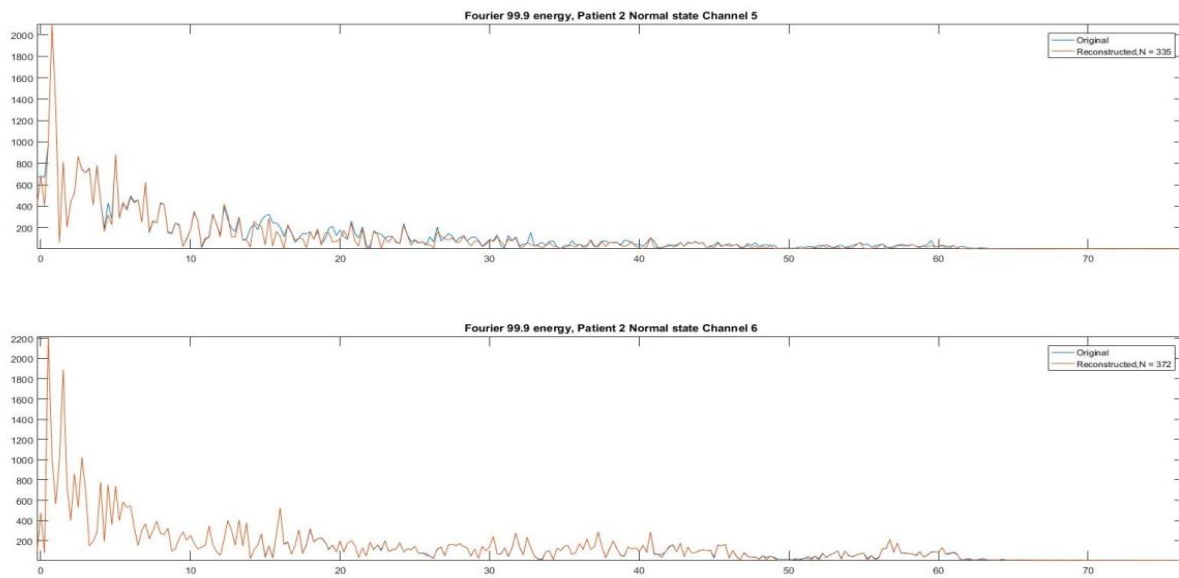


Figure 26: 100%, 99.9% Comparison in Frequency Spectrum

Then we present three images showing the Fourier Transform in 4 second windows and at a time before the start of the crisis. The results show that channel frequencies remained virtually the same. There is a reduction in the energy of these frequencies, which is normal, since we have reduced the energy of the signals. Greater loss was the first picture, which accounts for 99% of the initial energy, then the second with 99.5% and last with 99.9%.

We then provide the Fourier Transform for a point where there is judgment for all three cases.

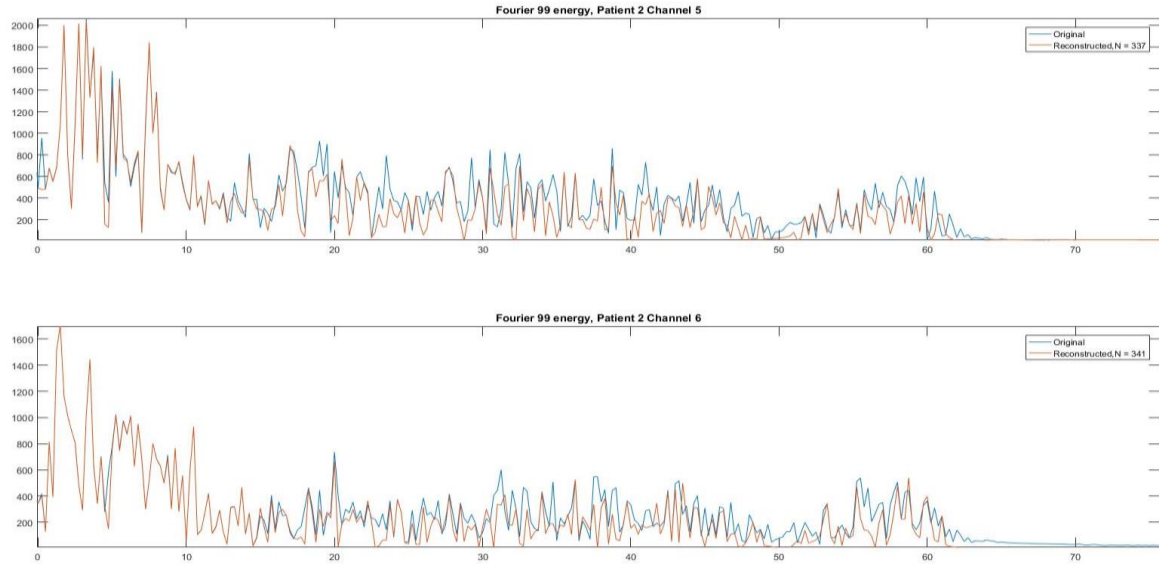


Figure 27: 100%, 99% Comparison in Frequency Spectrum

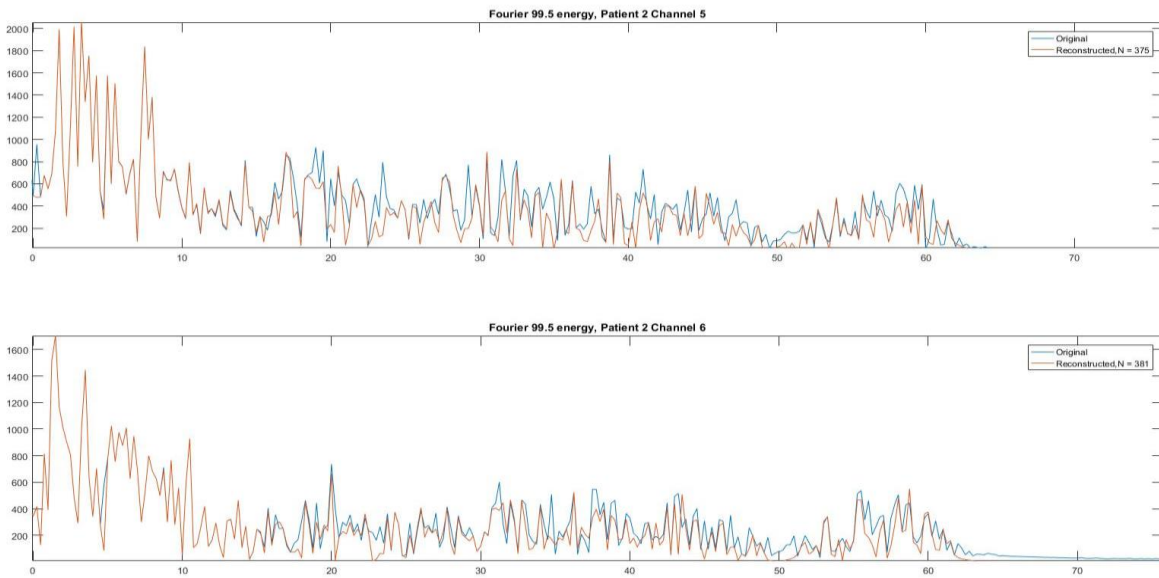


Figure 28: 100%, 99.5% Comparison in Frequency Spectrum

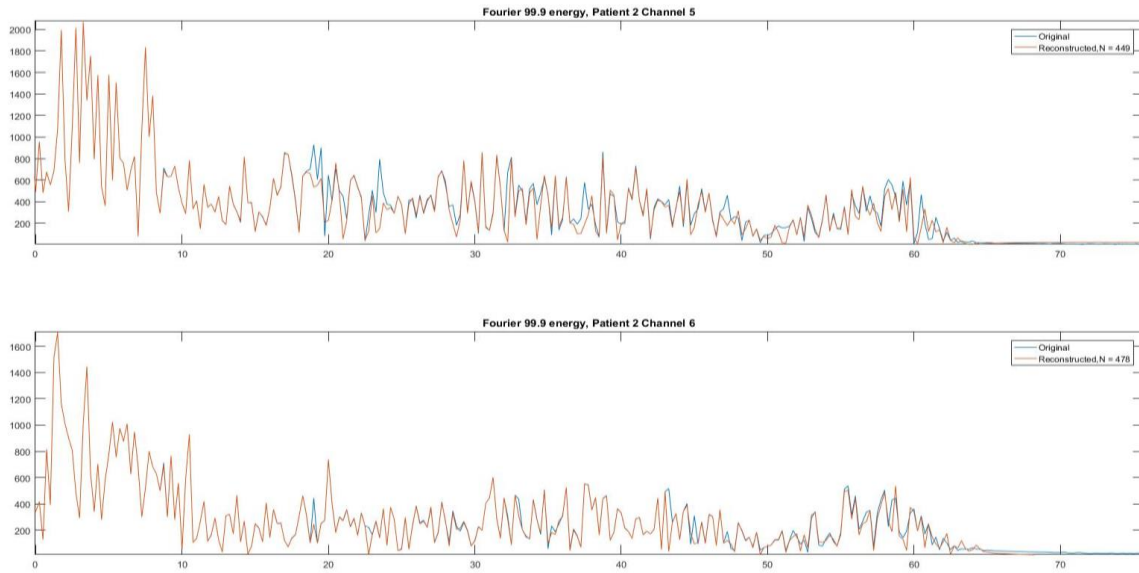


Figure 29: 100%, 99.9% Comparison in Frequency Spectrum

The results are like the previous ones, with the difference that the loss phenomenon focuses on high frequencies.

This research has shown us that after the information is reduced, the frequencies of the original signal are not lost. This information alone does not give enough information as to whether the important epileptic information has actually been maintained. For this another parameter, Approximate Entropy (ApEn) was tested.

We then estimated the approximate entropy size for various compression steps to control the performance achieved in crisis detection.

Reconstruction based on Approximate Entropy (ApEn)

Approximate entropy (ApEn) is a newly developed statistical quantification of regularity and complexity, which seems to be likely to apply to a wide variety of relatively small and noisy data. Approximate Entropy (ApEn) was suggested by Pincus [11]. It is essentially an improvement of the Kolmogorov-Sinai (K-S entropy) [12], which could not be used in statistical problems while there were significant discrepancies when there was noise in the time series to be examined. Pincus for the first time applied his method of estimating their regularity heart rate in healthy and diseased newborns [13]. In general, cardiology is one from the areas the method is widely applied [14]. Quickly however, the method has been applied to other physiological problems and areas, such as endocrinology but also neurology for analysis electroencephalograms [15]. Approximate entropy estimates the regularity of a time series in the sense of identifying similar patterns of observation groups that reappear within the time series.

Specifically, for each window of a specific length, the algorithm searches for the frequency of appearance of a sample sequence within the time series. The reappearance of a sequence is considered to occur when the differences of the observational groups do not exceed a certain threshold. The sequences that are similar (are smaller than the threshold) are involved in the creation of a quotient which then gives the value of approximate entropy. A low entropy value indicates that the time series are deterministic, while a high value indicates randomness. The development of ApEn was caused by the data length limitations commonly encountered, e.g. heart rate, EEG. The advantages of ApEn include:

Lower computational requirement: ApEn can be designed to work for small data samples and can be implemented in real time.

Less impact from noise: If the data is noisy, the ApEn measure can be compared to the noise level in the data to determine what quality of real information can be present in the data.

Details of the process are described as follows with the example of a single-channel EEG segment. $U=[u(1),u(2),...,u(N-1),u(N)]$ that includes N measurements. Define the parameters of the algorithm as, the parameter m of the number of sample points to be used for the comparison and the parameter r which is the similarity threshold between the groups of observations. Based on these parameters the calculation of the ApEn consists of the steps:

The vector sequence $x(i)=[u(i),u(i+1),...,u(i+m-1)]$, $i=1,2,...,N-m+1$ is formed. These vectors contain m consecutive values of the sequence U starting from the i^{th} point.

Defined as $d[x(i),x(j)]$ the distance between the vectors $x(i)$, $x(j)$ as the maximum absolute difference between the m vector components of the two vectors according to the equation

$$d[x(i),x(j)] = \max_{1 \leq k \leq m} \{|x(i+k-1) - x(j+k-1)|\}$$

For each vector $x(i)$, the number of indices $j(1 \leq j \leq N-m+1)$ is calculated for which the distance

$d[x(i),x(j)]$ is less than or equal to the filter threshold r , that is: $d[x(i),x(j)] \leq r$

If $N_i^m(r)$ the number of vectors that satisfy the above relationship, for all $i=1...N-m+1$ the parameters m are calculated such as:

$$C_i^m(r) = \frac{N_i^m(r)}{N-m+1}$$

The values $C_i^m(r)$ indicate the frequency with which the patterns resemble vector $x(i)$ appear in the U timeserie, with tolerance due to filtering on level r . Calculate the mean $\varphi^m(r)$ of the logarithmic values of the parameters C_i^m according to the equation

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

The above steps are repeated using as input parameter the value $m + 1$. That is, we are gradually calculating the new vector sequences, the values C_i^{m+1} and finally, the value φ^{m+1} . The ApEn is ultimately defined by the following relation:

$$ApEn = \varphi^m(r) - \varphi^{m+1}(r)$$

Equation 8: Approximate Entropy

Comparison window length m : This parameter specifies the number of observations group that are compared with the same number of observations in the rest of the series. The bigger is the comparison window, the more difficult it is to find groups of observations in the time series approaching the comparison window.

Filter threshold r : This parameter determines the similarity tolerance between a pair of windows that are compared. When comparing two windows we consider that they are similar if the differences in their components are no greater than the filter threshold r . The threshold is given as a function of the standard deviation of the time series because the size must be comparable to the variation of time series. Large values of r denote high similarity tolerance while low values r indicates the strict version with low tolerance on the similarity of comparison windows.

Length of time series N : This is the number of measurements that constitute the under-consideration time series. ApEn is essentially a statistic, so to have reliable results it needs to have a sufficient amount of measurements. In practical applications, ApEn have been used for $N > 60$. In this particular work the length of the EEG recording time series that is used is $N = 512$ measurements (2 seconds window).

To check if there was a substantial change in the original information, we applied ApEn to the complete sample, then to the reduced and to the reconstructed. Below we list a channel with a strong epileptic event. The first 10 minutes are Inter-Ictal (1-300, 300 segments) and the last 80 seconds are Ictal (301-340, 40 segments).

