



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Clustering Brain Regions by Similar Interaction Patterns Based on Multivariate Neural Signals for Identifying the Response to Antidepressants“

verfasst von / submitted by

Mykola Lazarenko

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2023 / Vienna 2023

Studienkennzahl lt. Studienblatt /
degree programme code as it appears on
the student record sheet:

UA 066 977

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Business Analytics

Betreut von / Supervisor:

Univ.-Prof. Dipl.-Inform. Univ. Dr. Claudia Plant

Mitbetreut von / Co-Supervisor:

RNDr. CSc. Katerina Schindlerova

Abstract

Major depressive disorder (MDD) or clinical depression is a mental disorder that is characterized by low self-esteem, persistent sadness, and loss of pleasure in activities that are normally enjoyable. Modern ways of treatment have a response of only up to 65% and the response can usually be seen after 4-6 weeks of treatment.

Understanding the interactions of brain regions of responders and non-responders can help to avoid trials of ineffective therapy and help in further research of the disease. Moreover, the choice of effective therapy leads to a decrease in costs for the treatment. We postulate, that patients who responded to antidepressant treatment have altered information patterns obtainable from EEG recordings compared to those who have not responded to it. One way of finding patterns in the data is by applying of clustering algorithms. A data basis of EEG recordings of patients with 19 electrodes is given.

In this work, we apply the Interaction K-means (IKM) clustering algorithm on multi-trial EEG data of patients with major depressive disorder. This algorithm works effectively on multi-variate time-series as well as it allows us to derive interaction patterns from the obtained results. We performed an exploratory analysis of the influence of preprocessing on the cluster purity of IKM on the EEG data set. Different preprocessing strategies to improve the clustering accuracy of the algorithm on the given data were proposed, namely bands extraction, Hilbert, Box-Cox and z-score transformations, and others. Also, the interaction between electrodes was explored. Additionally, the popular clustering methods were applied to the parameters (coefficients) derived from the data using the least squares method.

The best clustering result is the one using the combination of the Box-Cox and z-score transformation methods and taking into consideration only Cz, Fp1, F3, F7, C3, T3, P3, T5, O1 electrodes (left-located electrodes including the central Cz). The cluster purity for this setting was 60.5%. Almost the same result was shown for clustering only on the Delta band obtained from the data of the same set of electrodes. The cluster purity was 60.3%. The same cluster purity was observed for the derived Hilbert amplitude from the Delta band from the data of the same set of electrodes.

In the context of the current study, a CP of 60% for the IKM clustering algorithm applied to multi-trial EEG data of patients with major depressive disorder is a reasonable result. EEG data is known to be highly variable and noisy, and the underlying biological mechanisms of depression are complex and not fully understood. Therefore, achieving a CP of 0.6 or higher on this type of data is a challenging task, and the results obtained using the proposed preprocessing strategies are considered promising.

When interpreting the data used for the best clustering, two electrodes were identified as the most discriminative among the clusters. They are P3 and T5 that located in the left and back regions of the head. In the first cluster, P3 and T5 have the strongest relationship with the Fp1 and F3 (left and frontal-located electrodes) respectively. In the second cluster, P3 has a strong connection to F3, Fp1, F7 (left and frontal located) and Cz (centrally located) electrodes while T5 is strongly related to the O1 (left and back located) electrode.

Zusammenfassung

Die Major-Depression oder klinische Depression ist eine psychische Störung, die sich durch ein geringes Selbstwertgefühl, anhaltende Traurigkeit und den Verlust der Freude an Aktivitäten, die normalerweise angenehm sind, auszeichnet. Moderne Behandlungsmethoden haben eine Ansprechrate von nur bis zu 65%, und die Reaktion ist in der Regel erst nach 4-6 Wochen Behandlung zu sehen.

Das Verständnis der Interaktionen zwischen den Hirnregionen von Patienten, die auf die Behandlung ansprechen und solchen, die nicht darauf ansprechen, kann helfen, unnütze Therapieversuche zu vermeiden und weitere Forschung zur Krankheit zu ermöglichen. Darüber hinaus führt die Wahl einer effektiven Therapie zur Verringerung der Kosten für die Behandlung. Wir postulieren, dass Patienten, die auf eine antidepressive Behandlung angesprochen haben, im Vergleich zu denen, die nicht darauf angesprochen haben, veränderte Informationsmuster aufweisen, die aus EEG-Aufzeichnungen abgeleitet werden können. Eine Möglichkeit, Muster in den Daten zu finden, besteht darin, Clustering-Algorithmen anzuwenden. Eine Datengrundlage von EEG-Aufzeichnungen von Patienten mit 19 Elektroden liegt vor.

In dieser Arbeit wenden wir den Interaction K-means (IKM) Clustering-Algorithmus auf Multi-Trial-EEG-Daten von Patienten mit Major Depression an. Dieser Algorithmus funktioniert effektiv bei Multi-Variate Zeitreihen und ermöglicht es uns, Interaktionsmuster aus den erhaltenen Ergebnissen abzuleiten. Wir führten eine explorative Analyse des Einflusses der Vorverarbeitung auf die Clusterreinheit von IKM auf dem EEG-Datensatz durch. Verschiedene Vorverarbeitungsstrategien wurden vorgeschlagen, um die Clustering-Genauigkeit des Algorithmus auf den gegebenen Daten zu verbessern, nämlich Bandextraktion, Hilbert-, Box-Cox- und Z-Score-Transformationen und andere. Auch die Interaktion zwischen den Elektroden wurde untersucht. Darüber hinaus wurden die beliebten Clustering-Methoden auf die aus den Daten abgeleiteten Parameter (Koeffizienten) angewendet, die mit der Methode der kleinsten Quadrate berechnet wurden.

Das beste Clustering-Ergebnis ergibt sich aus der Kombination der Box-Cox- und Z-Score-Transformation und der Berücksichtigung nur der Cz-, Fp1-, F3-, F7-, C3-, T3-, P3-, T5-, O1-Elektroden (linksseitig gelegene Elektroden einschließlich der zentralen Cz). Die Clusterreinheit für diese Einstellung betrug 60,5%. Fast das gleiche Ergebnis wurde beim Clustering nur des Delta-Bandes aus den Daten desselben Elektrodensatzes erzielt. Die Clusterreinheit betrug 60,3%. Die gleiche Clusterreinheit wurde für die abgeleitete Hilbert-Amplitude aus dem Delta-Band aus den Daten desselben Elektrodensatzes beobachtet.

Im Kontext der vorliegenden Studie ist eine CP von 60% für den IKM-Clustering-Algorithmus, der auf Mehrfach-EEG-Daten von Patienten mit Major Depression angewendet wird, ein vernünftiges Ergebnis. EEG-Daten sind bekanntlich hoch variabel und rauschig und die zugrunde liegenden biologischen Mechanismen der Depression sind komplex und nicht vollständig verstanden. Daher ist es eine anspruchsvolle Aufgabe, auf diesem Datentyp eine CP von 0,6 oder höher zu erreichen, und die Ergebnisse, die mit den vorgeschlagenen Vorverarbeitungsstrategien erzielt wurden, gelten als vielversprechend.

Bei der Interpretation der Daten, die für das beste Clustering verwendet wurden, wurden zwei Elektroden als am diskriminierendsten unter den Clustern identifiziert. Es handelt sich um P3

und T5, die sich in den linken und hinteren Regionen des Kopfes befinden. Im ersten Cluster haben P3 und T5 die stärkste Beziehung zu den Fp1- bzw. F3-Elektroden (links und frontal gelegen). Im zweiten Cluster hat P3 eine starke Beziehung zu den F3-, Fp1-, F7- (links und frontal gelegen) und Cz (zentral gelegen) Elektroden, während T5 stark mit der O1 (links und hinten gelegen) Elektrode verbunden ist.

Acknowledgements

I want to express my gratitude to my co-supervisor Dr. Katerina Schindlerova and supervisor Prof. Dr. Claudia Plant. A particular thank you goes to Katerina for our brainstorming sessions, her valuable ideas, feedback, and input and, of course, incredible support during my journey of working on this thesis. Thank you to MSc. Lena Bauer for your beneficial input and suggestions. Additional thank you to the colleagues from the project which this thesis was part of, namely Dr. Jaroslav Hlinka, Dr. Milan Paluš, Dr. Martin Brunovsky for your input and feedback during our synchronisation meetings. A special shout out goes to Dr. Zherdin for his explanations about the details and principles of the implementation of the Interaction K-means algorithm.