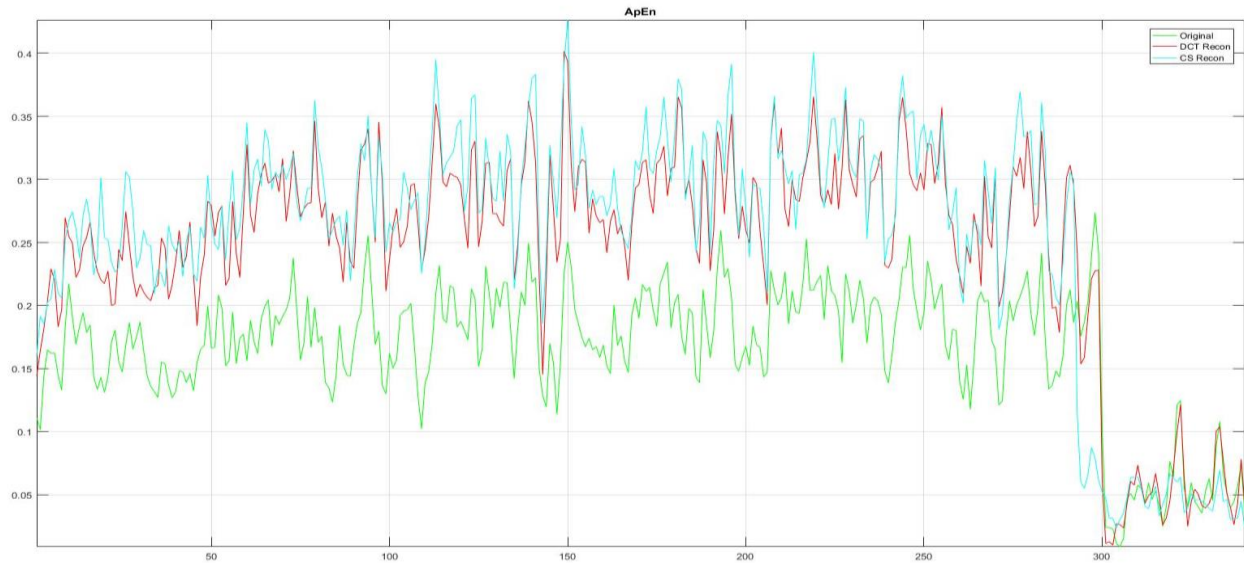


Figure 30: Approximate Entropy (ApEn)

The graph above shows the original channel in green. It is noted that in the normal state the values are between 0.13 and 0.25 while at the crisis point there are at 0.01 and 0.12. Then we get the results after the information is reduced through the DC Transform, red color. ApEn values show a slight increase of 0.05 in the reference values at normal points, whereas in crisis the increase was 0.01. However, the general form of ApEn remains the same. The energy of the above signal amounts to 99.5% of the original.

Finally, entropy on the signal was checked after using compressed sensing. The signal was broken into pieces and then reunited. After the merger, ApEn was tested. Note that the form remained the same, except that we had a normal increase of 0.1 while there was a decrease in rates during epilepsy. This is due to the fact that there has been a slight loss of information in the recovery process.

The other two energy reductions have been tested, but we have decided to select 99.5% of the total energy for further research as it has a good retention of information after use of CS and the number of non-zero data is 1 / 3 of the original. This means that it gives a good compression of data but at the same time it keeps the information of the crisis.

References

- [1 E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, pp. 21-30, 2008.
- [2 D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289-1306, 2006.
-] [3 B. Keaton, "Sampling theory Fourier theory made easy," [Online]. Available: <https://www.slideserve.com/barton/sampling-theory-fourier-theory-made-easy>.
- [4 M. W. Fakhri, "Researchgate," March 2017. [Online]. Available: https://www.researchgate.net/publication/314839832_Joint_Image_Compression_and_Encryption_Based_on_Compressed_Sensing_and_Entropy_Coding/figures?lo=1.
-] [5 S. Bhattacharya, "Compressed Sensing and Control," [Online]. Available: http://web.me.iastate.edu/sbhattac/research_cs.html.
- [6 R. G. Baraniuk, "SlidePlayer," [Online]. Available: <https://slideplayer.com/slide/5237107/>.
-] [7 E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *C. R. Acad. Sci. I*, vol. 346, pp. 589-592, 2008.
-] [8 E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203-4215, 2005.
-] [9 D. L. Donoho, Y. Tsaig, I. Drori and J.-L. Starck, "Sparse Solution of Underdetermined Linear Equations by Stagewise Orthogonal Matching Pursuit," Preprint, 2007.
-] [1 S. S. Chen, D. L. Donoho and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, pp. 33-61, 1998.
-] [1 S. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Natl. Acad. Sci. USA*, vol. 88, pp. 2297-2301, 1991.
-] [1 B. Bein, "Entropy," *Best Practice & Research Clinical Anaesthesiology*, vol. 20, no. 1, pp. 101-109, 2006.
-] [1 S. Pincus and E. Gladstone, "A Regularity Statistic For Medical Data Analysis," *Journal of Clinical Monitoring*, vol. 7, no. 4, 1991.
-] [1 T. Jartii, T. Kuusela, T. Kaila, K. Tahvanainen and I. Valimaki, "The dose-response effects of terbutaline on the variability, approximate entropy and fractal dimension of heart rate and blood pressure," *Br J Clin Pharmacol*, vol. 45, no. 3, pp. 277-285, 1998.
-] [1 S. Pincus and D. Keefe, "Quantification of hormone pulsatility via an approximate entropy algorithm," *The American Physiological Society*, vol. 262, no. 5, pp. 741-753, 1992.

Chapter Four: Analysis and Classification of epileptic seizures using Common Spatial Patterns

The aim of the noninvasive brain-computer interfaces (BCIs) is to translate brain activity of control sequences so that a subject, such as a person with disabilities, so that he will be able to commune with the outside world, such as a computer, without the usage of the peripheral nervous system. The wide usage of Electroencephalography for capturing the electrical potentials generated by the central nervous system and inferring the user's desires is because of the simplicity of the method, the inexpensiveness, and the high temporal resolution provided. [1]. Low signal-to-noise ratio (SNR) are the common trait of multichannel EEG signals and as a result they give a rather blurred image of the brain activity, making them unusable for BCI applications.

4.1 Common Spatial Patterns (CSP)

The common spatial patterns (CSP) method is a frequently utilized algorithm in order to extract the most discriminative information from EEG recordings. The first recommendation of the method was for the binary classification of EEG trials [2]. CSP is a spatial filtering method that seeks to optimize filters that will result in having the most difference in power/variance ratios in the feature space. The projections/filters are computed by a simultaneous diagonalization of the covariance matrices of the two classes that we want to individualize. The first few most discriminative filters are usually used for classification cause of the highest discrimination ability they procure. The reason is that the n th filter obtained has a relative variance of d_n for trials of class 1 and relative $1-d_n$ variance for trials of class 2. If it is close to 1, the filter maximizes variance for class 1, and since is close to 0, minimizes variance for class 2 and vice versa [3].

The Common Spatial Standard (CSP) is a signal processing method mainly used in EEG analysis to help distinguish between two categories of data by calculating a set of spatial filters that can maximize variation for a data category and at the same time minimize time. fluctuation for the other [4]. Specifically, the main use of this method is to retrieve the signals that best carries the brain activity for a specific task, such as the movement of a hand or foot, for Brain-Computer Interface (BCI) purposes. It can also be used to separate patterns from encephalographic signals [5].

Mathematically, the method relies on the simultaneous diagonalization of two matrices closely related to the covariance matrices [6]. To summarize briefly, given two distributions in a high-dimensional

space, the CSP algorithm finds directions (spatial filters) that maximize variance for one class and that at the same time minimize variance for the other class. [2].

The formal use of this method is by selecting windows that correspond to different activation sources, such as during the movement the hand and during the foot. Another possibility is the selection of windows that belong to two different frequency bands in order to identify components that have a particular frequency pattern. For efficiently using the CSP algorithm, several parameters have to be taken into accounts, such as the frequencies for band-pass filtering of EEG recordings and the time interval of the recordings taken on the stimuli. Usually, the frequency band is between 7-30 Hz for the encephalogram when the study is for BCI purposes. The reason being that the sensorimotor rhythms (SMRs) are oscillations at the aforementioned band frequency in the EEG recorded from the scalp. In our research, the bandwidth was from 0.5 Hz to 70 Hz, because epilepsy affects the whole spectrum.

Because of the dependence of the eigenvector matrix only on the covariance matrices acquired by the two classes, it is possible to obtain the same results if the process is applied in the frequency domain of the recording, namely after the use of the Fourier transformation on the recordings. None the less, it should be stated that the conventional CSP algorithm is relying on sample-based covariance matrix estimations and for that reason, the performance in EEG categorization can deteriorate when there is only a small quantity of samples available for training. Similar problems have arisen in a lot of other applications that rely on the quantity of the samples provided.

Mathematical Formulation

Common spatial pattern method was firstly suggested for classification of multi-channel EEG for performing one of 4 movements (pressing a micro-switch with the left or right index finger, flexing the toes of the right foot, or moving the tongue to the upper gum) by J. Müller-Gerking [2]. The main idea, behind the common spatial patterns, is to use a linear transformation to project the multi-channel EEG data into a low-dimensional spatial subspace with a projection matrix, of which each row is a filter that consists of weights for each channel. This transformation can maximize the variance of two-class signal matrices. The CSP method is based on determining projection matrix in such a way that the variance of its first row is maximal for the trials of class 1 and at the same time minimal for the trials of class 2. The last row is the opposite of the first meaning, it has maximal for the trials of class 2 and at the same time minimal for the trials of class 1.

Details of the process are described as follows with the example of classifying single-trial EEG for class 1 and class 2. Class 1 or 2 can be any to two different cases that we want to classify. X_1 and X_2 denote

the preprocessed EEG matrices under two conditions (e.g. hand and foot movement, left hand right hand movement) with dimensions $M \times N$, where M is the number of channels and N is the number of samples per channel. The normalized spatial covariance of the EEG can be represented as:

$$COV_1 = \frac{X_1 X_1^T}{\text{trace}(X_1 X_1^T)} \quad COV_2 = \frac{X_2 X_2^T}{\text{trace}(X_2 X_2^T)}$$

X^T is the transpose of X and $\text{trace}(A)$ computes the sum of the diagonal elements of A . The averaged normalized covariance $\overline{COV_1}$ and $\overline{COV_2}$ are computed by calculating the average values for all the trials of each class separately. The composite spatial covariance can be factorized as

$$COV = \overline{COV_1} + \overline{COV_2} = U_0 \times D \times U_0^T$$

where U_0 is the matrix of the eigenvectors and D is the diagonal matrix of the eigenvalues. The whitening transformation matrix is then defined as

$$P = D^{-1/2} \times U_0^T$$

And transforms the average covariance matrices as

$$S_1 = P \times \overline{COV_1} \times P^T \quad S_2 = P \times \overline{COV_2} \times P^T$$

S_1 and S_2 share the same eigenvectors and the aggregation of the corresponding eigenvalues for the two matrices will always be one,

$$S_1 = U \times D_1 \times U^T \quad S_2 = U \times D_2 \times U^T \quad D_1 + D_2 = I$$

The eigenvectors with the largest eigenvalues for S_1 have the smallest eigenvalues for S_2 and vice versa.

The transformation of whitened EEG onto eigenvectors corresponding to the largest eigenvalues in D_1 and D_2 is optimal for separating variance in two signal matrices. The projection matrix W , that also consist of the spatial filters is denoted as

$$W = U^T \times P$$

With the projection matrix W , the original EEG can be transformed into uncorrelated components

$$Z = W \times X$$

Z can be seen as EEG source components including mutual and unique components of different classes.

The original EEG X can be reconstructed by

$$X = W^{-1} \times Z$$

Where W^{-1} is the inverse matrix of W . The columns of W^{-1} are spatial patterns, which can be considered as EEG source distribution vectors. The first and last rows of W are the most important spatial filters that provide the largest variance of one class and the smallest variance of the other [2]. The explanation is below:

$$W^T \times \overline{COV_1} \times W = D_1 = \begin{pmatrix} d_1 & & \\ & \cdot & \\ & & \cdot \\ & & & d_{ch} \end{pmatrix} \quad W^T \times \overline{COV_2} \times W = D_2 = I - D_1 = \begin{pmatrix} 1-d_1 & & \\ & \cdot & \\ & & \cdot \\ & & & 1-d_{ch} \end{pmatrix}$$

Where $d_1 \geq d_2 \geq \dots \geq d_{ch}$, therefore, the first CSP filter (W_1) provides maximum variance for the class 1 and the last CSP filter (W_{ch}) provides the maximum variance for the second class.

4.2 Time window determination

In this section, we discuss about the time window determination. The length of the time window is an import factor when dealing with seizures. If you have a relative big window in comparison with the seizure then, the non-seizure samples of the window may have a negative effect in classifying the segment as ictal or inter-ictal. On the other hand, if the window is too small then, the information included may be unable to characterize the segment as ictal. So what consist a good time window for seizure detection? The window length m , specifies the number of observations group that are compared with the same number of observations in the rest of the series. Based on previous researches, a seizure can last from 4-5 seconds to 1-2 minutes. In order, to be able to have segments that consist purely of seizure data or at the very least, be more than the inter-ictal samples, we chose a 2 second time window, meaning 512 measurements-samples, because the sampling frequency used at the Hospital was 256Hz. With the help of a doctor we confirmed that the 2 second where enough to classify the segment as ictal. Another factor in determining the duration of the window was the total amount of seizures we had to conduct our research. The more sections to check and check, the more accurate the results of this research. The total amount of ictal segments, after determining the 2 seconds time window, was nearly one thousand segments. The original length was 4 seconds (1024 measurements), but the total amount of ictal segments was around 500, so in order to have more segments available, we halved that time. The procedure of creating the segments consist of a sliding temporal window of $\Delta t=2\text{sec}$ and a step of 0.5sec.

4.3 Frequency Analysis of EEG for Common Spatial Patterns filters (CSP)

The Common Spatial Pattern (CSP) algorithm is applied to the data, in order to categorize the segments between Ictal and Inter-Ictal, by maximizing the variances of one class and minimizing for the other. To use the algorithm, we use 2-second segments (512 samples) and group them into a three-dimensional array, the sides being: the data (2 seconds), the encephalogram's channels, and finally the segments to be checked.

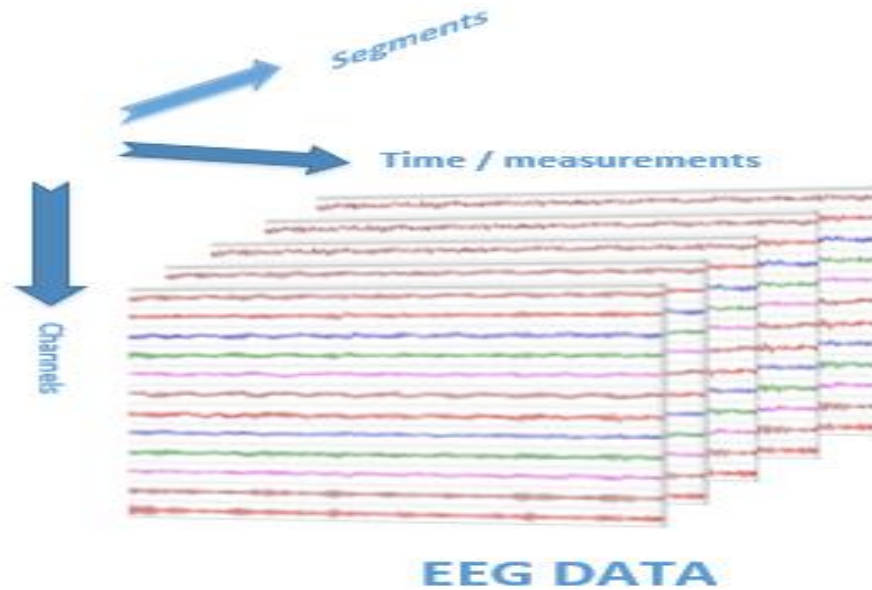


Figure 31: Data Organization for CSP

Then for each Patient, a sliding temporal window of $\Delta t = 2\text{sec}$ and a step of 2 sec was used. There is no overlap between the segments and they are continuous without significant gaps. Then we decompose each segment into the δ {0.5-4 Hz}, θ {4-8 Hz}, α {8-13 Hz}, β {13-30 Hz} and γ {30-70 Hz} rhythms. in order to investigate the spectral content of each sub-band. For each temporal window, the 5 EEG rhythms (δ , θ , α , β , γ) timeseries were extracted using the corresponding bandpass filtering on the aforementioned frequency bands. The filter used in this thesis is a finite impulse response (FIR), non-causal and with zero phase filter. The brain rhythms have specific bandwidth and they are continuous, we cannot afford to have a high roll off factor, because we will be overlapping the neighbor rhythms. For that we have a roll off area of 0.5Hz, to ensure the individuality of each rhythm-band. An example of the filtering procedure and rhythm decomposition is presented in Figure 2, where the full spectrum of the signal is displayed at the top. Then follows from top to bottom each rhythm. First, we have the δ rhythm between 0.5-4Hz, then θ {4-8Hz}, α {8-13 Hz}, β {13-30 Hz} and finally, γ {30-70 Hz} rhythm. It can be seen

that there has been used a bandstop filter around 50Hz, because of the fact that the machines operate with 50Hz frequency current provided by the Public Power Corporation of Greece, the machines create a very intense noise distortion, so we are obliged to use a bandstop filter between 48-52Hz.

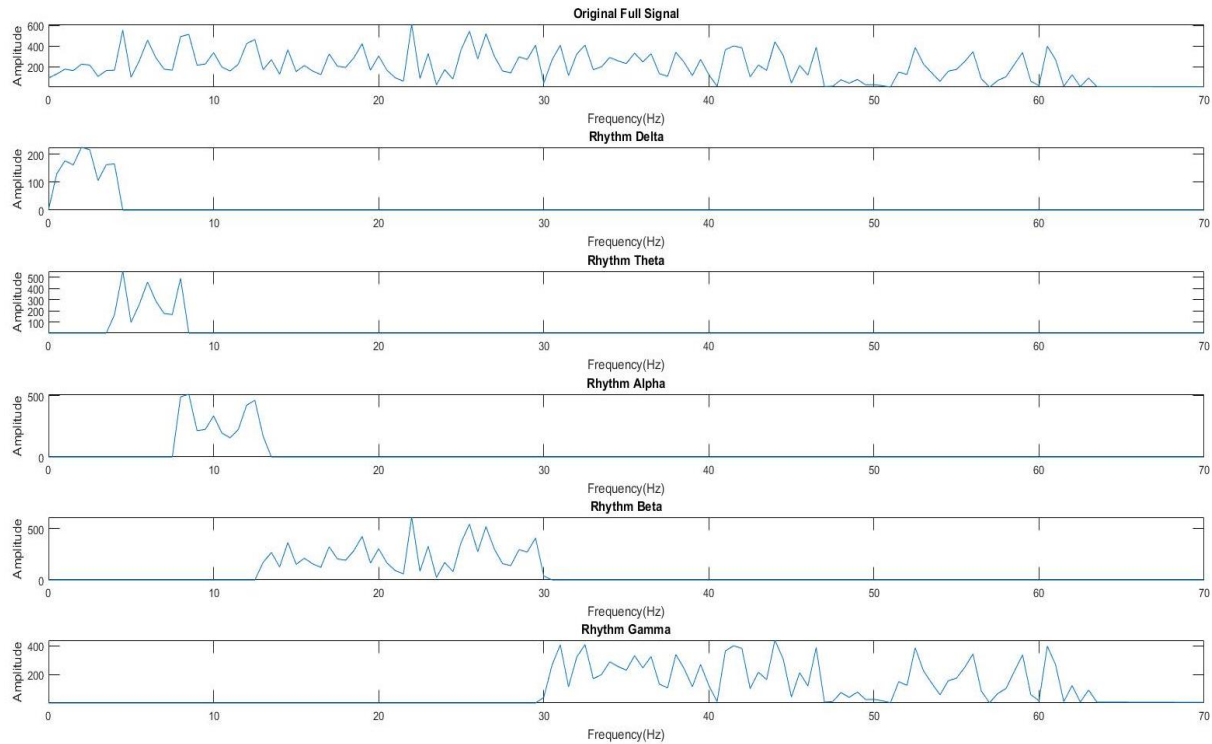


Figure 32: Spectrum of Each Rhythm

Here are the individual spectrums of each rhythm and the total initial spectrum of the signal. As can be seen, if we add the individual spectrums, then the initial one is obtained. In the next graph, we can see our data, for a specific segment, after the bandpass filtering is applied. As before, we follow the same order of illustration, at the top we have the original signal, next is the δ {0.5-4 Hz}, then θ {4-8Hz}, α {8-13 Hz}, β {13-30 Hz} and finally, γ {30-70 Hz} rhythm. After we create the specific data, for each segment, we continue to the feature extraction.

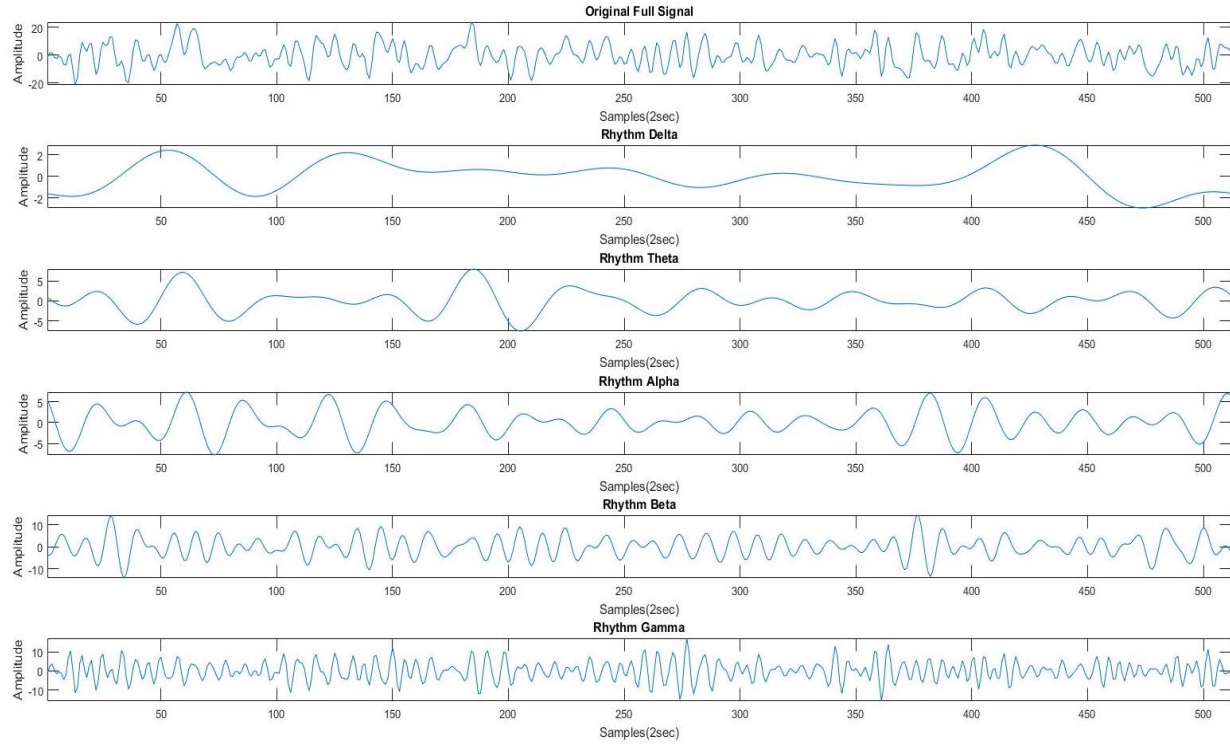


Figure 33: Time Domain of Whole Spectrum and Rhythms

4.4 Feature Extraction Using Common Spatial Patterns filters (CSP)

After the extraction of the frequencies, we proceed with the feature extraction using the CSP algorithm. First, we need to calculate the covariance matrix.

$$S = \frac{E \cdot E^T}{\text{trace}\{E \cdot E^T\}}$$

$\frac{E \cdot E^T}{\text{tr}(E \cdot E^T)}$, where E: matrix M*N (M-channels, N samples) for each segment

$\text{trace}\{E \cdot E^T\}$ = the sum of the diagonal elements of $(E \cdot E^T)$, which is also the sum of $(E \cdot E^T)$'s eigenvalues

This happens for each segment and then we calculate the average spatial covariance matrix for each class separately.

$$\overline{S_c} = \frac{1}{K} \sum_{k=1}^K S_{(c,k)}$$

, c: class, k: trial and K: total trials for the class c

Then we add the two tables and get the corresponding eigenvectors and eigenvalues of the whole.

$\overline{S}_1 + \overline{S}_2 = U \times \Lambda \times U^T$, where U the eigenvectors matrix and Λ the diagonal matrix of corresponding eigenvalues. For the construction of the final projection matrix, several matrices need to be calculated.

Initially we have $P = \Lambda^{-1/2} \times U^T$ then we get the average spatial covariance matrices and we calculate

the $\Sigma_1 = P \times \overline{S}_1 \times P^T$ and $\Sigma_2 = P \times \overline{S}_2 \times P^T$ matrices. Furthermore $\Sigma_1 = B \times \Lambda_1 \times B^T$ where B is the

eigenvectors matrix and Λ_1 the diagonal matrix of corresponding eigenvalues for Σ_1 . The final

projection matrix is defined as $W_0 = B^T \times P$.

The first and last columns of W_0 are the most important spatial patterns that explain the largest variance of one task and the smallest variance of the other.

For each patient, the overall departments train the algorithm in 2 classes (Ictal and Inter-Ictal). Next, an M-channel by M-channel matrix is generated which is used to extract the final 19 CSP features for each segment. Specifically, for feature extraction, a segment E is first projected as $Z = W_0^T \times E$.

Then, a N-dimensional feature vector y is formed from the variance of the rows of Z as

$$y_n = \log \left(\frac{\text{var}(Z_n)}{\sum_{n=1}^N \text{var}(Z_n)} \right)$$

Where y_n is the n^{th} component of y , Z_n is the n^{th} row of Z , and $\text{var}(Z_n)$ is the variance of the vector

Z_n . The transformation to logarithmic values is done in order to make the distribution of the elements in

y_n normal. The extracted features are equal to the number of the total number of common spatial filters, which is the same number as the number of channels available. In our study, the number of channels was 19, so the total amount of features for each rhythm and the whole spectrum was 19 measurements and a total of 95 for each segment. The features are then to be used as inputs to machine learning algorithms in order to determine if the segments belong to an ictal or an inter-ictal time window.

Data balancing procedure

The seizure detection problem is highly non-balanced, i.e. data from ictal periods are much fewer than inter-ictal periods, a balanced dataset should be ensured for formulating a proper model and increasing its performance. Thus, all samples from ictal and the same samples from inter-ictal periods were selected for the subsequent analysis. In our study, these procedures lead to a feature matrix X [984x95] of 984 time-window cases for each class and 19 channels x 5 rhythms = 95 features. Specifically, the columns of X are (19 features for δ , 19 for θ , 19 for α , 19 for β and 19 for total spectrum) and the rows represent each temporal window EEG segment.

Specifically the rows of the matrix Y represent the CSP features (19 features for Whole Spectrum - 19 features for Delta Rhythm - 19 features for Theta Rhythm - 19 features for Alpha Rhythm - 19 features for Beta Rhythm - 19 features for Gamma Rhythm), and the columns represent the 2 second Segments of the Encephalogram (Segment 1,...,Segment n). The first half of the total (1, 2, ..., $n/2$) Segments are inter-ictal recordings and the rest (the other half) ($n/2 + 1, +2, \dots, n$) Segments are Ictal recordings. Because of the balancing process the inter-ictal and ictal segments are equal in number, and $n=2 \times \text{ictal}$ or inter-ictal segments.

Matrix Y		Seg-1	...	Seg-n/2	Seg-n/2 +1	...	Seg-n
19 Features	WS	Inter-Ictal	Inter-Ictal	Inter-Ictal	Ictal	Ictal	Ictal
19 Features	DR	Inter-Ictal	Inter-Ictal	Inter-Ictal	Ictal	Ictal	Ictal
19 Features	THR	Inter-Ictal	Inter-Ictal	Inter-Ictal	Ictal	Ictal	Ictal
19 Features	AR	Inter-Ictal	Inter-Ictal	Inter-Ictal	Ictal	Ictal	Ictal
19 Features	BR	Inter-Ictal	Inter-Ictal	Inter-Ictal	Ictal	Ictal	Ictal
19 Features	GR	Inter-Ictal	Inter-Ictal	Inter-Ictal	Ictal	Ictal	Ictal

Table II: Feature matrix with the extracted features, consists of the patient-subject segments after the balancing procedure

Next we take the transpose of the matrix and we add an extra column which represents the class for each segment (1 for Inter-ictal segment and 2 for ictal segment). This column is to be used by the machine learning algorithm to estimate the accuracy of the classification, the specificity and the sensitivity. The new matrix is formed as follows and it is used for classification.

Matrix X	19 Features WS	19 Features DR	19 Features THR	19 Features AR	19 Features BR	19 Features GR	Class
Seg-1	Inter-Ictal	Inter-Ictal	Inter-Ictal	Inter-Ictal	Inter-Ictal	Inter-Ictal	1
...	Inter-Ictal	Inter-Ictal	Inter-Ictal	Inter-Ictal	Inter-Ictal	Inter-Ictal	1
Seg-n/2	Inter-Ictal	Inter-Ictal	Inter-Ictal	Inter-Ictal	Inter-Ictal	Inter-Ictal	1
Seg-n/2 +1	Ictal	Ictal	Ictal	Ictal	Ictal	Ictal	2
...	Ictal	Ictal	Ictal	Ictal	Ictal	Ictal	2
Seg-n	Ictal	Ictal	Ictal	Ictal	Ictal	Ictal	2

Table III: Final Feature Matrix for each Segment

Fig. 34 presents the most significant spatial filters learned by CSP for the first patient in our dataset and represent the most important spatial filters that explain the largest variance of inter-ictal and ictal class. Next we can see the results in the data after applying all the filters that we learned through the training of the CSP algorithm. The scalps show the mean values for all the inter-ictal and ictal data segments of the patient. The scale for the data is based on the mean highest and lowest values of the patient.

CSP maps

The CSP maps are constructed using the FieldTrip toolbox for EEG/MEG-analysis (REF; Donders Institute for Brain, Cognition and Behaviour, Radboud University, the Netherlands. <http://fieldtriptoolbox.org>) [7]. The figure 4 shows the most important spatial filter for each class, for the first patient of our study. To create these maps, we use the measurements of the first and last row of W and with the use of the fieldtrip toolbox we create the scalp maps. Each measurement represents a weight for each of the 19 channels. The left represents the filter that maximizes the variance for inter-ictal segments and minimizes the variance for ictal and the right vice versa. The first patient has focal epilepsy in the right frontal lobe. The right scalp CSP map shows similar results to the doctor's deduction. The figure 5 shows the CSP feature map for the inter-ictal and the ictal classes. The specific feature maps are created from averaging the features of each class and can serve as a reference for the application of the spatial filters as we can notice that the values of the ictal period are higher than the corresponding mean values of the inter-ictal features. The features being averaged are from the whole spectrum of the patient and not from a particular rhythm. Similar maps can be obtained from every segment and rhythm of each patient.