Contents

Abstract	2
Zusammenfassung.....	3
Acknowledgements.....	5
1 Introduction	11
1.1 Motivation	11
1.2 Objectives and Contribution.....	11
1.3 Thesis Structure.....	12
2 Background.....	13
2.1 Major Depressive Disorder	13
2.2 Electroencephalography (EEG).....	14
2.3 Time-series	16
3 Clustering Algorithms	17
3.1 k-means	17
3.2 DBSCAN.....	18
3.3 Hierarchical Clustering	19
3.4 Interaction k-means	20
3.4.1 Definitions.....	20
3.4.2 Definition of Models.....	20
3.4.3 Aggregative Pre-Computing	21
3.4.4 Distance Errors Used	22
3.4.5 IKM Algorithm	22
3.4.6 Model Finding.....	23
3.5 Dimensions Selection.....	24
3.6 Interpretation of the Results	25
3.7 Criteria of the Quality of Clustering	26
3.7.1 Cluster Purity	26
3.7.2 Rand Index	27
3.7.3 Information Criterion	27
4 Data Sets and Methods	28
4.1 Data Characteristics.....	28
4.1.1 Data of Alcoholic and Non-alcoholic Patients.....	28
4.1.2 Data of Depressed Patients	28
4.2 Data Exploration	29

4.2.1	3-dimensional Trajectories of the EGG Data of Depressed Patients	29
4.2.2	Comparison of Topomaps of Depression and Alcoholic EEG Data.....	30
4.2.3	Coefficients Correlation Matrices.....	33
4.2.4	Distances Between Coefficients	37
4.3	General Data Preprocessing	38
4.4	Used Tools.....	39
5	Preprocessing Methods.....	40
5.1	Alcoholic and Control EEG Data	40
5.2	EEG Data of Depressed Patients	40
5.2.1	Determination of Optimal Number of Clusters	40
5.2.2	Untransformed Data.....	41
5.2.3	Sinus Non-linear Function	41
5.2.4	Downsampling	42
5.2.5	Dimensions selection	44
5.2.6	Box-Cox Transformation	45
5.2.7	Discrete Wavelet Transformation	45
5.2.8	z-normalization	46
5.2.9	z-score	47
5.2.10	Z-transform	48
5.2.11	Exponential Smoothing.....	49
5.2.12	EEG Bands Extraction	52
5.2.13	Hilbert Transform	53
5.2.14	Selected Electrodes	55
5.2.15	Clustering on Coefficients	56
5.2.16	Different Distance Error Types.....	58
5.2.17	Mixed Transformations.....	59
5.3	Interpretation of the Results	60
6	Conclusion and Outlook	65
	Bibliography	67
	Appendices.....	71
	Appendix A. Cluster Coefficients Correlation Matrices	71
	Appendix B. Each Object Coefficients Correlation Matrices	75
	Appendix C. Clustering Results on the Different Parts of Electrodes Using Mixed Transformations	80
	Appendix D. Results Interpretation Details	83

List of Figures

Figure 2.1 – Structural and functional brain abnormalities in patients with major depressive disorder (Illustration from Rot, Mathew, and Charney 2009).	14
Figure 2.2 – 19 electrode locations of (electroencephalography) recording. (“A” electrodes here refer to the ear electrodes that are attached to the mastoid bone behind the ear.) (Illustration from Kruk et al. 2014).	15
Figure 2.3 – 19 time-series of the EEG of the depressed patient.	16
Figure 3.1 – Three clusters as an outcome of the k-means algorithm run with k=3. The separation was based on two variables: texture and radius (Illustration from Harezlak, 2022).	18
Figure 3.2 – Two generated clusters using the DBSCAN algorithm (Illustration from DiFrancesco, Bonneau, and Hutchinson 2020).	19
Figure 3.3 – The results of hierarchical clustering presented as a tree of clusters called dendrogram (Illustration from Halkidi, 2009).	19
Figure 3.4 – Interaction pattern between different time series. The signal dim12 is represented by a linear combination of other dimensions, namely dim4, dim6, and dim5 (Illustration from Plant et al. 2013).	20
Figure 3.5 – A diagram representing the IKM algorithm (adopted from Plant et al. 2013).	23
Figure 3.6 – A diagram representing the model finding algorithm (Adopted from Böhm et al. 2009).	23
Figure 3.7 – A diagram representing the dimension selection algorithm.	25
Figure 3.8 – A diagram representing the interpretation algorithm (Adopted from Plant et al. 2013).	26
Figure 4.1 – EEG data of the first visit of the subject 1 who responded to the treatment.	29
Figure 4.2 – EEG data of the second visit of the subject 1 who responded to the treatment.	30
Figure 4.3 – Snapshots of a topomap of an alcoholic subject.	31
Figure 4.4 – Snapshots of a topomap of an alcoholic subject.	31
Figure 4.5 – Snapshots of a topomap of the first visit of a depressed subject.	32
Figure 4.6 – Snapshots of a topomap of the second visit of a depressed subject.	32
Figure 4.7 – Correlation matrix for the cluster coefficients of ideal clusters. The coefficients are calculated for the first visit non-responders and the second visit responders.	34
Figure 4.8 – Correlation matrix for the cluster coefficients of ideal clusters. The coefficients are calculated for the first visit responders and the second visit responders.	35
Figure 4.9 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the first visit non-responders and the second visit responders.	36
Figure 5.1 – The plot of the clusters error value with regard to a number of clusters.	41
Figure 5.2 – The Wavelet Daubechies 4 (db4) waveform (Illustration from Belkhou, Jbari, and Belarbi 2017).	46
Figure 5.4 – Marked in green are electrodes chosen for the clustering to obtain the best clusters purity results.	60
Figure 5.5 – Error values of electrodes models of “ideal” clusters. Positive values of errors here mean that most subjects were allocated in the wrong cluster.	61

Figure 5.6 – Error values of Cz, Fp1, F3, F7, C3, T3, P3, T5, O1 models applying Box-Cox and z-score transformations for clustering. The smaller the error the better the model of the electrode discriminate among the clusters.....	62
Figure 5.7 – The electrodes for “non-responders” cluster that discriminate the most (i.e., having the highest error value) among two clusters of the best clustering setting and their relationship to other electrodes.....	63
Figure 5.8 – The electrodes for “responders” cluster that discriminate the most among two clusters and their relationship to other electrodes.....	64
Figure A-1 – Correlation matrix for cluster coefficients of ideal clusters. The coefficients are calculated for the first visit non-responders and the second visit non-responders.....	71
Figure A-2 – Correlation matrix for cluster coefficients of ideal clusters. The coefficients are calculated for the first visit non-responders and the first visit responders.....	72
Figure A-3 – Correlation matrix for cluster coefficients of ideal clusters. The coefficients are calculated for the first visit responders and the second visit non-responders.....	73
Figure A-4 – Correlation matrix for cluster coefficients of ideal clusters. The coefficients are calculated for the second visit non-responders and the second visit responders.....	74
Figure B-1 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the first visit non-responders and the second visit non-responders.....	75
Figure B-2 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the first visit non-responders and the first visit responders.....	76
Figure B-3 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the first visit responders and the second visit responders.....	77
Figure B-4 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the first visit responders and the second visit non-responders.	78
Figure B-5 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the second visit non-responders and the second visit responders.	79
Figure D-1 – Relationships between electrodes of the first cluster with non-responders of “ideal” clusters.....	83
Figure D-2 – Relationships between electrodes of the second cluster with responders of “ideal” clusters.....	84
Figure D-3 – Relationships between electrodes of the first cluster obtained with the clustering setting of applying Box-Cox and z-score transformations on the left hemisphere electrodes + Cz.....	85
Figure D-4 – Relationships between electrodes of the second cluster obtained with the clustering setting of applying Box-Cox and z-score transformations on the left hemisphere electrodes + Cz.....	86