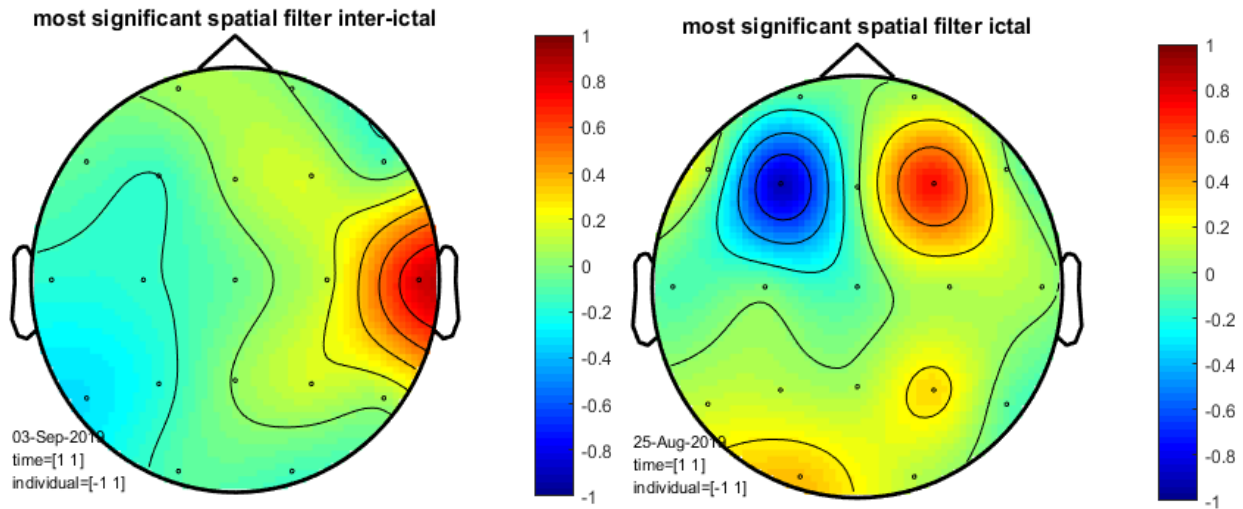


Figure 34: Inter-ictal (left graph) and ictal (right graph) most significant spatial filters for the patient (PAT_13)

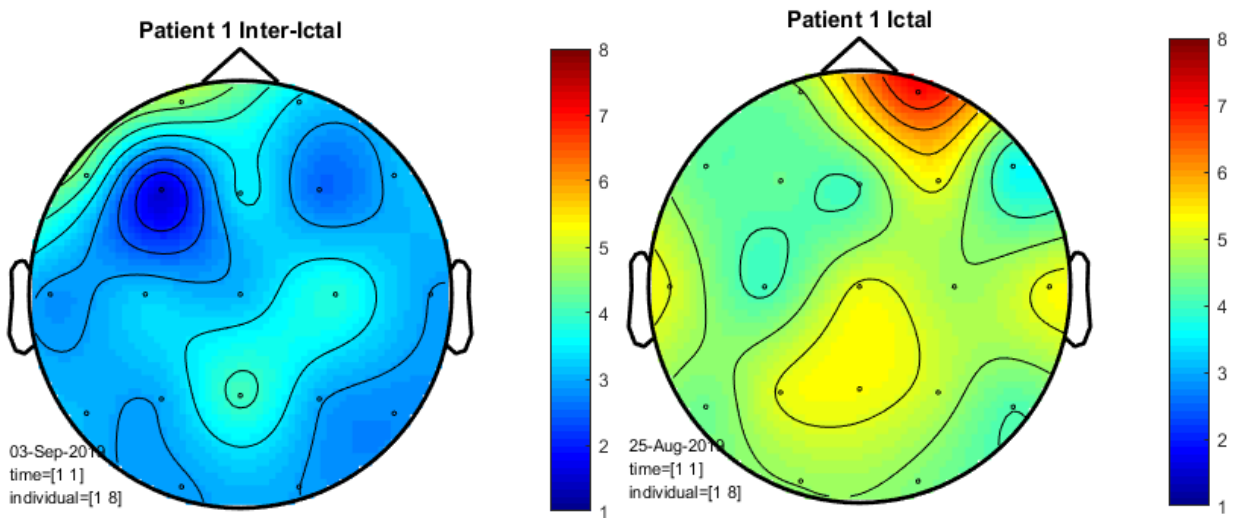


Figure 35: The graphs represent the mean CSP features map for inter-ictal (left graph) and ictal (right graph).

4.5 Feature Selection Methods

Nowadays, the growth of the high-throughput technologies has resulted in exponential growth in the harvested data with respect to both dimensionality and sample size. Efficient and effective management of these data becomes increasingly challenging. Traditionally manual management of these

datasets to be impractical. Therefore, data mining and machine learning techniques were developed to automatically discover knowledge and recognize patterns from these data [8] [9]. Dimensionality reduction is one of the most popular techniques to remove noisy (i.e. irrelevant) and redundant features. Dimensionality reduction techniques can be categorized mainly into feature extraction and feature selection.

Feature extraction approaches project features into a new feature space with lower dimensionality and the new constructed features are usually combinations of original features. Examples of feature extraction techniques include Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA) and Canonical Correlation Analysis (CCA). On the other hand, the feature selection approaches aim to select a small subset of features that minimize redundancy and maximize relevance to the target such as the class labels in classification. Representative feature selection techniques include Information Gain, Relief, Fisher Score and Lasso. Both Feature extraction and feature selection are capable of improving learning performance, lowering computational complexity, building better generalizable models, and decreasing required storage.

For the classification problem, feature selection aims to select a subset of highly discriminant features. In other words, it selects features that are capable of discriminating samples that belong to different classes. For the problem of feature selection for classification, due to the availability of label information, the relevance of features is assessed as the capability of distinguishing different classes. Furthermore, feature selection serves one more purpose. It detects not only the features with increased distinctive ability, but also those with low or zero contribution to the class separation. The removal of the later from the features set is equally important, since they act as noise for classification, usually leading to reduction of the classification efficiency. The three general categories of feature selection methods are categorized as:

- Filter based: We specify some metric and based on that filter features. An example of such a metric could be correlation/chi-square [10].
- Wrapper-based: Wrapper methods consider the selection of a set of features as a search problem. Example: Recursive Feature Elimination [11].
- Embedded: Embedded methods use algorithms that have built-in feature selection methods. For instance, Lasso and RF have their own feature selection methods [12].

Filter algorithms use a threshold to select the best features, while wrapper algorithms, through a variety of strategies, select subsets of the original features, evaluate their distinctive ability using a classifier, and result in the best subset, which then use for classification.

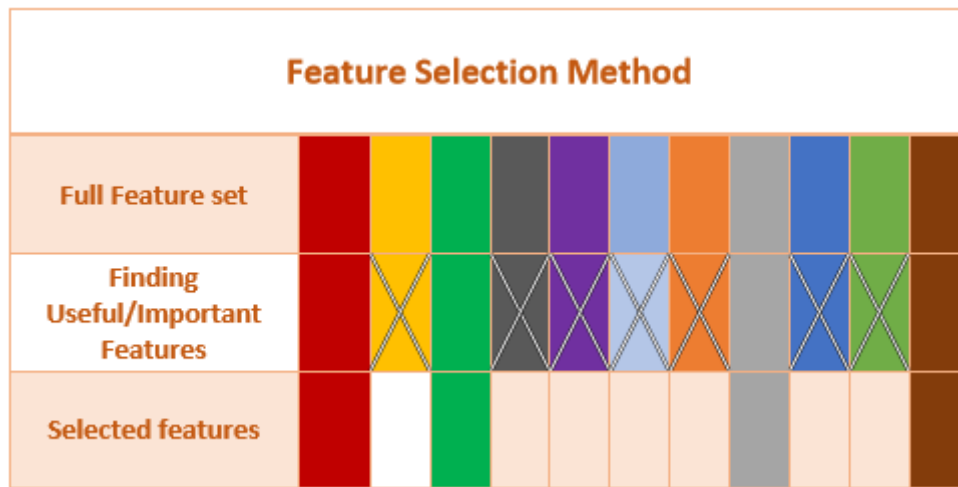


Figure 36: Feature Selection

Relief is a feature weight based algorithm inspired by instance-based learning. Relief has two critical components, the averaged vector relevance and the threshold. Relevance is the averaged vector of the value for each feature over m sample triplets. Each element of relevance corresponding to a feature shows how relevant the feature is to the target concept. The threshold is determining whether the feature should be selected or not. Relief is a simple algorithm which relies entirely on a statistical method. The algorithm employs few heuristics, and is less often fooled. It is efficient and its computational complexity is polynomial. Relief is also noise tolerant and is unaffected by feature interaction [13].

The RFE-SVM algorithm which is a wrapper method for feature selection method using Support Vector Machines. RFESVM method ranks all the features according to some score function and eliminates one or more features with the lowest scores. This process is repeated until the highest classification accuracy is obtained. Due to its successfully use in selecting informative genes for cancer classification, SVM-RFE gained a great popularity and is well known as one of the most effective feature selection method. However, the RFESVM is a greedy method that only hopes to find the best possible combination for classification [14].

Sequential forward selection (SFS) is the simplest greedy search algorithm. The way it operates is starting from an empty set, sequentially add the feature x that maximizes $S(Y + x)$ when combined with the features Y that have already been selected prior. SFS performs best when the optimal subset is small, because at the beginning a large number of feature states can be potentially evaluated. When the set is almost full, the region examined by SFS is narrower since most of the features have already been selected.

The main disadvantage of SFS is that it is unable to remove features that become obsolete after the addition of other features [15].

The mRMR is a feature selection approach that tends to select features with a high correlation with the class (relevance) and a low correlation between themselves (redundancy). Relevance can be calculated by using the F-statistic (for continuous features) or mutual information (for discrete features) and redundancy can be calculated by using Pearson correlation coefficient (for continuous features) or mutual information (for discrete features). After that, features are selected one by one by applying a greedy search to maximize the objective function, which is a function of relevance and redundancy. Two commonly used types of the objective function are MID (Mutual Information Difference criterion) and MIQ (Mutual Information Quotient criterion) representing the difference or the quotient of relevance and redundancy, respectively. For temporal data, mRMR feature selection approach requires some preprocessing techniques that flatten temporal data into a single matrix in advance. This may result in a loss of possibly important information among temporal data [16].

4.6 Classification Schemes

Classification is the problem of identifying to which of a set of categories (subpopulations) a new observation belongs, on the basis of a training set of data containing observations (or instances) whose category membership is known [17]. Many real-world problems can be modeled as classification problems such as assigning a given email into “spam” or “non-spam” classes and assigning a diagnosis to a given patient as described by observed characteristics of the patient (gender, blood pressure, presence or absence of certain symptoms, etc.).

In the training phase, data is analyzed into a set of features based on the feature generation models such as the vector space model for text data. These features may either be categorical (e.g. A, B), ordinal (e.g. large, small), integer-valued or real-valued (e.g. a measurement of height). Some algorithms work only in terms of discrete data such as ID3 and require that real-valued or integer-valued data be discretized into groups (e.g. less than 5, between 5 and 10, or greater than 10). After representing data through these extracted features, the learning algorithm will utilize the label information as well as the data itself to learn a map function f (or a classifier) from features.

In the prediction phase, data is represented by the feature set extracted in the training process, and then the map function (or the classifier) learned from the training phase will perform on the feature represented data to predict the labels. Note that the feature set used in the training phase should be the

same as that in the prediction phase. There are many classification methods in the literature. These methods can be categorized broadly into Linear classifiers, support vector machines, decision trees and Neural networks. A linear classifier makes a classification decision based on the value of a linear combination of the features. Examples of linear classifiers include Fisher's linear discriminant, logistic regression, the naive Bayes classifier and so on. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Therefore, support vector machine constructs a hyperplane or set of hyperplanes by maximizing the margin. In decision trees, a tree can be learned by splitting the source set into subsets based on a feature value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning. The recursion is completed when the subset at a node has all the same values of the target feature, or when splitting no longer adds value to the predictions.

K-nearest Neighbors is a lazy learning algorithm and is used to assign a data point to clusters based on similarity measurement. It uses a supervised method for classification. The steps to the algorithm are to choose the number of k and a distance metric. The most common choice of k is 5. Next, according to their distance, it finds the k -nearest neighbors of the sample that you want to classify. After finding the k nearest neighbors it weights the contribution of each of the k neighbors and assigns the class label by majority vote [18].

Named after Thomas Bayes from the 1700s who first coined this in the Western literature. Naive Bayes classifier works on the principle of conditional probability as given by the Bayes theorem. It is used for a variety of tasks such as spam filtering and other areas of text classification. Advantages of Naive Bayes Classifier are that it is very simple and easy to implement, it needs less training data, it can handle both continuous and discrete data, it is highly scalable with the number of predictors and data points and as it is fast, it can be used in real-time predictions. On the other hand, if the categorical variable belongs to a category that it was not followed up during the training set, then the model will give it a probability of 0 which will prevent it from making any prediction [19].

Artificial Neural Network is a set of connected input/output units where each connection has a weight associated with. It started by neurobiologists to develop and test computational analogs of neurons. During the learning phase, the network learns by adjusting the weights so as to be able to predict the correct class label of the input tuples. There can be multiple hidden layers in the model depending on the complexity of the function which is going to be mapped by the model. Having more hidden layers will

enable to model complex relationships such as deep neural networks. On the other hand, when exists many hidden layers, it takes significantly more time to train and adjust weights [20].

4.7 Cross Validation methods

Validation

This process of deciding whether the numerical results quantifying hypothesized relationships between variables, are acceptable as descriptions of the data, is known as validation. Generally, an error estimate for the test model is performed after training, known as residual evaluation. In this process, a numerical estimate of the difference between the predicted and the original responses is made and it is also called training error. The problem is that it provides an idea about the functionality of our model to the specific input of data. It is possible that the model is underfitting or overfitting the data. The problem that occurs based on this, is that the evaluation technique does not give an indication of how well the learning algorithm will perform to an independent or unseen data set. Getting this idea about our model is known as Cross Validation [21].

K-Fold Cross Validation

As there is never enough data to train your model, by removing a part of the data for validation purposes arises the problem of the remaining data being underfitting. The reduction of the training data may lead in losing important patterns in the data set. K Fold cross validation does is a method that we use our data for both the training and testing purposes and we can test the whole data without having to worry about underfitting. In K Fold cross validation, the data is divided into partitioned into k equal sized subsamples. The holdout method is repeated k times, so that each time one of the subsets is for testing the model while the others are used for training the model. The k results can then be averaged to produce a single estimation to acquire the total effectiveness of our model. As mentioned before, the cross-validation process is repeated k times, with each of the k subsamples used exactly once as the validation data. With this method we can significantly reduce bias as we are using all of the data for fitting, and also significantly reduces variance as the data are also being used in the validation set. The most common choice of k is 5 or 10, but it is not fixed and can take any value depending on the data [22]. If the number of k equals to

the number of the segments\observation then the k-fold cross validation is equivalent to leave-one-out cross-validation. [23]

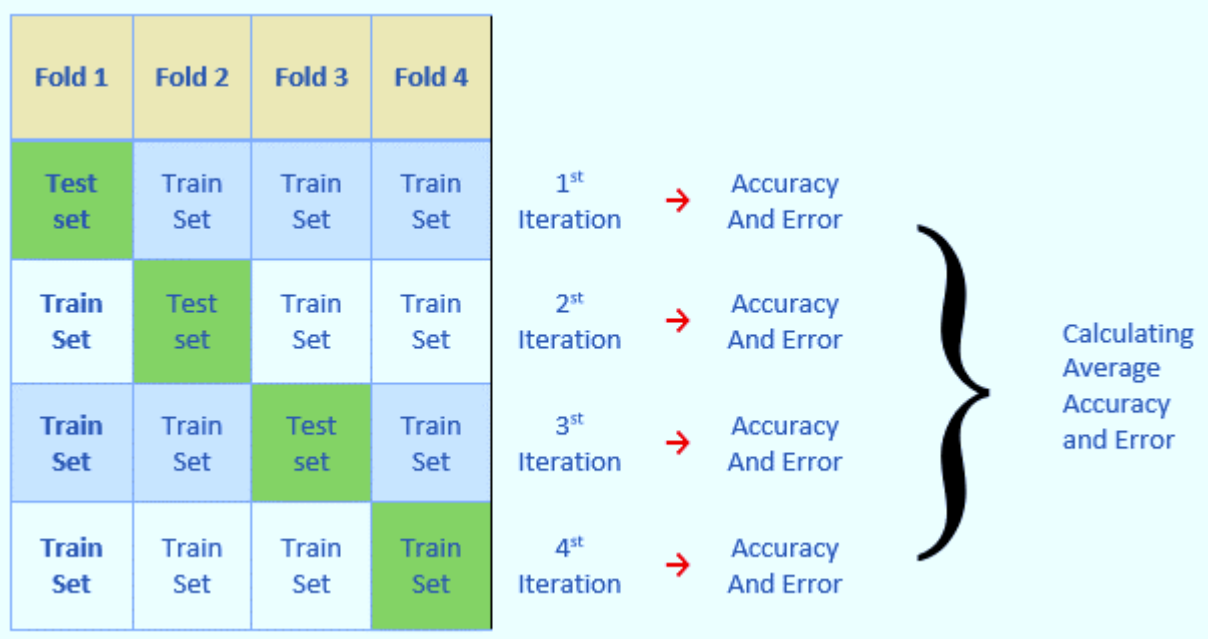


Figure 37: 4-fold cross validation

References

- [1] Y. Wang, S. Gao and X. Gao, "Common Spatial Pattern Method for Channel Selection in Motor Imagery Based Brain-computer Interface," in *IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005.
- [2] J. Müller-Gerking, G. Pfurtscheller and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical neurophysiology*, vol. 110, pp. 787-798, 1999.
- [3] S. Lemm, B. Blankertz, G. Curio and K.-R. Müller, "Spatio-spectral filters for improving the classification of single trial EEG," *IEEE transactions on biomedical engineering*, vol. 52, pp. 1541-1548, 2005.
- [4] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal processing magazine*, vol. 25, pp. 41-56, 2007.
- [5] T. N. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, F. M. Alotaibi, K. Aljibrin and S. R. Alrshod, "A common spatial pattern approach for scalp EEG seizure detection," in *5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, 2016.
- [6] K. Fukunaga, "Introduction to statistical pattern recognition," *Boston: Academic Press*, 1990.
- [7] R. Oostenveld, P. Fries, E. Maris and J. Schoffelen, "FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data.," *Computational Intelligence and Neuroscience*, no. 2011, 2011.
- [8] K. Giannakaki, G. Giannakakis, C. Farmaki and V. Sakkalis, "Emotional State Recognition Using Advanced Machine Learning Techniques on EEG Data," in *IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, 2017.
- [9] A. H. Shueb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010.
- [10] L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution," *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 856-863, 2003.
- [11] R. Kohavi and G. John, "Relevance wrappers for feature subset selection," *Artif. Intell.*, vol. 1, no. 97, p. 273-324, 1997.
- [12] T. Lal, O. Chapelle, J. Weston and A. Elisseeff, "Embedded Methods," *Feature Extraction: Foundations and Applications.*, pp. 137-165, 2006.
- [13] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," *AAAI-92 Proceedings*, 1992.
- [14] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene selection for cancer classification using support vector machines," *MACHLEARN: Machine Learning*, no. 46, pp. 389-422, 2002.
- [15] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 2006.
- [16] C. DING and H. PENG, "MINIMUM REDUNDANCY FEATURE SELECTION FROM MICROARRAY GENE EXPRESSION DATA," *Journal of Bioinformatics and Computational Biology*, vol. 2, no. 3, 2003.
- [17] M. S. d. Alencar, *Communication, Management and Information Technology: International*, London, UK: Taylor & Francis group, 2017.
- [18] N. S. Altman, "An Introduction to Kernel and Nearest Neighbor Nonparametric Regression," *The American Statistician*, vol. 3, no. 46, pp. 175-185, 1992.

- [19] M. E. Maron, "Automatic Indexing: An Experimental Inquiry," *Journal of the ACM.*, vol. 3, no. 8, pp. 404-417, 1961.
- [20] S. C. Kleene, "Representation of Events in Nerve Nets and Finite Automata," *Annals of Mathematics Studies*, 1956.
- [21] R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, vol. 2, no. 12, p. 1137–1143, 1995.
- [22] G. J. McLachlan, C. Ambrose and K.-A. Do, "Analyzing microarray gene expression data," *Wiley*, 2004.
- [23] T. Hastie, R. Tibshirani and J. Friedman, *Elements of Statistical Learning: data mining, inference, and prediction*. 2nd Edition, Springer, 2009.
- [24] Y. Zhang, G. Zhou, J. Jin, X. Wang and A. Cichocki, "Optimizing spatial patterns with sparse filter bands for motor-imagery based brain–computer interface," *Journal of neuroscience methods*, vol. 255, pp. 85-91, 2015.
- [25] W. Wu, Z. Chen, X. Gao, Y. Li, E. N. Brown and S. Gao, "Probabilistic common spatial patterns for multichannel EEG analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, pp. 639-653, 2014.
- [26] N. Tomida, T. Tanaka, S. Ono, M. Yamagishi and H. Higashi, "Active data selection for motor imagery EEG classification," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 458-467, 2014.
- [27] T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaibi and S. R. Alrshoud, "Epileptic seizure prediction using CSP and LDA for scalp EEG signals," *Computational intelligence and neuroscience*, vol. 2017, 2017.
- [28] Y. Zhang, Y. Wang, G. Zhou, J. Jin, B. Wang and X. Wang, "Multi-kernel extreme learning machine for EEG classification in brain-computer interfaces," *Expert Systems with Applications*, vol. 96, pp. 302-310, 2018.

Chapter Five: Proposed Approach and Results

5.1 Clinical Protocol and procedures

Inclusion criteria and ethics

Subjects participating in this study are patients diagnosed with non-idiopathic focal epilepsy. If a sustained seizure freedom is not ensured, then even the occurrence of at least one seizure event makes the subject eligible for inclusion in the study. There were some cases for which long-term video EEG and synchronous ECG were recorded, but when evaluated by the two neuropsychiatrists (see Section 0), they did not present any epileptic seizure. These cases were excluded from our analysis. The study's protocol has been approved by the appropriate scientific board of the University Hospital of Heraklion under the reference number 5631/15-5-14. Informed consent was obtained from all patients following a detailed explanation of the study objectives and protocol to each patient and/or caregiver. All caregivers/patients provided written informed consent prior to being monitored. This study complies with the obligations and procedures of confidentiality and privacy, to implement the security and protection policies of the personal data and to comply with the Rules established by the Greek legislation and by the European Parliament's Regulation (EU) 2016/679 and of 27 April 2016 (**General Data Protection Regulation - GDPR**)

Procedure

A patient that meets the criteria, as evaluated by an expert neuropsychiatrist, was admitted to the hospital. Their medical health record were created including clinical data about demographics, medical history, family history, medication, epilepsy classification, etc. An EEG cap with 10/20 electrode system was placed in the head of the patient, a camera was placed opposite the patient's bed and additional sensors for recording the breath rate and SpO2 were utilized. Video and surface EEG were recorded simultaneously for each patient during routine long-term hospital monitoring. The EEG signals

were recorded at 19 scalp loci of the international 10–20 system (channels Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2), with all electrodes referenced to the earlobe. An electrode placed in the middle of distance between Fp1 and Fp2 on the subject's forehead served as ground. EEG data were sampled at 256Hz. The long-term EEG recordings were independently evaluated and annotated for epileptic seizures and pathological findings by two expert neuropediatricians.

Dataset

The clinical dataset has been recorded at the University Hospital of Heraklion and contained 10 participants (3 females, 7 males). Their age was 6.8 ± 5.9 years at the moment of monitoring. The recorded dataset included 63 seizures in total. The seizures classification into standardized types and subtypes was performed according to the criteria of the International League Against Epilepsy (ILAE) [1]. Table I presents patients demographic data as well as selected clinical data. It should be noted that the dataset contains only cases of focal seizures in order to share similar clinical characteristics. Four additional cases were excluded from the study due to excessive body motion induced artifacts in the ECG or other errors, such as the removal of an electrode from the chest.

Patient Code	# seizures	Age	Gender	Epilepsy
PAT_11	1	3	Male	focal right frontal lobe
PAT_13	1	9	Female	focal right frontal lobe
PAT_14	8	7	Female	focal right frontal lobe
PAT_15	1	13	Male	focal fronto-polar lobe
PAT_24	14	10	Female	focal left frontal lobe
PAT_27	7	13	Male	focal right frontal lobe
PAT_29	1	0	Male	focal bifrontal lobe
PAT_32	1	13	Male	focal right frontotemporal lobe
PAT_34	28	5	Male	focal right frontal lobe
PAT_35	1	4	Male	focal
Total seizures	63			

Table IV: Study population Demographics and Seizure Types

EEG preprocessing and feature extraction

The EEG recordings were digitized in sampling frequency $f_s=256$ Hz which is used by the University Hospital of Heraklion for EEG recordings. Because the subjects were free to move and also during the seizures they exhibited various movements and actions the artifacts related to these activities (body movements, eye blinks, spikes, head movements, chewing, general discharges) contaminated the EEG recordings with unwanted noise components [2]. There was a need to suppress these artifacts and spikes, and the process of suppression was performed with the usage of Independent Component Analysis (ICA) [3]. Another step was the removal of the noise caused by mechanical interferences. The recording machines, amplifiers, and various others are operating with an electrical current of 50Hz, causing noise to the EEG recording at the specific frequency [4]. Finally, the application of the CSP method mentioned in the previous chapter took place. We segmented the preprocessed EEG recordings, into segments of 2sec

with the usage of a sliding temporal window. We then extracted the rhythms (δ , θ , α , β and γ) of each temporal window and applied the CSP [5] method to each rhythm and total spectrum separately and extracted the features that were to be used for classifying the segments as inter-ictal or ictal.

5.2 Proposed Pipeline and Use of Techniques

Feature Selection

The outline followed in this thesis for classification was to create combinations of different feature selection methods, classifications schemes and parameters, in order to evaluate their performance and determine which combination gives the best results for the specific data. The feature selection methods used are Correlation Coefficient, ReliefF algorithm [6], Minimum Redundancy Maximum Relevance algorithm (mRMR) with MID and MIQ schemes [7], linear Support Vector Machines Recursive Feature Elimination algorithm (SVM-RFE) [8], Sequential Forward Selection (SFS) [9]. There was also the case of selecting all variables (no feature selection). The result of each feature selection method was a ranking of the most important features. Based on this ranking, we created combinations of features: the first combination was the 1 top ranked feature, the second combination was the 2 top ranked features, the third combination was the 3 top ranked features, and so on. Each combination was used to train an SVM classifier and estimated its performance. The combination achieving the best performance, was finally used as the outcome of each feature selection method.

Classification

The discrimination between the two states under investigation (interictal, ictal) was performed by evaluating the CSP features through classification schemes comparison. The classification schemes used in this study are the trivial classifier, Naïve Bayes classifier, k-Nearest Neighbors with $k=5$, Artificial Neural Networks with hidden layer size 10. The trivial classifier is used in order to determine the accuracy of random classification, which depends on the class distribution of the data, and serves as a reference point for the performance of the other classifiers. In order to assess the performance of each classification scheme the performance accuracy metric was used which is given by the equation

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (10)$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

where TP is the true positive, TN the true negative, FP is the false positive and FN the false negative cases. Sensitivity is the ability of a classification scheme to correctly identify those with the disease (true positive rate), whereas specificity is the ability of the classification scheme to correctly identify those without the disease (true negative rate). The classification schemes (combinations of classifier and parameters) were cross-validated in order to evaluate their performance and select the best combination. A standard 10-fold cross validation method was utilized for each classification scheme for testing the performance of each system.

		The Truth		
		Has the disease	Does not have the disease	
Test Score:	Positive	True Positives (TP) <div>a</div>	False Positives (FP) <div>b</div>	$PPV = \frac{TP}{TP + FP}$
	Negative	False Negatives (FN) <div>c</div>	True Negatives (TN) <div>d</div>	$NPV = \frac{TN}{TN + FN}$
		Sensitivity $\frac{TP}{TP + FN}$ Or, $\frac{a}{a + c}$	Specificity $\frac{TN}{TN + FP}$ $\frac{d}{d + b}$	

Figure 38: Sensitivity – Specificity – source([10])

5.3 Results

The feature selection was applied aiming to find the optimal feature subset for improving discrimination ability. The automatic feature selection methods used are the Correlation Coefficient (with $R=0.1$), the ReliefF, the mRMR_MID, the mRMR_MIQ, backward greedy search, the linear SVM-RFE and the Sequential Forward Selection (SFS). The top ranked features were inserted iteratively in the feature subset evaluating its each candidate subset's performance in terms of 10-fold SVM classification accuracy used as the objective function. The next 2 tables provide information in the form of concentrating the algorithms used in order to extract the final percentages of accuracy, sensitivity, and specificity for both cases. At the top are displayed the features/CSP coefficients used for each case. Next follows the feature selection methods used to extract the final features/CSP coefficients that were used for classification, and finally, the classifiers and the number of folds for cross-validation process.