List of Tables

Table 4.1 – Structure of the depressed patients' dataset. Only first three rows are presented.	29
Table 4.2 – Frobenius distance between coefficients of ideal clusters.	37
Table 4.3 – Frobenius distance between the coefficients of each object allocated to ideal clusters.	38
Table 5.1 – Results of clustering on the alcoholic-non-alcoholic EEG dataset with 60 objects and 64 electrodes (dimensions)	40
Table 5.2 – Results of clustering on the not-preprocessed EEG data of depressed patients.	41
Table 5.3 – Results of clustering on the data with applied sinus function	42
Table 5.4 – Results of clustering on the initial one-third of the data	42
Table 5.5 – Results of clustering on the middle one-third of the data	43
Table 5.6 – Results of clustering on the last one-third of the data	43
Table 5.7 – Results of clustering on the data with every 5 th element removed.	43
Table 5.8 – Results of clustering on the data with every 10 th element removed.	43
Table 5.9 – Results of clustering on the data with every 5 th element left.	43
Table 5.10 – Results of clustering on the data with every 10 th element left.	44
Table 5.11 – Results from applying dimensions selection modification of the IKM algorithm.	44

Table 5.12 – Results of clustering on the data with applied Box-Cox transformation	45
Table 5.13 – Results of clustering on the data with applied Discrete Wavelet transformation.	46

Table 5.14 – Results of clustering on the data with applied z-normalization	47
Table 5.15 – Results of clustering on the data with applied z-score measure	48
Table 5.16 – Results of clustering on the data with extracted magnitude applying Z-transform.	49

Table .517 – Results of clustering on the data with extracted phase applying Z-transform	49
Table 5.18 – Results of clustering on the data that was simple exponentially smoothed	51
Table 5.19 – Results of clustering on the data that was double exponentially smoothed	51
Table 5.20 – Results of clustering on the data that was triple exponentially smoothed	51
Table 5.21 – Results of clustering on the bands phase obtained from Hilbert transformation.	54

Table 5.22 – Results of clustering on the bands amplitude obtained from Hilbert transformation.	55

Table 5.23 – Results of clustering on the specific electrodes on the non-transformed data	56
Table 5.24 – Results of clustering on the coefficients obtained from the different parts of the data	57
Table 5.25 – Results of clustering on the reduced coefficients matrices	58
Table 5.26 – Results of clustering on the coefficients matrices of the reduced dataset	58
Table 5.27 – Results of clustering on the combined dataset with the different distance errors.	59

Table 5.28 – Results of clustering on the combined dataset with the different distance errors considering only right located electrodes.	59
Table C-1 – Results of clustering on the different parts of electrodes and using mixed transformations.	80

1 Introduction

1.1 Motivation

Major depressive disorder impacts the lives of many people every day. It has serious health, social, and economic consequences for humanity. Also, there is a lack of objective methods to diagnose this disease. Thus, this study intends to provide methods for better understanding the major depressive disorder and to support clinicians in the process of diagnosing it to reduce both monetary and time costs.

This Master's thesis was written as a part of the joint international research project "Learning Synchronization Patterns in Multivariate Neural Signals for Prediction of Response to Antidepressants" of the Data Mining and Machine Learning Research Group of the University of Vienna, of the Institute of Computer Science, the Czech Academy of Sciences and the National Institute of Mental Health of the Czech Republic funded by public funding.

1.2 Objectives and Contribution

The main objective of the thesis was to cluster depressed patients, who were given antidepressant medications in two clusters: responders and non-responders. Additionally, this study intended to find interaction patterns between electrodes. The main clustering algorithm used was Interaction K-means for time series as proposed by Plant et al. in 2013.

The contribution of this thesis was applying various preprocessing techniques to the EEG data basis to improve the results of the clustering. We applied such transformations as non-linear transformation, Box-Cox, discrete wavelet transformation, z-normalization, z-score, Z-transform, bands extraction, Hilbert, exponential smoothing, and others.

After obtaining the results from different clustering settings, the best one was using the combination of the Box-Cox and z-score transformation methods and taking into consideration only Cz, Fp1, F3, F7, C3, T3, P3, T5, O1 electrodes (left located electrodes including the central Cz). Thus, we concluded that these specific transformations together with only part of the electrodes works well for separating the data of depressed patients based on their response to the antidepressant treatment.

Most of other preprocessing strategies did not provide any satisfactory results. We hypothesize that it can be since trajectories of time series of depressive patients are relatively uniform in the comparison, for example, to the epilepsy patients having seizures, which makes any differentiation/separability of subsets of patients to be hard.

Given the challenging nature of multi-trial EEG data from patients with major depressive disorder, achieving a cluster purity (CP) value of 60% for the Interaction K-means (IKM) clustering algorithm is a reasonable outcome. EEG data is notoriously susceptible to high levels of variability and noise, and the underlying biological mechanisms behind depression are complex and not yet fully understood. Consequently, the results of the proposed preprocessing strategies are considered promising in achieving a CP of 0.6 or higher.

1.3 Thesis Structure

The thesis has six chapters. In the introduction chapter, the motivation, objectives, and contribution as well as the thesis structure are presented. The second chapter explains the main concepts covered in this work such as major depressive disorder, electroencephalography, and time series. The third chapter covers the clustering principles of k-means, DBSCAN and hierarchical clustering. This chapter also covers the details of Interaction k-means (IKM) algorithm, its enhanced version where the dimension selection process was applied as well as the interpretation algorithm that is used to interpret the results of IKM. Also, the third chapter explains the measures used to measure the quality of clustering. The fourth chapter shows in detail the characteristics of the data used for this study, data analysis results and how the data was processed to be ready to be used by the clustering algorithms. The fifth chapter presents the results of the repeated experiment using IKM on the EEG data of alcoholics and non-alcoholics, what data processing techniques were used to improve the results of IKM clustering and the results themselves and the chapter shows interpretation of the results obtained using IKM. Last but not least, the sixth section contains the conclusions of the work and ideas for possible future work.

2 Background

2.1 Major Depressive Disorder

Major depressive disorder (MDD) or clinical depression is a mental disorder characterized by low self-esteem, persistent sadness, and loss of pleasure in normally enjoyable activities. 16.6% of the population are estimated to have MDD one or more times during their lifetime (Kessler et al. 2005).

The diagnosis of the disease is based on the person's experience, behavior, and mental status examination such as Montgomery–Åsberg Depression Rating Scale (MADRS). No laboratory test can identify MDD (Patton, Glick, and American Dental Association 2016). MDD affects females twice as often as males and the common time of the disease offset is in a person's 20s (Kessler and Bromet 2013).

Currently, no mechanism can explain all aspects of the disease (Otte et al. 2016). It may be caused by a combination of psychological, environmental, and genetic factors with around 40% being genetic (Diagnostic and Statistical Manual of Mental Disorders: DSM-5TM, 5th Ed 2013). Chronic health problems, a family history of the condition, major life changes, and the persistent use of drugs are possible risk factors (Diagnostic and Statistical Manual of Mental Disorders: DSM-5TM, 5th Ed 2013).

MDD is typically treated with psychotherapy, antidepressant medication or brain stimulation such as repetitive transcranial magnetic stimulation (rTMS). Medication treatment seems to be effective but mostly in patients with severe depression (Fournier et al. 2010).

MDD is associated with structural and functional abnormalities of some parts of the brain, which were linked with MDD (Rot, Mathew, and Charney 2009). Figure 2.1 represents the different effects of the disease on the brain regions. The anterior cingulate, prefrontal, and orbitofrontal cortices, subgenual cingulate, ventral striatum, amygdala, and hippocampus may show a decrease in volume in patients with depression, whereas brain activity may increase in the amygdala and orbitofrontal cortex.