PARAMETERS	
Features	19 CSP Coefficients Total Power 19 CSP Coefficients δ Rhythm 19 CSP Coefficients θ Rhythm 19 CSP Coefficients α Rhythm 19 CSP Coefficients β Rhythm
Feature selection methods:	<div> <div> selectallvariables Correlation Coefficient, $R=0.1$ ReliefF mRMR_MID mRMR_MIQ backward greedy search, linear SVM-RFE Sequential Forward Selection (SFS) </div> <div> +SVM for selecting top features </div> </div>
Classifiers:	Trivial classifier Naïve Bayes KNN, $K=5$ Artificial Neural Networks, hidden layer size: 10

Cross-Validation:	number of folds: k=10
--------------------------	------------------------------

Table V: Features and Methods used in the analysis of the EEG signals (γ rhythm out)

PARAMETERS	
Features	19 CSP Coefficients Total Power 19 CSP Coefficients δ Rhythm 19 CSP Coefficients θ Rhythm 19 CSP Coefficients α Rhythm 19 CSP Coefficients β Rhythm 19 CSP Coefficients γ Rhythm
Feature selection methods:	<div> <div> selectallvariables Correlation Coefficient, $R=0.1$ ReliefF mRMR_MID mRMR_MIQ backward greedy search, linear SVM-RFE Sequential Forward Selection (SFS) </div> <div> +SVM for selecting top features </div> </div>
Classifiers:	Trivial classifier Naïve Bayes KNN, $K=5$ Artificial Neural Networks, hidden layer size: 10
Cross-Validation:	number of folds: k=10

Table VI: Features and Methods used in the analysis of the EEG signals

This procedure revealed that for the problem under investigation, the best feature selection method was ReliefF selecting a features subset of 49 features from the feature set for the case with 114 features/CSP coefficients included and all the 95 features for the other case, meaning without the features of the γ rhythm. The figure that follows is for visualization purposes only, to better understand the usage of the SVM classifier and do not represent our real data. It represents the 2 best features for classification, in our case, those two would be the features 105 and 97, as presented by Table VIII: Results of the analysis.

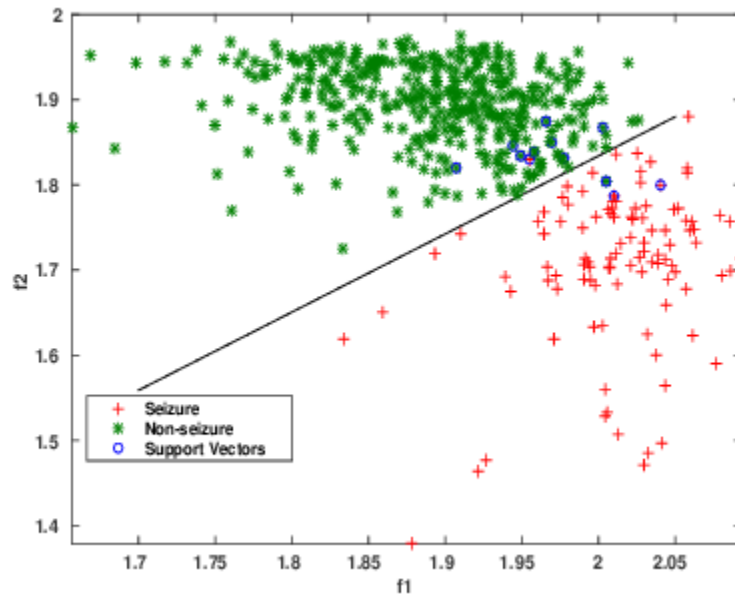


Figure 39: Classification between non-ictal (green color) and ictal period (red color) representing the first 2 best features (for visualization purposes) – source([11]).

Then, the classification phase was applied. The trivial classifier's accuracy was 50%, defining the random classification accuracy. The best accuracy, sensitivity and specificity was achieved with the ReliefF algorithm, the SVM classifier (with Gaussian kernel) and 10-fold cross-validation achieving classification of 92% accuracy, 89.4% sensitivity and 94.5% specificity for all the rhythms included (19 features for δ , 19 for θ , 19 for α , 19 for β , 19 for γ and 19 for total spectrum) to a total of 114 features and the best accuracy, sensitivity and specificity achieved for all the rhythms without the gamma features included (19 features for δ , 19 for θ , 19 for α , 19 for β and 19 for total spectrum) to a total of 95 features were with the selection of all 95 features, the KNN classifier (with $k=5$) and 10-fold cross-validation achieving classification of 91.2% accuracy, 90.6% sensitivity and 91.7% specificity.

RESULTS						
FeatSel Method	Selected Vars_N	SelectedVars	Classifier	Accuracy		
all variables	95	all	Trivial classifier	0.5000		
FeatSel Method	Selected Vars_N	SelectedVars	Classifier	Accuracy	Sensitivity	Specificity
all variables	95	all	KNN, K=5	0.9121	0.9066	0.9176
CorrCoef, R=0.1	95	50 8 19 44 48 2 12 57 42 54 63 18 27 86 56 71 16 69 65 6 15 31 82 90 84 72 17 46 10 88 11 21 52 80 23 35 75 51 67 4 7 76 13 93 61 14 38 73 94 37 59 40 74 91 29 92 87 33 55 68 47 70 53 3 83 66 45 28 25 49 95 26 9 36 64 5 32 89 1 78 34 85 81 30 43 62 79 20 41 24 77 22 58 60 39	SVM, gaussian kernel	0.9111	0.8872	0.9350
all variables	95	all	SVM, gaussian kernel	0.9106	0.8861	0.9350

ReliefF	94	93 47 84 90 30 55 88 49 80 38 16 2 53 10 36 69 50 35 39 22 71 56 48 21 4 27 66 52 34 29 85 54 44 45 78 94 76 91 87 42 40 65 74 24 25 7 75 51 11 81 8 6 92 12 46 31 15 70 14 1 20 68 23 62 17 72 86 95 9 33 41 57 60 3 58 82 19 83 32 73 43 59 61 5 26 89 13 37 63 64 28 79 18 67	SVM, gaussian kernel	0.9080	0.8841	0.9318
mRMR_MID	33	56 88 50 47 91 42 19 44 82 31 54 80 72 10 57 75 69 30 61 20 46 84 51 8 93 48 24 86 49 87 12 68 2	SVM, gaussian kernel	0.9079	0.8810	0.9349
CorrCoef, R=0.1	95	50 8 19 44 48 2 12 57 42 54 63 18 27 86 56 71 16 69 65 6 15 31 82 90 84 72 17 46 10 88 11 21 52 80 23 35 75 51 67 4 7 76 13 93 61 14 38 73 94 37 59 40 74 91 29 92 87 33 55 68 47 70 53 3 83 66 45 28 25 49 95 26 9 36 64 5 32 89 1 78 34 85 81 30 43 62 79 20 41 24 77 22 58 60 39	KNN, K=5	0.9070	0.8994	0.9146

ReliefF	94	93 47 84 90 30 55 88 49 80 38 16 2 53 10 36 69 50 35 39 22 71 56 48 21 4 27 66 52 34 29 85 54 44 45 78 94 76 91 87 42 40 65 74 24 25 7 75 51 11 81 8 6 92 12 46 31 15 70 14 1 20 68 23 62 17 72 86 95 9 33 41 57 60 3 58 82 19 83 32 73 43 59 61 5 26 89 13 37 63 64 28 79 18 67	KNN, K=5	0.9070	0.8983	0.9157
backw. greedy search, lin. SVM-RFE	50	25 21 5 10 24 22 82 35 26 34 28 23 37 3 31 12 27 6 78 64 32 20 29 1 86 77 81 16 33 71 54 4 69 49 50 39 7 79 36 83 47 89 68 41 14 58 18 15 57 2	SVM, gaussian kernel	0.9065	0.8872	0.9259
mRMR_MIQ	47	56 88 50 19 42 91 69 44 47 80 31 72 8 82 54 61 86 11 10 93 46 17 75 84 57 63 76 90 51 2 49 15 1 95 12 65 45 67 35 71 48 13 52 83 7 16 94	SVM, gaussian kernel	0.9060	0.8842	0.9278
SFS	17	42 35 84 44 12 94 82 54 93 23 61 68 1 83 67 39 89	SVM, gaussian kernel	0.8989	0.8710	0.9269

mRMR_MIQ	47	56 88 50 19 42 91 69 44 47 80 31 72 8 82 54 61 86 11 10 93 46 17 75 84 57 63 76 90 51 2 49 15 1 95 12 65 45 67 35 71 48 13 52 83 7 16 94	KNN, K=5	0.8989	0.8923	0.9055
mRMR_MID	33	56 88 50 47 91 42 19 44 82 31 54 80 72 10 57 75 69 30 61 20 46 84 51 8 93 48 24 86 49 87 12 68 2	KNN, K=5	0.8979	0.8781	0.9176
SFS	17	42 35 84 44 12 94 82 54 93 23 61 68 1 83 67 39 89	KNN, K=5	0.8913	0.8812	0.9014
backw. greedy search, lin. SVM-RFE	50	25 21 5 10 24 22 82 35 26 34 28 23 37 3 31 12 27 6 78 64 32 20 29 1 86 77 81 16 33 71 54 4 69 49 50 39 7 79 36 83 47 89 68 41 14 58 18 15 57 2	KNN, K=5	0.8913	0.8802	0.9024
all variables	95	all	ANN, hidden Layer Size=10	0.7947	0.8435	0.7460

CorrCoef, R=0.1	95	50 8 19 44 48 2 12 57 42 54 63 18 27 86 56 71 16 69 65 6 15 31 82 90 84 72 17 46 10 88 11 21 52 80 23 35 75 51 67 4 7 76 13 93 61 14 38 73 94 37 59 40 74 91 29 92 87 33 55 68 47 70 53 3 83 66 45 28 25 49 95 26 9 36 64 5 32 89 1 78 34 85 81 30 43 62 79 20 41 24 77 22 58 60 39	ANN, hidden Layer Size=10	0.7799	0.8221	0.7376
mRMR_MID	33	56 88 50 47 91 42 19 44 82 31 54 80 72 10 57 75 69 30 61 20 46 84 51 8 93 48 24 86 49 87 12 68 2	NB	0.7734	0.9004	0.6464
ReliefF	94	93 47 84 90 30 55 88 49 80 38 16 2 53 10 36 69 50 35 39 22 71 56 48 21 4 27 66 52 34 29 85 54 44 45 78 94 76 91 87 42 40 65 74 24 25 7 75 51 11 81 8 6 92 12 46 31 15 70 14 1 20 68 23 62 17 72 86 95 9 33 41 57 60 3 58 82 19 83 32 73 43 59 61 5 26 89 13 37 63 64 28 79 18 67	ANN, hidden Layer Size=10	0.7657	0.8183	0.7132

mRMR_MIQ	47	56 88 50 19 42 91 69 44 47 80 31 72 8 82 54 61 86 11 10 93 46 17 75 84 57 63 76 90 51 2 49 15 1 95 12 65 45 67 35 71 48 13 52 83 7 16 94	ANN, hidden Layer Size=10	0.7632	0.8323	0.6942
mRMR_MIQ	47	56 88 50 19 42 91 69 44 47 80 31 72 8 82 54 61 86 11 10 93 46 17 75 84 57 63 76 90 51 2 49 15 1 95 12 65 45 67 35 71 48 13 52 83 7 16 94	NB	0.7623	0.9046	0.6199
mRMR_MID	33	56 88 50 47 91 42 19 44 82 31 54 80 72 10 57 75 69 30 61 20 46 84 51 8 93 48 24 86 49 87 12 68 2	ANN, hidden Layer Size=10	0.7607	0.8363	0.6851
backw. greedy search, lin. SVM-RFE	50	25 21 5 10 24 22 82 35 26 34 28 23 37 3 31 12 27 6 78 64 32 20 29 1 86 77 81 16 33 71 54 4 69 49 50 39 7 79 36 83 47 89 68 41 14 58 18 15 57 2	ANN, hidden Layer Size=10	0.7602	0.8120	0.7084
SFS	17	42 35 84 44 12 94 82 54 93 23 61 68 1 83 67 39 89	NB	0.7480	0.9035	0.5925

SFS	17	42 35 84 44 12 94 82 54 93 23 61 68 1 83 67 39 89	ANN, hidden Layer Size=10	0.7434	0.8231	0.6636
CorrCoef, R=0.1	95	50 8 19 44 48 2 12 57 42 54 63 18 27 86 56 71 16 69 65 6 15 31 82 90 84 72 17 46 10 88 11 21 52 80 23 35 75 51 67 4 7 76 13 93 61 14 38 73 94 37 59 40 74 91 29 92 87 33 55 68 47 70 53 3 83 66 45 28 25 49 95 26 9 36 64 5 32 89 1 78 34 85 81 30 43 62 79 20 41 24 77 22 58 60 39	NB	0.7429	0.9045	0.5812
ReliefF	94	93 47 84 90 30 55 88 49 80 38 16 2 53 10 36 69 50 35 39 22 71 56 48 21 4 27 66 52 34 29 85 54 44 45 78 94 76 91 87 42 40 65 74 24 25 7 75 51 11 81 8 6 92 12 46 31 15 70 14 1 20 68 23 62 17 72 86 95 9 33 41 57 60 3 58 82 19 83 32 73 43 59 61 5 26 89 13 37 63 64 28 79 18 67	NB	0.7419	0.9025	0.5814
all variables	95	all	NB	0.7419	0.9034	0.5804

backw.	50	25 21 5 10 24 22 82 35	NB	0.7369	0.8893	0.5845
greedy		26 34 28 23 37 3 31 12				
search, lin.		27 6 78 64 32 20 29 1				
SVM-RFE		86 77 81 16 33 71 54 4				
		69 49 50 39 7 79 36 83				
		47 89 68 41 14 58 18				
		15 57 2				

Table VII: Results of the analysis (γ rhythm out)

Results						
FeatSel Method	Selected Vars_N	Selected Vars	Classifier	Accuracy		
all variables	114	all	Trivial classifier	0.5000		
FeatSel Method	Selected Vars_N	Selected Vars	Classifier	Accuracy	Sensitivity	Specificity
ReliefF	49	105 97 107 111 84 108 47 54 101 90 34 49 66 88 109 50 44 30 85 12 93 40 87 29 10 104 38 31 103 76 52 46 68 75 42 36 56 69 22 78 102 7 91 33 27 48 35 71 106	SVM, gaussian kernel	0.9197	0.8943	0.9451
all variables	114	all	SVM, gaussian kernel	0.9172	0.8953	0.9390
all variables	114	all	KNN, K=5	0.9162	0.9055	0.9269
backw. greedy search, lin.SVM-RFE	49	10 25 21 5 24 35 34 28 26 22 20 1 16 29 32 78 3 7 23 82 77 37 64 31 6 7 86 33 81 4 12 36 39 69 45 58 54 99 18 14 2 43 40 15 41 38 13 9 105	SVM, gaussian kernel	0.9157	0.8944	0.9369
ReliefF	49	105 97 107 111 84 108 47 54 101 90 34 49 66 88 109 50 44 30 85 12 93 40 87 29 10 104 38 31 103 76 52 46 68 75 42 36 56 69 22 78 102 7 91 33 27 48 35 71 106	KNN, K=5	0.9136	0.9025	0.9247

mRMR_MIQ	50	97 51 12 44 111 50 69 106 47 42 19 105 82 61 54 108 56 10 8 101 63 76 80 46 57 86 72 110 17 103 91 31 84 65 49 104 75 11 71 90 35 67 2 107 48 102 45 95 13 16	KNN, K=5	0.9131	0.9056	0.9207
CorrCoef, R=0.1	82	50 8 19 2 42 44 48 97 57 12 54 63 105 18 86 27 82 71 65 69 84 107 101 90 106 80 16 6 46 110 88 31 56 104 111 10 17 102 72 67 15 21 61 4 75 73 51 103 52 76 93 23 35 108 7 94 59 14 112 11 109 13 40 37 113 38 114 29 99 91 92 87 74 55 100 96 47 33 68 3 95 83	SVM, gaussian kernel	0.9125	0.8891	0.9359
mRMR_MIQ	50	97 51 12 44 111 50 69 106 47 42 19 105 82 61 54 108 56 10 8 101 63 76 80 46 57 86 72 110 17 103 91 31 84 65 49 104 75 11 71 90 35 67 2 107 48 102 45 95 13 16	SVM, gaussian kernel	0.9111	0.8893	0.9329
mRMR_MID	32	97 56 50 61 10 106 47 44 12 111 54 82 105 57 108 19 42 69 75 46 20 101 63 8 51 72 80 48 76 102 31 49	SVM, gaussian kernel	0.9106	0.8790	0.9421
SFS	15	42 12 109 3 84 85 10 99 6 7 78 105 89 113 22	SVM, gaussian kernel	0.9080	0.8902	0.9258