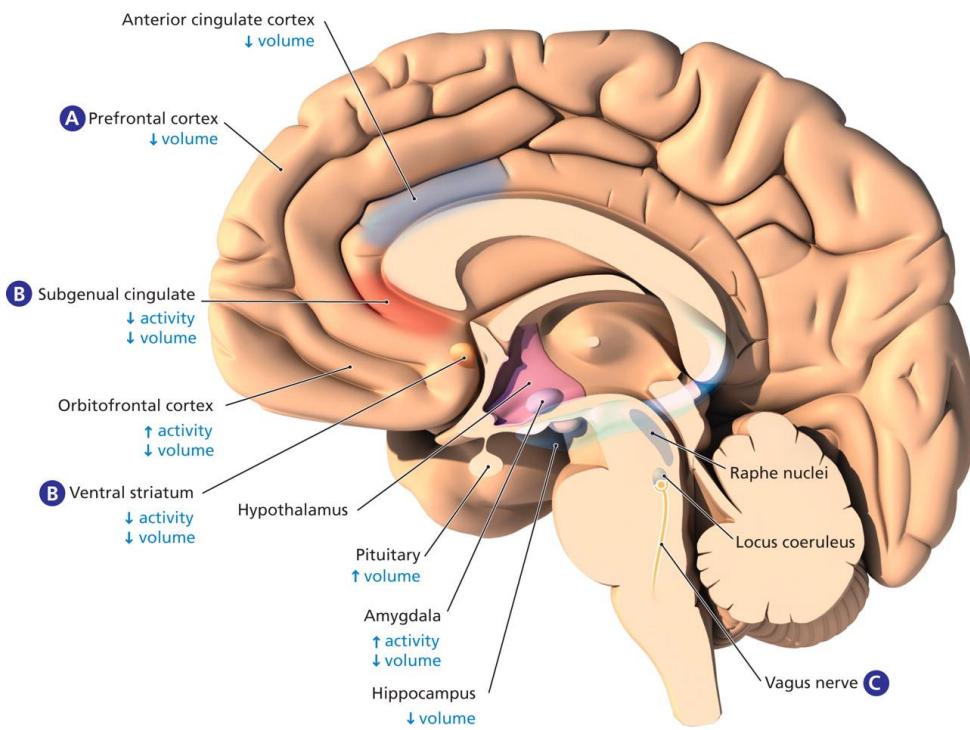


Figure 2.1 – Structural and functional brain abnormalities in patients with major depressive disorder (Illustration from Rot, Mathew, and Charney 2009).

2.2 Electroencephalography (EEG)

Electroencephalography (EEG) is a method of medical diagnostics and research, which measures the spontaneous electrical activity of the brain. Voltage fluctuation in individual brain cells which react to the brain activities causes the potential difference. Therefore, with EEG it is possible to observe what brain activity happened at different times in the various regions. This method is non-invasive.

Electrodes are placed along the scalp according to the International 10-20 system or its variations (Figure 2.2). The international 10-20 system is an internationally agreed method to describe and apply the location of electrodes (Herwig, Satrapia, and Schönfeldt-Lecuona 2003).

The names of electrodes are based on the location where they are placed on the scalp. These locations include: frontopolar (Fp), anterior frontal (AF), frontal (F), central (C), temporal (T), parietal (P), occipital (O), frontocentral (FC), frontotemporal (FT), centroparietal (CP), temporal-posterior temporal (TP), and parietooccipital (PO) (Acharya et al. 2016).

The electrodes are able to capture spatial information from a several square centimeters area of the scalp surface. However, the temporal resolution of these electrodes is often superior to their spatial resolution, indicating that they are better suited for detecting changes in activity over time rather than precise spatial localization.

Frequencies of the EEG recording can be up to 1 kHz. The data coming from each electrode is a long time series with thousands of values.

Different brain activities happen to relate to different frequencies of the waveforms. The frequency ranges are usually identified as delta (0-4 Hz), theta (4-7 Hz), alpha (8-12 Hz), beta(13-30Hz) and gamma (30-100 Hz) (Saby and Marshall 2012).

The amplitude in lower frequency bands is bigger, however, the bands are noisier. The opposite applies to higher-frequency bands. Their amplitudes are smaller so is the noise.

Delta (<4 Hz) frequencies are found frontally in adults. Normally, they relate to slow-wave sleep in adults and some continuous attention tasks (Kirmizi-Alsan et al. 2006).

Theta frequencies (4-7 Hz) can be found in different locations that are not related to a task at hand. They happen when there is drowsiness in adults and teens, idling, and active repression of a response or action (Kirmizi-Alsan et al. 2006).

Alpha frequencies (8-12 Hz) are usually present in posterior regions of the head, both sides. At rest, they are found in central sites (C3 and C4 electrodes). They are associated with a relaxed and reflecting state having eyes closed. Alpha waves are also more common in a coma.

Beta frequencies (13-30 Hz) are low-amplitude waves that are found on both sides having symmetrical distribution. They identify active thinking, focus, high alert, and anxious brain states.

Finally, gamma frequencies (30-100 Hz) are mainly found in the somatosensory cortex. They represent cross-modal sensory processing such as a combination of two different senses like sound and sight as well as short-term memory matching of recognized objects, tactile sensations, or sounds.

With EEG, the immediate reaction to the stimulation can be seen. Moreover, this method is relatively simple as it does not require any personal computer support since the voltage changes can be written on continuous-form paper.

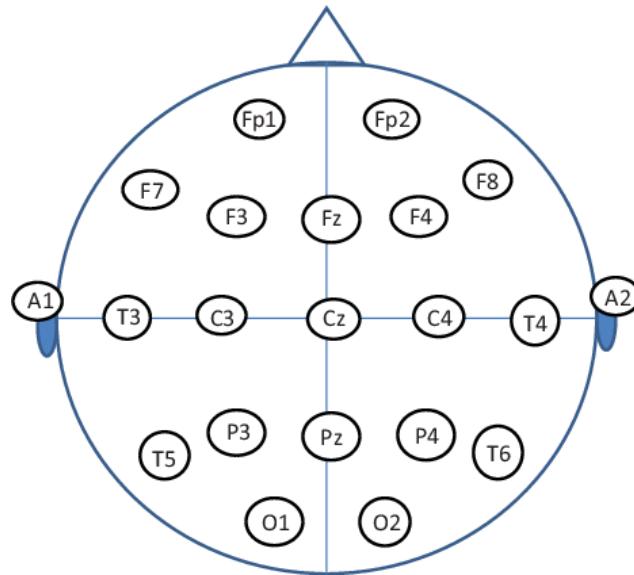


Figure 2.2 – 19 electrode locations of (electroencephalography) recording. (“A” electrodes here refer to the ear electrodes that are attached to the mastoid bone behind the ear.) (Illustration from Kruk et al. 2014).

2.3 Time-series

Time series data, which represents a sequence of values or events observed at specific time points with a consistent interval between them (Data Mining 2012), is often found in various scientific fields such as weather forecasting, scientific experiments, stock market analysis, and medical diagnostics. The use of sensors and other data collection tools has led to a large amount of time series data that needs to be processed quickly or in real time. A scientific discipline dealing with time series called time series analysis aims to both model and forecasts this data. Modelling involves identifying patterns in the data while forecasting attempts to predict future values.

Time series can be decomposed into four components: trend, cycles, seasonal, and irregular. The trend represents the overall direction of the data over a long period and is typically extracted using a moving average or least squares method. Cycles refer to long-term oscillations around the trend. Seasonal patterns are repetitive deviations that occur at specific times of the year and can be detected using autocorrelation analysis or seasonal index numbers. The irregular component represents random behavior introduced by uncertainties in the world.

In many applications, time series data is multivariate, meaning that each time point is represented by a vector of multiple time series variables. Multivariate time series are useful for describing the interactions between these variables.

Figure 2.3 represents the 19-dimensional time series from the EEG of a depressed patient. Each time series in this figure represents one electrode. In Figure, around 11 first seconds of a recording are represented.

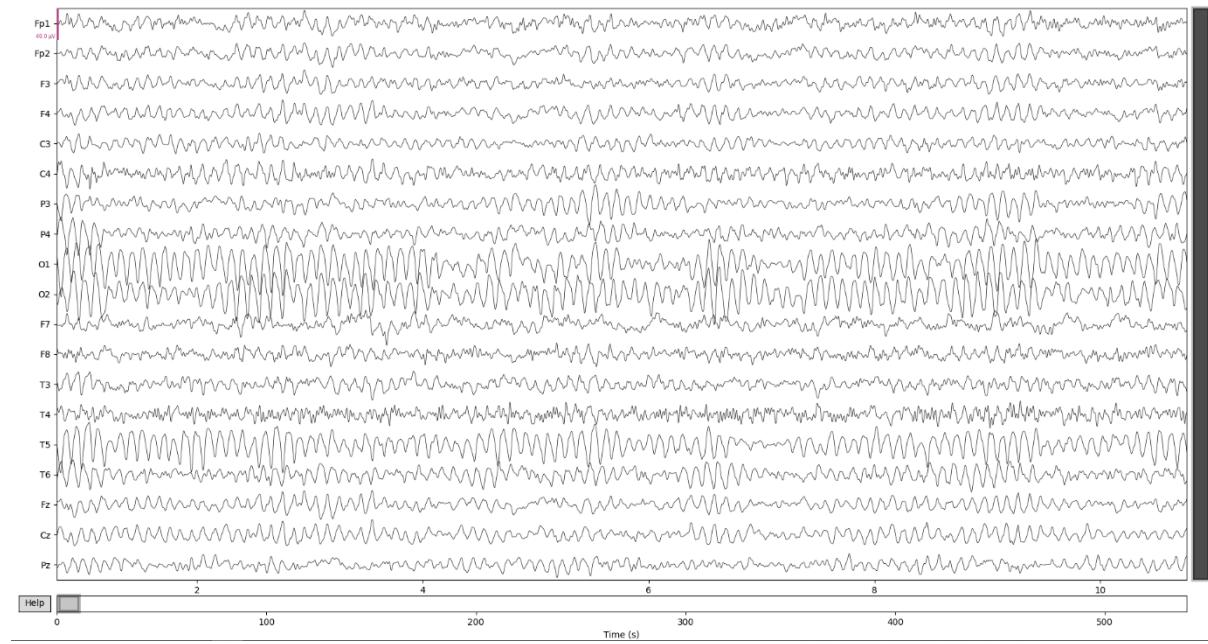


Figure 2.3 – 19 time-series of the EEG of the depressed patient.

3 Clustering Algorithms

The ultimate goal of any clustering method is to divide a set of objects into groups, or clusters, such that objects within the same cluster are like each other, while objects from different clusters are highly dissimilar. The aim of this work is to cluster depression patients into two clusters and get the interaction patterns between subjects within these clusters. The main clustering algorithm used for this is the interaction K-means (IKM), introduced in (Plant et al. 2013). This section covers the general principles of k-means, DBSCAN and hierarchical clustering as well as the details of IKM algorithm.

3.1 k-means

K-means clustering is a widely used method for partitioning a set of objects into clusters based on their similarity. This method was first proposed by (MacQueen 1967). It involves calculating the mean, or centroid, of each cluster and iteratively reassigning objects to their nearest centroid. The process begins with a randomly chosen set of k objects as the initial centroids for the clusters. Then, the distance from each object to each centroid is calculated, and each object is assigned to the cluster with the nearest centroid. The centroids are then recalculated based on the objects in their respective clusters. This process is repeated until the error of the current partitioning is minimized. Usually, the function that is taken for calculating the error is the square-error function (Zherdin, 2016):

$$E = \sum_{i=1}^k \sum_{o \in C_i} |o - m_i|^2$$

So, minimizing the sum of the squared distances of every object o to its mean m_i in each cluster C_i leads to the optimal solution. The iteration process in clustering should be stopped when the sequence of errors converges, and the clusters cease to change further.

Figure 3.1 represents three clusters separated using the k-means algorithm. The colors and shapes encode objects that are relevant to the specific clusters. Clustering was done considering two features: texture and radius.

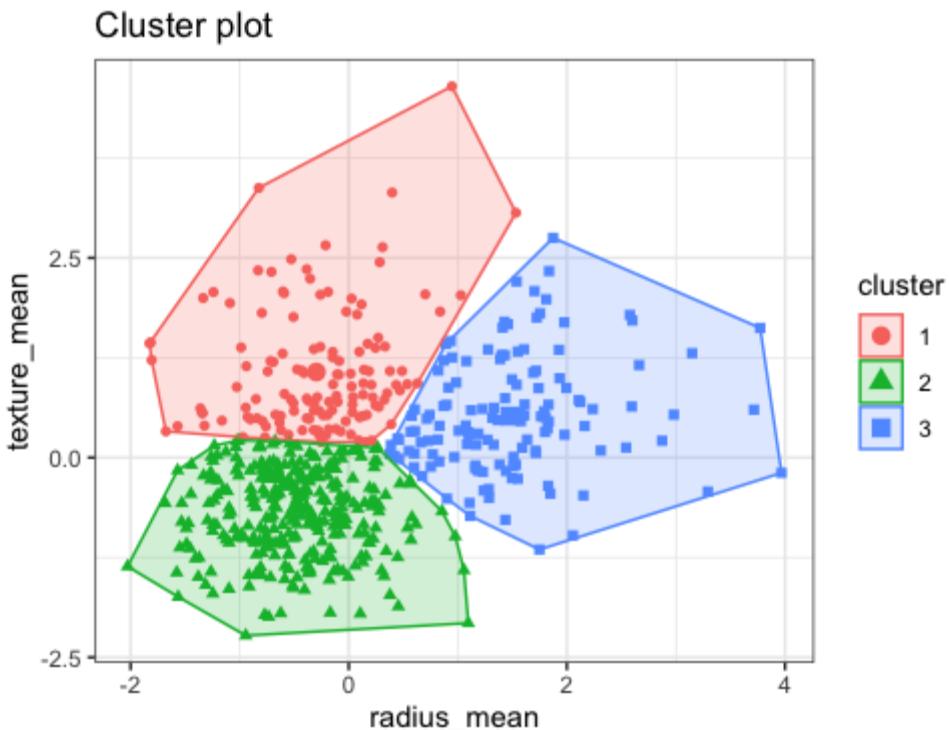


Figure 3.1 – Three clusters as an outcome of the k-means algorithm run with $k=3$. The separation was based on two variables: texture and radius (Illustration from Harezlak, 2022).

3.2 DBSCAN

The main value of another clustering algorithm Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is that it can find clusters of different shapes (Ester, Kriegel, and Xu, 1996). The basic concept is that clusters are determined by areas with a high concentration of objects, while regions with low object density define the boundaries between the clusters.

In the density-based method, a cluster is defined as the largest set of objects that are connected by density. The main objects, which have a sufficient number of objects in their vicinity as determined by a predetermined threshold, are used as the foundation for the density connectivity. An object is considered directly density-reachable concerning the threshold from a core object if it lies within the threshold-defined neighbourhood of the core object. As a result, a density-based cluster, which is a maximal set of density-connected objects, is obtained. The other objects that do not belong to any cluster are considered to be a noise (Zherdin, 2016).

Figure 3.2 shows the basic principles of the DBSCAN algorithm. There, red are main points, blue are border points and grey are noise points. The main points are those that meet the criteria of the cluster. In contrast, border points are those which do not fulfil the cluster criteria but are located within the vicinity of a main point. The DBSCAN algorithm follows two guidelines: (1) Points that are within the defined search radius of a main point are deemed to be part of its cluster, and (2) main points that have shared border points are part of the same cluster, as illustrated by p1 and p2 in Cluster 1 (DiFrancesco, Bonneau, and Hutchinson 2020).

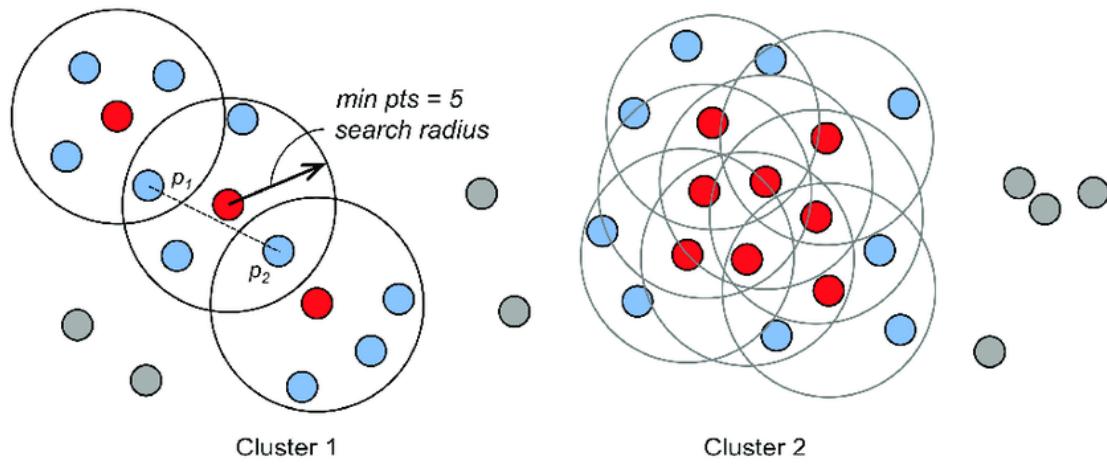


Figure 3.2 – Two generated clusters using the DBSCAN algorithm (Illustration from DiFrancesco, Bonneau, and Hutchinson 2020).