CorrCoef, R=0.1	82	50 8 19 2 42 44 48 97 57 12 54 63 105 18 86 77 82 71 65 69 84 107 101 90 106 80 16 6 46 110 88 31 56 104 111 10 17 102 72 67 15 21 61 4 75 73 51 103 52 76 93 23 35 108 7 94 59 14 112 11 109 13 40 37 113 38 114 29 99 91 92 87 74 55 100 96 47 33 68 3 95 83	KNN, K=5	0.9065	0.8932	0.9197
mRMR_MID	32	97 56 50 61 10 106 47 44 12 111 54 82 105 57 108 19 42 69 75 46 20 101 63 8 51 72 80 48 76 102 31 49	KNN, K=5	0.9019	0.8922	0.9115
backw. greedy search, lin. SVM-RFE	49	10 25 21 5 24 35 34 28 26 22 20 1 16 29 32 78 3 27 23 82 77 37 64 31 6 7 86 33 81 4 12 36 39 69 45 58 54 99 18 14 2 43 40 15 41 38 13 9 105	KNN, K=5	0.8968	0.8912	0.9024
SFS	15	42 12 109 3 84 85 10 99 6 7 78 105 89 113 22	KNN, K=5	0.8923	0.8649	0.9197
mRMR_MID	32	97 56 50 61 10 106 47 44 12 111 54 82 105 57 108 19 42 69 75 46 20 101 63 8 51 72 80 48 76 102 31 49	ANN, hidden Layer Size=10	0.7897	0.8405	0.7388
mRMRJVID	32	97 56 50 61 10 106 47 44 12 111 54 82 105 57 108 19 42 69 75 46 20 101 63 8 51 72 80 48 76 102 31 49	NB	0.7870	0.8922	0.6819

CorrCoef, R=0.1	82	50 8 19 2 42 44 48 97 57 12 54 63 105 18 86 27 82 71 65 69 84 107 101 90 106 80 16 6 46 110 88 31 56 104 111 10 17 102 72 67 15 21 61 4 75 73 51 103 52 76 93 23 35 108 7 94 59 14 112 11 109 13 40 37 113 38 114 29 99 91 92 87 74 55 100 96 47 33 68 3 95 83	ANN, hidden Layer Size=10	0.7836	0.8262	0.7409
mRMR_MIQ	50	97 51 12 44 111 50 69 106 47 42 19 105 82 61 54 108 56 10 8 101 63 76 80 46 57 86 72 110 17 103 91 31 84 65 49 104 75 11 71 90 35 67 2 107 48 102 45 95 13 16	ANN, hidden Layer Size=10	0.7820	0.8364	0.7275
all variables	114	all	ANN, hidden Layer Size=10	0.7815	0.8170	0.7459
backw. greedy search, lin. SVM-RFE	49	10 25 21 5 24 35 34 28 26 22 20 1 16 29 32 78 3 27 23 82 77 37 64 31 6 7 86 33 81 4 12 36 39 69 45 58 54 99 18 14 2 43 40 15 41 38 13 9 105	ANN, hidden Layer Size=10	0.7744	0.8374	0.7114
ReliefF	49	105 97 107 111 84 108 47 54 101 90 34 49 66 88 109 50 44 30 85 12 93 40 87 29 10 104 38 31 103 76 52 46 68 75 42 36 56 69 22 78 102 7 91 33 27 48 35 71 106	ANN, hidden Layer Size=10	0.7729	0.8203	0.7255

mRMR_MIQ	50	97 51 12 44 111 50 69 106 47 42 19 105 82 61 54 108 56 10 8 101 63 76 80 46 57 86 72 110 17 103 91 31 84 65 49 104 75 11 71 90 35 67 2 107 48 102 45 95 13 16	NB	0.7713	0.8903	0.6524
ReliefF	49	105 97 107 111 84 108 47 54 101 90 34 49 66 88 109 50 44 30 85 12 93 40 87 29 10 104 38 31 103 76 52 46 68 75 42 36 56 69 22 78 102 7 91 33 71 48 35 71 106	NB	0.7627	0.8912	0.6343
CorrCoef, R=0.1	82	50 8 19 2 42 44 48 97 57 12 54 63 105 18 86 27 82 71 65 69 84 107 101 90 106 80 16 6 46 110 88 31 56 104 111 10 17 102 72 67 15 21 61 4 75 73 51 103 52 76 93 23 35 108 7 94 59 14 112 11 109 13 40 37 113 38 114 29 99 91 92 87 74 55 100 96 47 33 68 3 95 83	NB	0.7566	0.8974	0.6158
all variables	114	all	NB	0.7475	0.8964	0.5987
SFS	15	42 12 109 3 84 85 10 99 6 7 78 105 89 113 22	NB	0.7434	0.9014	0.5854
backw. greedy search, lin.SVM-RFE	49	10 25 21 5 24 35 34 28 26 22 20 1 16 29 32 78 3 27 23 82 77 37 64 31 6 7 86 33 81 4 12 36 39 69 45 58 54 99 18 14 2 43 40 15 41 38 13 9 105	NB	0.7429	0.8892	0.5966

SFS	15	42 12 109 3 84 85 10 99 6 7 78 105 89 113 22	ANN, hidden Layer Size=10	0.7313	0.8180	0.6447
-----	----	---	------------------------------------	--------	--------	--------

Table VIII: Results of the analysis

The results were very encouraging, as they achieved very good classification accuracy, specificity and sensitivity percentages. The usage of the Common Spatial Patterns proved to be able to provide features that can help in the classification of the ictal and inter-ictal periods of time. From the observation of the results, we can extract that Naive Bayes and Artificial Neural Network with 10 hidden layers, had the lowest accuracy achieved from the four classifiers by a percent of at least 10% lower than Support-vector machine and K-nearest neighbors classifiers. The reason is the average percentage of specificity that was near 72% for the Artificial Neural Network and around 62% for the Naive Bayes. The percentages are from matrix VIII with all the features included, but similar results were extracted from the other case of the 95 features (without γ rhythm). The reason for this could be investigated in a later study, as well as the usage of for selection algorithms and more classifiers. The Support-vector machine and K-nearest neighbors classifiers had similar percentages of accuracy but we can extract from the matrices that SVM achieved the highest specificity of 94.5% for the 114 features and KNN the highest sensitivity of 90.6% for the 95. In general, the SVM had better overall specificity in both matrices and KNN better overall sensitivity. The percentages of the accuracy of classification were higher when we used the features from all the rhythms (δ , θ , α , β and γ). That is could be investigated in future studies, of the role of γ rhythm in seizures and the discrimination ability in classification. The reason for the speculation is that a lot of its features/CSP coefficients provide better classification results.

References

- [1] A. T. Berg, S. F. Berkovic, M. J. Brodie, J. Buchhalter, J. H. Cross and W. v. E. Boas, "Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009," *Epilepsia*, vol. 51, pp. 676-685, 2010.
- [2] J. W. Bang, J.-S. Choi and K. R. Park, "Noise Reduction in Brainwaves by Using Both EEG Signals and Frontal Viewing Camera Images," *Sensors (Basel)*, p. 6272–6294, 2013.
- [3] A. Hyvärinen, "Independent component analysis: recent advances," *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, pp. 1-19, 2013.
- [4] G. Repovš, "Dealing with Noise in EEG Recording and Data Analysis," *Infor Med Slov*, pp. 18-25, 2010.
- [5] J. Müller-Gerking, G. Pfurtscheller and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical neurophysiology*, vol. 110, pp. 787-798, 1999.
- [6] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," *AAAI-92 Proceedings*, 1992.
- [7] C. DING and H. PENG, "MINIMUM REDUNDANCY FEATURE SELECTION FROM MICROARRAY GENE EXPRESSION DATA," *Journal of Bioinformatics and Computational Biology*, vol. 2, no. 3, 2003.
- [8] I. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene selection for cancer classification using support vector machines," *MACHLEARN: Machine Learning*, no. 46, pp. 389-422, 2002.
- [9] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 2006.
- [1] G. Diaz, "GrepMed," [Online]. Available:
 0] <https://www.grepmed.com/images/1480/positivepredictivevalue-negativepredictivevalue-epidemiology-specificity-calculation-sensitivity-statistics>.
- [1] M. Sharma, "Researchgate," Classification of ictal and non-ictal EEG signals by LS-SVM classifier for
 1] the best two features f1=FD of SB-17 and f2= FD of SB-2., [Online]. Available:
https://www.researchgate.net/publication/315582822_A_new_approach_to_characterize_epileptic_seizures_using_analytic_time-frequency_flexible_wavelet_transform_and_fractal_dimension/figures?lo=1.

Chapter Six: Conclusion

6.1 Conclusion

The aim of this diploma thesis is to investigate the preservation of epileptic data after the compression and reconstruction of the EEG using the Compressed Sensing method and to develop a seizure detection model based on CSP features. An issue in the area of seizure detection using EEG signal analysis is that information from the multivariate signals coming from different EEG channels should be taken into account in order to discriminate a possible ictal period. This information presented in EEG head maps that can be represented using specific coefficients is very useful towards this direction. There is a signal processing technique for efficiently acquiring and reconstructing a signal, by finding solutions to underdetermined linear systems. We propose the Discrete Cosine basis to sparsify the EEG. Next we make the DCT signal sparser by reducing its energy. After the reduction we apply the CS technique and reconstruct the reduced DCT signal by solving the Basis Pursuit problem. In the end we check the frequency spectrum and applying the Approximate Entropy. The results indicate that the epileptic data are preserved with a small reduction and can be used to identify the epilepsy. On the next phase of the thesis we use the CSP method to create a seizure detection model. The CSP method is an advanced signal processing method that provides information using the multivariate signals data and using spatial filters extracts meaningful spatial coefficients that describe cerebral activity. These CSP features using automatic feature selection and classification techniques led to a best achieved time-window classification of 91.1% using the combination ReliefF and SVM (Gaussian Kernel) and 10-fold cross validation. These results indicate that CSP features could be used in combination with other feature for seizure detection, although their effectiveness might be improved.

6.2 Future Work

In the future, we propose to combine the 19-channel CSP features of each rhythm in a specific metric that will represent the underlying EEG map with the appropriate normalizations needed. The validation of the proposed methodology could be validated with a greater number of participants in a follow up study. It is also important that the features of the method can be used with other features and improve the detection model.

References

- [1 "Epilepsy: a public health imperative: summary," World Health Organization report, 2019.
]
- [2 K. M. Fiest, K. M. Sauro, S. Wiebe, S. B. Patten, C.-S. Kwon and J. Dykeman, "Prevalence and incidence
] of epilepsy A systematic review and meta-analysis of international studies," *Neurology*, vol. 88, pp. 296-303, 2017.
- [3 ILAE, "Proposal for revised clinical and electroencephalographic classification of epileptic seizures.
] From the Commission on Classification and Terminology of the International League against Epilepsy," *Epilepsia*, vol. 22, pp. 489-501, 1981.
- [4 R. S. Fisher, "The new classification of seizures by the International League Against Epilepsy 2017,"
] *Current neurology and neuroscience reports*, vol. 17, p. 48, 2017.
- [5 R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. E. Jr, L. Forsgren, J. A.
] French, M. Glynn, D. C. Hesdorffer, B. Lee, G. W. Mathern, S. L. Moshe and E. Perucca, "A practical clinical definition of epilepsy," *Epilepsia*, vol. 55, no. 4, pp. 475-482, 2014.
- [6 I. E. Scheffer, S. Berkovic, G. Capovilla, M. B. Connolly, J. French and L. Guilhoto, "ILAE classification of
] the epilepsies. Position paper of the ILAE Commission for Classification and Terminology," *Epilepsia*, vol. 58, no. 4, pp. 512-521, 2017.
- [7 R. S. Fisher, J. H. Cross, C. D'Souza, J. A. French, S. R. Haut, N. Higurashi, E. Hirsch, F. E. Jansen, L. Lagae,
] S. L. Moshe and J. Peltova, "Instruction manual for the ILAE 2017 operational classification of seizure types," *Epilepsia*, vol. 58, no. 4, pp. 531-542, 2017.
- [8 B. E. Swartz, "Timeline of the history of EEG and associated fields," *Electroencephalography and
] clinical Neurophysiology*, vol. 106, pp. 173-176, 1998.
- [9 P. Gloor, Hans Berger on the Electroencephalogram of Man. The Fourteen Original Reports on the
] Human Electroencephalogram, vol. Electroencephalography and Clinical Neurophysiology, New York: Elsevier, 1969.
- [1 J. Malmivuo and R. Plonsey, Bioelectromagnetism - Principles and Applications of Bioelectric and
0] Biomagnetis, New York: Oxford University Pres, 1995.
- [1 H. H. Jasper, "The 10–20 electrode system of the International Federation," *Electroencephalography
1] and clinical Neurophysiology*, vol. 10, pp. 370-375, 1958.
- [1 G. H. Klem, H. O. Luders, H. H. Jasper and C. Elger, "The ten±twenty electrode system of the
2] International Federation," *Recommendations for the Practice of Clinical Neurophysiology: Guidelines of the International Federation of Clinical Physiolog*, no. 52, pp. 3-6, 1999.
- [1 "Instrumentation Forum," [Online]. Available: [https://instrumentationforum.com/t/electrode-10-20-3\] system/5486](https://instrumentationforum.com/t/electrode-10-20-3] system/5486).
- [1 G. E. Chatrian, E. Lettich and P. L. Nelson, "Ten percent electrode system for topographic studies of
4] spontaneous and evoked EEG activity," *American Journal EEG Technology*, vol. 25, pp. 83-92, 1985.
- [1 "Neupsy KeyFastest Neupsy Insight Engine," [Online]. Available: [https://neupsykey.com/eeg-5\] instrumentation/](https://neupsykey.com/eeg-5] instrumentation/).
- [1 "American Electroencephalographic Society, Guideline thirteen: Guidelines for standard electrode
6] position nomenclature," *J Clin Neurophysiol*, vol. 1994, pp. 111-113, 1994.
- [1 L. G. Tassinary, T. H. Geen, J. T. Cacioppo and R. Edelberg, "Issues in biometrics: Offset potentials and
7] the electrical stability of Ag/AgCl electrodes," *Psychophysiology*, vol. 27, pp. 236-242, 1990.