3.3 Hierarchical Clustering

Hierarchical clustering is a method for constructing a hierarchical decomposition of objects, which is achieved through a dendrogram (El Boucheiry and de Souza 2020). There are two main types of hierarchical clustering, the agglomerative and the divisive methods. Agglomerative methods have a bottom-up approach, starting with individual clusters for each object and merging them iteratively until only one cluster remains. The divisive method starts with one cluster and iteratively partitions it into smaller clusters.

In the agglomerative method, the steps are as follows: start with a set of clusters, each corresponding to an object. Then, merge the two closest clusters into a new parent inner node in the dendrogram. Repeat this process of finding and merging the closest clusters until only one cluster remains, which forms the dendrogram root. The methods differ in the way the distance between clusters is calculated (Zherdin, 2016). Figure 3.3 shows the process of the hierarchical clustering.

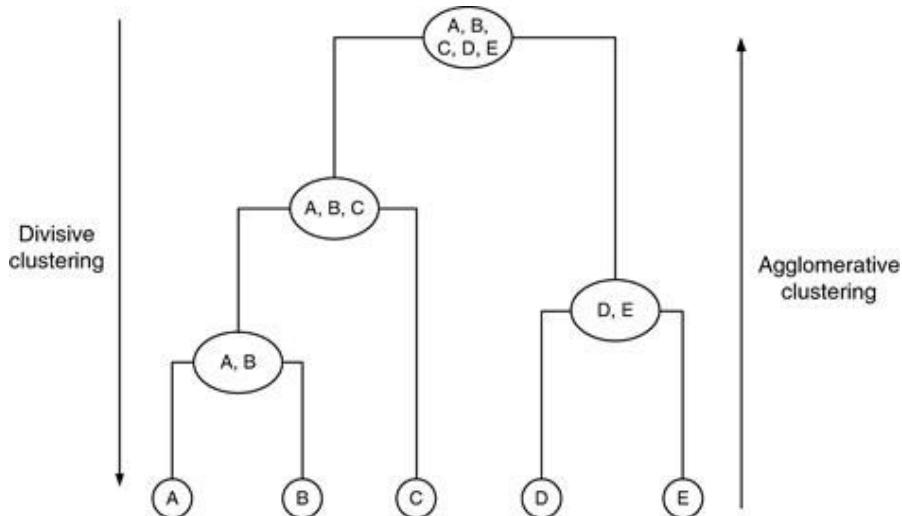


Figure 3.3 – The results of hierarchical clustering presented as a tree of clusters called dendrogram (Illustration from Halkidi, 2009).

3.4 Interaction k-means

The Interaction k-means algorithm is a partitioning clustering algorithm suitable to detect clusters of objects represented by time series with similar interaction patterns and was proposed by (Plant et al. 2013). The interaction patterns in this context are the relationship between different features of an object.

In Figure 3.4, a data object is shown as a multivariate time series. This data object has multiple dimensions, some of which exhibit a pattern of interaction. Specifically, the time series of dimension dim12 can be represented as a linear combination of dimensions dim4, dim5, and dim6. It is worth noting that not all dimensions of the data object necessarily interact in this way. The figure only shows the dimensions that are involved in the interaction pattern for the sake of simplicity (Plant et al. 2013).

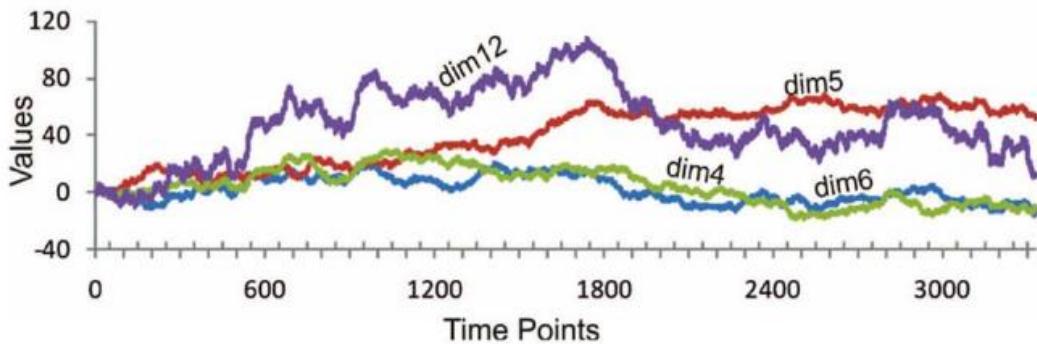


Figure 3.4 – Interaction pattern between different time series. The signal dim12 is represented by a linear combination of other dimensions, namely dim4, dim6, and dim5 (Illustration from Plant et al. 2013).

3.4.1 Definitions

Before explaining the specifics of the IKM algorithm, the formal definitions shall be made. All definitions in this section are taken from (Plant et al. 2013). The input to the algorithm is a set of objects. Each object O consists of multivariate time series. Object can be any entity, for example, an EEG data of a patient, the data from a sensor located in a house, the data of sensors that are used to track a subject motion and so on. Each time series has d number of dimensions and consists of n time points.

Additionally, n^* shows the number of time points for concatenated objects in a specific dimension. For example, if there are 3 objects each having the time series of length 200 in the dimension, then n^* for this specific dimension is 600.

3.4.2 Definition of Models

The IKM calculates models for finding the interaction patterns between them as well as being able to allocate objects in the clusters. The algorithm can work with two types of models: linear and non-linear.

Linear model is defined as next:

$$A = V \cdot P + E,$$

- where $A \in \mathbb{R}^{n^*}$ is the vector of values for the dependent variable,
- $V \in \mathbb{R}^{n^* \times |v|}$ is the matrix of independent variables, $|v|$ here is the number of these independent variables,
- $P \in \mathbb{R}^{|v|}$ is the vector of coefficients of the model,
- $E \in \mathbb{R}^{|v|}$ is the vector of errors. Errors are not assumed to be normal.

Non-linear model is defined similarly to the linear one. The only difference is the application of the non-linear function f to the multiplication of the matrix of independent variables V and the vector of coefficients P of the model:

$$A = f(V, P) + E,$$

where A, V, P, E are as defined for the linear model.

3.4.3 Aggregative Pre-Computing

The runtime of the IKM algorithm is largely determined by the number of objects being processed. Typically, only a few iterations are required, which keeps the overall complexity relatively low. The update step, however, is more computationally intensive as it involves matrix inversion and a greedy stepwise algorithm. Overall, the runtime of IKM scales linearly with the number of objects (Plant et al. 2013).

The use of aggregative pre-computing, as inspired by (Chen et al. 2006), allows the algorithm to be virtually independent of the length of the time series. This means that the runtime of the algorithm is not significantly affected by the length of the time series being processed.

According to (Chen et al. 2006) we compute $\theta_{ij}^* = \theta_{ji}^* := \sum_{k=1}^{m^*} Z_{ki} \cdot Z_{kj}$, $Z := [V|A]$ for aggregation matrix Θ^* . Here and further the “ \cdot ” operation is the operation of multiplication.

From this equation, the next terms can be derived (Plant et al. 2013):

$$V^T \cdot V = (\theta_{ij}^*), \text{ for } 1 \leq i \leq d-1, 1 \leq j \leq d-1 ,$$

$$V^T \cdot A = (\theta_{ij}^*), \text{ for } i = d, 1 \leq j \leq d-1$$

To find the E the following reduction is applied (Plant et al. 2013):

$$\begin{aligned} E^2 &= \|A - VP\|^2 = (A - VP)^T(A - VP) = A^T A - A^T VP - (VP)^T A + (VP)^T (VP) = \\ &= A^T A - (V^T A)^T P - P^T (V^T A) + P^T (V^T V) P, \text{ with } (V^T A) \in \mathbb{R}^{(d-1)} \text{ and } (V^T V) \in \\ &\quad \mathbb{R}^{(d-1) \times (d-1)} \text{ and } (A^T A) = \theta_{d,d} \in \mathbb{R}. \end{aligned}$$

The matrix Θ^* is calculated as (Plant et al. 2013):

$$\Theta^* = \sum_{i=1}^{|O|} \theta_i,$$

where $|O|$ is the number of objects in the cluster.

For object i , Θ^* is precalculated as (Plant et al. 2013):

$$(\theta_i)_{kj} = \sum_{l=1}^m Z_{lk} \cdot Z_{lj}$$

3.4.4 Distance Errors Used

In addition to the calculation of an error for the algorithm in the form of a Euclidian distance, other types of distances were used for the calculations (Balazia et al. 2022):

- *Total (absolute) norm distance:* $E = \|A - VP\|$ with the resulting matrix from the difference $\|B\| = \sum_{i,j=1}^p |b_{ij}|$
- *Max norm distance:* $E = \|A - VP\|_M$ with the resulting matrix from the difference $\|B\|_M = \max_{i,j=1,\dots,p} |b_{ij}|$
- *Hamming distance:* $E = \frac{(\| \max(A, VP) \|_2 - \| \min(A, VP) \|_2)}{n(n-1)}$ where $\max(\cdot, \cdot)$ and $\min(\cdot, \cdot)$ are matrices of element-wise minima and maxima, respectively. $\|\cdot\|_2$ represents the Frobenius norm of the matrix which is defined as $\|B\|_2 = \sqrt{\sum_{i,j=1}^p (b_{ij})^2}$
- *Jaccard distance:* $E = \frac{\| \min(A, VP) \|_2}{\| \max(A, VP) \|_2}$ where $\max(\cdot, \cdot)$ and $\min(\cdot, \cdot)$ as above.

3.4.5 IKM Algorithm

The diagram of the IKM algorithm is presented in Figure 2.5. All algorithms in this section are adopted from (Plant et al. 2013) or from (Böhm et al. 2009).

Before the initiation of the algorithm, the next variables must be setup:

- the number of algorithm initiations
- the maximum number of steps that algorithm should go through
- the k number of clusters.

Then, the next sequence of actions is performed:

1. Randomly allocate objects with time-series to k different clusters.
2. Find coefficients of regression models for each cluster.
3. Calculate the error of each object regarding each cluster. Errors are calculated considering every model in a cluster. For this purpose, the A vector of dependent variables is eliminated from the object and the model created for this dependent variable is considered.
4. Compare the errors of an object and allocate it to the cluster where its error is the smallest.
5. Check if the algorithm reached the maximum number of steps or if clusters have not changed. If not, then repeat the steps 3-5.
6. If the maximum number of steps has been reached or clusters changed, then calculate the objective function.
7. Check if the number of initiations has not been exceeded. If not, then repeat the steps 1-7. If the number has been exceeded, the algorithm is completed.

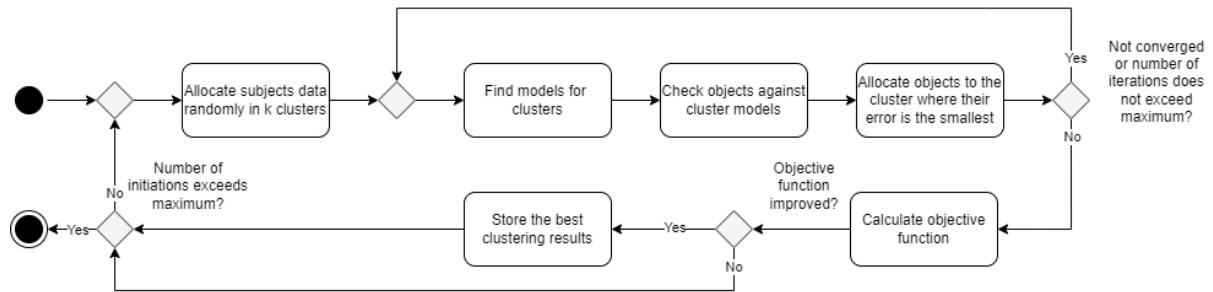


Figure 3.5 – A diagram representing the IKM algorithm (adopted from Plant et al. 2013).

3.4.6 Model Finding

Linear models involve solving multiple linear regression problems. One method for solving these problems is by using multiple least square regression. However, it is important to be aware of the issue of overfitting, which can occur when many dimensions are included in the model. This can lead to a model that is highly specific and not generalizable and may be difficult to interpret. To avoid overfitting, it is important to carefully select the set of explanatory variables for each model (Plant et al. 2013). For this selection, a greedy stepwise algorithm (Larose, 2015) is used. The Bayesian Information Criterion (BIC) is used as the evaluation criterion (George 2000). It is done by balancing the fit of the model to the data with the complexity of the model.

BIC is calculated using the next formula (Plant et al. 2013):

$$BIC = -2 \cdot LL(n^*, \sigma) + \log(n^*)(h\nu + 1),$$

- where $-2 \cdot LL(n^*, \sigma)$ term is goodness-of-fit and
 $LL(n^*, \sigma) = -\left(\frac{n^*}{2} \cdot \log(2\pi) + \frac{n^*}{2} + \frac{n^*}{2} \cdot \log(\sigma)\right)$ is a log-likelihood,
 - σ is expressed as $\sigma = \frac{E}{n^*}$,
 - $\log(n^*)(|v| + 1)$ is a penalty term that punishes overly complicated models.

This algorithm is used in step 2 of the IKM algorithm described in the previous section. The diagram of the greedy stepwise algorithm is presented in Figure 3.6.

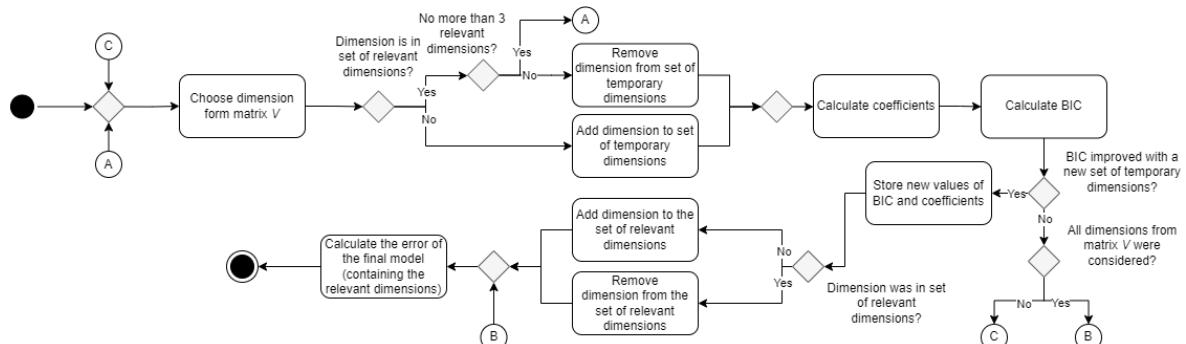


Figure 3.6 – A diagram representing the model finding algorithm (Adopted from Böhm et al. 2009).

The model-finding algorithm finds coefficients P for a dependent variable A . Before running the algorithm, the initial BIC value which is the big number enough (in this work the biggest float number for Python has been used) must be setup. Additionally, the relevant and temporary sets of dimensions must be created.

Algorithm from (Böhm et al. 2009) performs the next actions:

1. A dimension from matrix V is chosen.
2. If the dimension is in the set of the relevant dimensions and the set does not contain already 3 dimensions, then come back to step 1. If the dimension is in the relevant set of dimensions and the set contains already 3 dimensions, then remove the chosen dimension from the set of temporary dimensions. If the dimension is not in the set of relevant dimensions, then add this dimension to the set of the temporary dimensions.
3. Calculate coefficients for the dimensions that are in the set of the temporary dimensions.
4. Compute BIC is using the coefficients to evaluate the model.
5. Compare BIC to the previous value.
6. If BIC improved (the value is smaller than the previous one), then store it for the future comparison. Also, the coefficients for the current model are stored as the final coefficients. If BIC has not improved but not all dimensions from the matrix V were considered, then steps 1-5 are repeated until all possible combinations of dimensions were used. If all dimensions were used, then calculate the error for the model containing only the relevant dimensions.
7. After new values have been stored, check if the dimension was before in the set of the relevant dimensions. If yes, then remove the dimension from the set. If not, then add the dimension.
8. Lastly, calculate the error for the model with containing only the relevant dimensions.

As mentioned, the algorithm terminates when the BIC value was not improved. It means that decreasing variance by adding more dimensions did not improve the model or all possible dimensions were added and thus the variance was maximally decreased.