- [1 "Epilepsy Foundation," Basic Science of EEG, [Online]. Available: 8] <https://www.epilepsy.com/learn/professionals/diagnosis-treatment/basic-science-eeg>.
- [1 "Guideline 6: A Proposal for Standard Montages to Be Used in Clinical EEG," American Clinical 9] Neurophysiology Society, 2006.
- [2 M. M. Kabiraj and N. M. Biary, "Basic principles and interpretations of electroencephalography," 0] *Neurosciences*, vol. 2, no. 8, pp. 131-143, 2003.
- [2 "IEEE Computer Society, Digital Library," [Online]. Available: 1] <https://www.computer.org/csdl/magazine/co/2012/07/mco2012070087/13rUYgT7Bk>.
- [2 "American Clinical Neurophysiology Society Guideline 6: Minimum Technical Standards for EEG 2] Recording in Suspected Cerebral Death," *J Clin Neurophysiol*, vol. 4, no. 33, pp. 324-327, 2016.
- [2 J. V. d. Drift and o. Magnus, "The value of the EEG in the differential diagnosis of cases with cerebral 3] lesion," *Electroencephalogr Clin Neurophysiol.*, no. 11, pp. 733-746, 1959.
- [2 W. A. Cobb, "Rhythmic slow discharges in the electroencephalogram," *J Neurol Neurosurg Psychiatry*, 4] no. 8, pp. 65-78, 1945.
- [2 E. J. Candes and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing 5] Magazine*, pp. 21-30, 2008.
- [2 D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289-1306, 2006. 6]
- [2 B. Keaton, "Sampling theory Fourier theory made easy," [Online]. Available: 7] <https://www.slideserve.com/barton/sampling-theory-fourier-theory-made-easy>.
- [2 M. W. Fakhr, "Researchgate," March 2017. [Online]. Available: 8] https://www.researchgate.net/publication/314839832_Joint_Image_Compression_and_Encryption_Based_on_Compressed_Sensing_and_Entropy_Coding/figures?lo=1.
- [2 S. Bhattacharya, "Compressed Sensing and Control," [Online]. Available: 9] http://web.me.iastate.edu/sbhattac/research_cs.html.
- [3 R. G. Baraniuk, "SlidePlayer," [Online]. Available: <https://slideplayer.com/slide/5237107/>. 0]
- [3 E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *C. R. 1] Acad. Sci. I*, vol. 346, pp. 589-592, 2008.
- [3 E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, pp. 4203- 2] 4215, 2005.
- [3 D. L. Donoho, Y. Tsaig, I. Drori and J.-L. Starck, "Sparse Solution of Underdetermined Linear Equations 3] by Stagewise Orthogonal Matching Pursuit," Preprint, 2007.
- [3 S. S. Chen, D. L. Donoho and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. 4] Comput.*, vol. 20, pp. 33-61, 1998.
- [3 S. Pincus, "Approximate entropy as a measure of system complexity," *Proc. Nati. Acad. Sci. USA*, vol. 5] 88, pp. 2297-2301, 1991.
- [3 B. Bein, "Entropy," *Best Practice & Research Clinical Anaesthesiology*, vol. 20, no. 1, pp. 101-109, 2006. 6]
- [3 S. Pincus and E. Gladstone, "A Regularity Statistic For Medical Data Analysis," *Journal of Clinical 7] Monitoring*, vol. 7, no. 4, 1991.

- [3 T. Jartii, T. Kuusela, T. Kaila, K. Tahvanainen and I. Valimaki, "The dose-response effects of terbutaline
8] on the variability, approximate entropy and fractal dimension of heart rate and blood pressure," *Br J Clin Pharmacol*, vol. 45, no. 3, pp. 277-285, 1998.
- [3 S. Pincus and D. Keefe, "Quantification of hormone pulsatility via an approximate entropy algorithm,"
9] *The American Physiological Society*, vol. 262, no. 5, pp. 741-753, 1992.
- [4 Y. Wang, S. Gao and X. Gao, "Common Spatial Pattern Method for Channel Selection in Motor Imagery
0] Based Brain-computer Interface," in *IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005.
- [4 S. Lemm, B. Blankertz, G. Curio and K.-R. Muller, "Spatio-spectral filters for improving the classification
1] of single trial EEG," *IEEE transactions on biomedical engineering*, vol. 52, pp. 1541-1548, 2005.
- [4 B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe and K.-R. Muller, "Optimizing spatial filters for robust
2] EEG single-trial analysis," *IEEE Signal processing magazine*, vol. 25, pp. 41-56, 2007.
- [4 T. N. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, F. M. Alotaibi, K. Aljibrin and S. R. Alrshod, "A common
3] spatial pattern approach for scalp EEG seizure detection," in *5th International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, 2016.
- [4 K. Fukunaga, "Introduction to statistical pattern recognition," *Boston: Academic Press*, 1990.
4]
- [4 J. Müller-Gerking, G. Pfurtscheller and H. Flyvbjerg, "Designing optimal spatial filters for single-trial
5] EEG classification in a movement task," *Clinical neurophysiology*, vol. 110, pp. 787-798, 1999.
- [4 R. Oostenveld, P. Fries, E. Maris and J. Schoffelen, "FieldTrip: Open Source Software for Advanced
6] Analysis of MEG, EEG, and Invasive Electrophysiological Data.," *Computational Intelligence and Neuroscience*, no. 2011, 2011.
- [4 K. Giannakaki, G. Giannakakis, C. Farmaki and V. Sakkalis, "Emotional State Recognition Using
7] Advanced Machine Learning Techniques on EEG Data," in *IEEE 30th International Symposium on Computer-Based Medical Systems (CBMS)*, 2017.
- [4 A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," in
8] *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, 2010.
- [4 L. Yu and H. Liu, "Feature selection for high-dimensional data: A fast correlation-based filter solution,"
9] *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pp. 856-863, 2003.
- [5 R. Kohavi and G. John, "Relevance wrappers for feature subset selection," *Artif. Intell.*, vol. 1, no. 97,
0] p. 273-324, 1997.
- [5 T. Lal, O. Chapelle, J. Weston and A. Elisseeff, "Embedded Methods," *Feature Extraction: Foundations
1] and Applications.*, pp. 137-165, 2006.
- [5 K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm,"
2] *AAAI-92 Proceedings*, 1992.
- [5 I. Guyon, J. Weston, S. Barnhill and V. Vapnik, "Gene selection for cancer classification using support
3] vector machines," *MACHLEARN: Machine Learning*, no. 46, pp. 389-422, 2002.
- [5 S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, Academic Press, 2006.
4]
- [5 C. DING and H. PENG, "MINIMUM REDUNDANCY FEATURE SELECTION FROM MICROARRAY GENE
5] EXPRESSION DATA," *Journal of Bioinformatics and Computational Biology*, vol. 2, no. 3, 2003.
- [5 M. S. d. Alencar, *Communication, Management and Information Technology: International*, London,
6] UK: Taylor & Francis group, 2017.

- [5 N. S. Altman, "An Introduction to Kernel and Nearest Neighbor Nonparametric Regression," *The American Statistician*, vol. 3, no. 46, pp. 175-185, 1992.
- [5 M. E. Maron, "Automatic Indexing: An Experimental Inquiry," *Journal of the ACM.*, vol. 3, no. 8, pp. 404-417, 1961.
- [5 S. C. Kleene, "Representation of Events in Nerve Nets and Finite Automata," *Annals of Mathematics Studies*, 1956.
- [6 R. Kohavi, "A study of cross-validation and bootstrap for accuracy estimation and model selection," *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, vol. 2, no. 12, p. 1137-1143, 1995.
- [6 G. J. McLachlan, C. Ambroise and K.-A. Do, "Analyzing microarray gene expression data," *Wiley*, 2004.
- [6 T. Hastie, R. Tibshirani and J. Friedman, *Elements of Statistical Learning: data mining, inference, and prediction*. 2nd Edition, Springer, 2009.
- [6 A. T. Berg, S. F. Berkovic, M. J. Brodie, J. Buchhalter, J. H. Cross and W. v. E. Boas, "Revised terminology and concepts for organization of seizures and epilepsies: report of the ILAE Commission on Classification and Terminology, 2005-2009," *Epilepsia*, vol. 51, pp. 676-685, 2010.
- [6 J. W. Bang, J.-S. Choi and K. R. Park, "Noise Reduction in Brainwaves by Using Both EEG Signals and Frontal Viewing Camera Images," *Sensors (Basel)*, p. 6272-6294, 2013.
- [6 A. Hyvärinen, "Independent component analysis: recent advances," *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, pp. 1-19, 2013.
- [6 G. Repovš, "Dealing with Noise in EEG Recording and Data Analysis," *Infor Med Slov*, pp. 18-25, 2010.
- [6 G. Diaz, "GrepMed," [Online]. Available: <https://www.grepmed.com/images/1480/positivepredictivevalue-negativepredictivevalue-epidemiology-specificity-calculation-sensitivity-statistics>.
- [6 M. Sharma, "Researchgate," Classification of ictal and non-ictal EEG signals by LS-SVM classifier for the best two features f1=FD of SB-17 and f2= FD of SB-2., [Online]. Available: https://www.researchgate.net/publication/315582822_A_new_approach_to_characterize_epileptic_seizures_using_analytic_time-frequency_flexible_wavelet_transform_and_fractal_dimension/figures?lo=1.
- [6 "World Health Organization (WHO) report, fact sheet No 999," February 2018.
- [7 R. S. Fisher, W. v. E. Boas, W. Blume, C. Elger, P. Genton, P. Lee and e. al., "Epileptic Seizures and Epilepsy: Definitions Proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE)," *Epilepsia*, vol. 46, pp. 470-472, 2005.
- [7 G. Giannakakis, V. Sakkalis, M. Pediaditis and M. Tsiknakis, "Methods for seizure detection and prediction: an overview," *Modern Electroencephalographic Assessment Techniques, Neuromethods*, vol. 91, pp. 131-157, 2015.
- [7 V. Sakkalis, G. Giannakakis, C. Farmaki, A. Mousas, M. Pediaditis and P. Vorgia, "Absence seizure epilepsy detection using linear and nonlinear EEG analysis methods," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013.
- [7 G. Giannakakis, V. Sakkalis, M. Pediaditis, C. Farmaki, P. Vorgia and M. Tsiknakis, "An approach to absence epileptic seizures detection using Approximate Entropy," in *35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2013.

- [7 G. Giannakakis, M. Tsiknakis and P. Vorgia, "Focal Epileptic Seizures Anticipation based on Patterns of
4] Heart Rate Variability Parameters," *Computer Methods and Programs in Biomedicine*, vol. 178, pp. 123-133, 2019.
- [7 Y. Zhang, G. Zhou, J. Jin, X. Wang and A. Cichocki, "Optimizing spatial patterns with sparse filter bands
5] for motor-imagery based brain-computer interface," *Journal of neuroscience methods*, vol. 255, pp. 85-91, 2015.
- [7 W. Wu, Z. Chen, X. Gao, Y. Li, E. N. Brown and S. Gao, "Probabilistic common spatial patterns for
6] multichannel EEG analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, pp. 639-653, 2014.
- [7 N. Tomida, T. Tanaka, S. Ono, M. Yamagishi and H. Higashi, "Active data selection for motor imagery
7] EEG classification," *IEEE Transactions on Biomedical Engineering*, vol. 62, pp. 458-467, 2014.
- [7 P. Xu, X. Xiong, Q. Xue, P. Li, R. Zhang and Z. Wang, "Differentiating between psychogenic nonepileptic
8] seizures and epilepsy based on common spatial pattern of weighted EEG resting networks," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 1747-1755, 2014.
- [7 T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaibi and S. R. Alrshoud, "Epileptic seizure prediction using CSP
9] and LDA for scalp EEG signals," *Computational intelligence and neuroscience*, vol. 2017, 2017.
- [8 Y. Zhang, Y. Wang, G. Zhou, J. Jin, B. Wang and X. Wang, "Multi-kernel extreme learning machine for
0] EEG classification in brain-computer interfaces," *Expert Systems with Applications*, vol. 96, pp. 302-310, 2018.
- [8 Z. Hong-xin, W. Hai-qing, L. Xiao-ming, L. Ying-hua and Z. Li-kun, "Implementation of compressive
1] sensing in ECG and EEG signal processing," *ScienceDirect*, vol. 17, pp. 122-126, 2010.
- [8 G. Kutyniok, "Compressed Sensing: Theory and Applications," 2012.
2]
- [8 E. Candes, J. Romberg and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from
3] highly incomplete Fourier information," *IEEE Trans. Inform. Theory*, vol. 52, pp. 489-509, 2006.
- [8 S. Pincus and R. Viscarello, "Approximate Entropy: A regularity Measure for Fetal Heart Rate Analysis,"
4] *Obstetrics Gynecology*, vol. 79, no. 2, pp. 249-255, 1992.
- [8 A. Fleisher, S. Pincus and S. Rosenbaum, "Approximate entropy of Heart Rate as a Correlate of
5] Postoperative Ventricular Dysfunction," *Anesthesiology*, vol. 78, no. 4, pp. 683-692, 1993.
- [8 S. Dawes, M. Moulden, O. Sheil and C. Redman, "Approximate Entropy, a Statistic of Regularity,
6] Applied to Fetal Heart Rate Before and During Labor," *Obstetrics Gynecology*, vol. 80, no. 5, pp. 763-768, 1992.
- [8 M. Tulpo, T. Makikallio, E. Timo, T. Seppanen and H. Huikuri, "Quantitative beat-to-beat analysis of
7] heart rate dynamics exercise," *The American Physiological Society*, pp. 244-251, 1996.
- [8 H. Timo, M. Makikallio, T. Seppanen, M. Niemela, A. Juhani, M. Tulppo, M. Heikki and F. Huikuri,
8] "Abnormalities in Beat to Beat Complexity of Heart rate Dynamics," *Patients With a Previous Myocardial Infarction*, vol. 28, no. 4, pp. 1005-1011, 1996.
- [8 J. Palazzolo, F. Estafanous and P. Murray, "Entropy measures of heart rate variation in conscious
9] dogs," *Am. J. Physiol*, vol. 274, no. 5, pp. 1099-1105, 1998.
- [9 E. J. Candès, Y. Eldar, D. Needell and P. Randall, "Compressed Sensing with Coherent and Redundant
0] Dictionaries," *Appl. Comput. Harmon. Anal.*, vol. 31, pp. 59-73, 2011.
- [9 E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Found. of Comput.*
1] *Math.*, vol. 9, pp. 717-772, 2008.

- [9 F. Krahmer and R. Ward, "New and improved Johnson-Lindenstrauss embeddings via the Restricted 2] Isometry Property," *SIAM J. Math. Anal.*, vol. 43, pp. 1269-1281, 2011.
- [9 G. Pfander, H. Rauhut and J. Tropp, "The restricted isometry property for timefrequency structured 3] random matrices," Preprint, 2011.
- [9 M. Steriade and R. Llinas, "The functional states of the thalamus and the associated neuronal 4] interplay," *Physiological Reviews*, vol. 68, pp. 649-742, 1988.
- [9 O. Knudsen, "Biological Membranes: Theory of Transport, Potentials and Electric Impulses," 5] Cambridge University Press, 2002.
- [9 A. E. Society, "Guideline thirteen: Guidelines for standard electrode position nomenclature," *J Clin 6] Neurophysiol*, vol. 11, pp. 111-113, 1994.
- [9 A. E. Society, "Guideline thirteen: Guidelines for standard electrode position nomenclature," *J Clin 7] Neurophysiol*, vol. 1994, pp. 111-113, 1994.
- [9 S. J. M. Smith, "EEG in the diagnosis, classification and management of patients with epilepsy," *J 8] Neurol Neurosurg Psychiatry*, no. 76, 2005.