3.5 Dimensions Selection

To improve the results of the IKM, in particular, the clusters purity, the IKM was enhanced with the additional flow. The idea of this enhanced algorithm was to delete the dimensions the mean of which has the maximum or the minimum length to the mean of the whole dataset. Next are some definitions used in this algorithm:

- $\mu_d = \frac{\sum_{i=1}^n v_{dn}}{n}$, where μ_d – is the mean across one dimension, v_{dn} —is the value of a dimension (electrode) at time point n
- $\mu = \frac{\sum_{d=1}^n \mu_d}{d}$, where μ is the mean across all time points, n – is the total number of dimensions
- $D_d = \sqrt{(\mu - \mu_d)^2}$, where D_d is the distance between the mean across all time points and the mean across one dimension

The dimensions selection algorithm was implemented in two configurations:

- Removing the dimensions where their mean has the maximum distance to the mean across all time points one by one
- Removing the dimensions where their mean has the minimum distance to the mean across all time points one by one

The enhanced algorithm is presented in Figure 3.7.

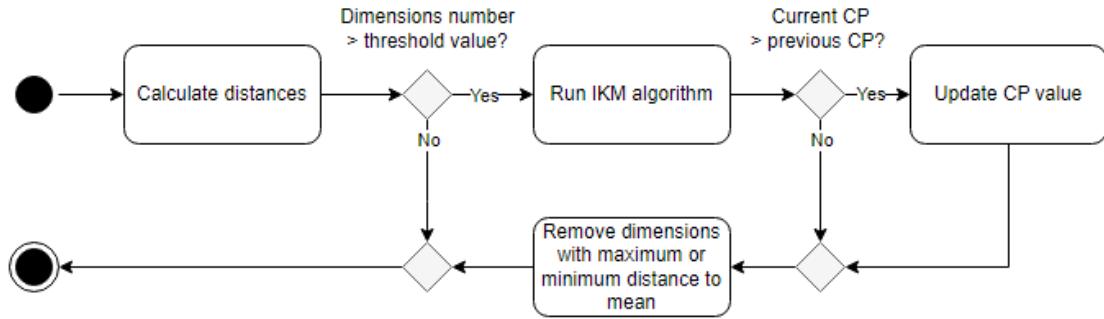


Figure 3.7 – A diagram representing the dimension selection algorithm.

Algorithm performs the next actions:

1. Calculate the distances D_d for each dimension d .
2. Check if the number of the dimensions higher than the threshold value. If yes, run the IKM algorithm as presented in Figure 2.5. Otherwise, terminate the algorithm.
3. Check if the clusters purity improved.
4. If yes, update the clusters purity value. Otherwise, remove dimensions from each object that have either maximum or minimum distance (depends on the configuration of the algorithm) to the mean across all time points.

The algorithm is terminated when the number of the dimensions is lower than some threshold value.

3.6 Interpretation of the Results

One benefit of using the IKM algorithm is the ability to interpret the interaction patterns that are detected. To make this easier, the focus shall be on a subset of the models that are most effective at distinguishing between different clusters. For each pair of clusters, a leave-one-out validation method is used to select the models that are most effective at discriminating between the clusters, using objects from the corresponding clusters (Plant et al. 2013). The diagram of the interpretation algorithm is presented in Figure 3.8.

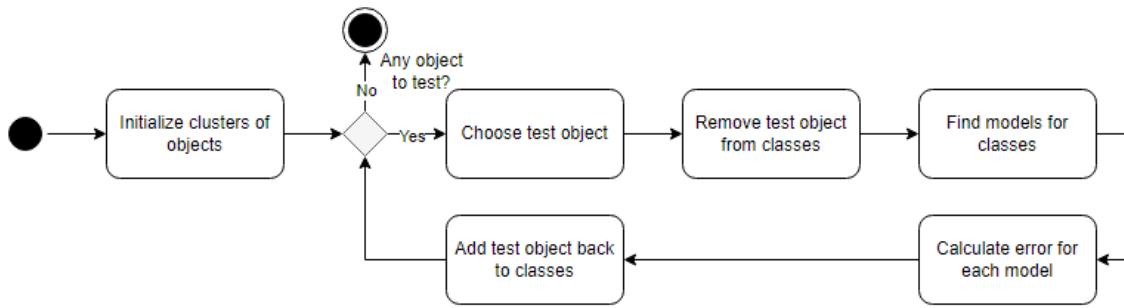


Figure 3.8 – A diagram representing the interpretation algorithm (Adopted from Plant et al. 2013).

Algorithm performs the next actions:

1. Add objects to their corresponding clusters.
2. Check if there are any objects that were not considered yet. If these objects are present, then continue the execution of the algorithm. Otherwise, terminate the algorithm.
3. Choose a test object from the clusters and remove it from them.
4. Find models for the clusters without a test object.
5. Calculate an error for each model and sum them up to the previous calculated errors for each model, if any. Before adding the error the previous value it is multiplicated by the sign value. This sign value can be either 1 or -1. If we consider the object with regard to the cluster to which it belongs, then the sign is 1. If we consider the object belonging to the opposite cluster, then the sign is -1.
6. Add test object back to the clusters.

The algorithm terminates when there are no objects left in the clusters that are needed to be tested. Then, the errors of the models are sorted ascendingly. The best separation models should have the lowest value. It means that the error of an object in the “appropriate” cluster has a low enough value and the error from the cluster to which the object does not belong is high enough. If the error value is positive it means that there is a “noise” in the cluster.

3.7 Criteria of the Quality of Clustering

There are various criteria to measure the quality of clustering. As in (Plant et al. 2013), in this work, we use to evaluate clustering effectiveness the next measures:

- Cluster purity (CP)
- Rand Index (RI)
- Information Criterion (IC)

3.7.1 Cluster Purity

Cluster purity (CP) is the ratio of the size of the class that has majority in a cluster with respect to the size of this class in a dataset (Xiaozhe Wang 2007). The CP is defined in the next way:

$$CP = \frac{Count_M}{Count_O} * 100\%$$

- where $Count_M$ is the number of elements of the dominant class in a considered cluster
- $Count_O$ is the overall number of elements of that class in a dataset.

The resulting metric is obtained by averaging it among all clusters. The goal of a clustering is to achieve as high value of CP as possible.

3.7.2 Rand Index

The Rand Index (RI) is a metric used to evaluate clustering results, which is determined by counting the number of pairs of objects that belong to the same class and cluster, as well as the number of pairs of objects that belong to different classes and different clusters (Halkidi, Batistakis, and Vazirgiannis 2001). The RI is defined as (Halkidi, Batistakis, and Vazirgiannis 2001):

$$RI = (SS + SN) / (SS + SN + DS + DD),$$

- where SS is the number of occurrences of points that belong to the same cluster and to the same class,
- SD is the number of occurrences of points that belong to the same cluster and to different classes,
- DS is the number of occurrences of points that belong to different clusters and to the same class,
- DD is the number of occurrences of points that belong to different clusters and to different classes

The goal of a clustering is to achieve as high RI as possible.

3.7.3 Information Criterion

An Information Criterion (IC) is a measure of entropy based on the empirical relationship between class labels and cluster labels. Conceptually, IC can be thought of as the number of bits required to encode the class labels of all objects, taking into account the information provided by the cluster labels (Dom 2013). IC is calculated using the next formula (Dom 2013):

$$IC = \tilde{H}(C|K) + \frac{1}{n} \sum_{k=1}^{|K|} \log \binom{h(k)+|C|-1}{|C|-1},$$

- where $\tilde{H}(C|K) = - \sum_{c=1}^{|C|} \sum_{k=1}^{|K|} \frac{h(c,k)}{n} \log \left(\frac{h(c,k)}{h(k)} \right)$ is a conditional empirical entropy,
- C is class labels,
- K is cluster labels ,
- $h(c, k)$ is the number of objects labelled class c that are assigned to cluster k , $h(c) \equiv \sum_k h(c, k)$ and $h(k) \equiv \sum_c h(c, k)$.

The goal of a clustering is to achieve as low IC as possible.

4 Data Sets and Methods

4.1 Data Characteristics

This section explains the datasets used when working on this thesis.

4.1.1 Data of Alcoholic and Non-alcoholic Patients

As reference data, the EEG recordings of alcoholic and non-alcoholic patients were used (EEG Database. 1999. UCI Machine Learning Repository.). The subjects in the study were exposed to either one or two visual stimuli, with the second stimulus being either identical or different from the first. The dataset consists of data from two subjects - one alcoholic and one non-alcoholic – and includes 10 runs of EEG recordings for each of the three experimental conditions. Additionally, the data is collected from 64 electrodes.

Moreover, this dataset was used to repeat the identical experiment as described by (Plant et al. 2013) the aim of which was to cluster the EEG runs into groups based on whether the subject was alcoholic or non-alcoholic. The clustering is into two groups, as similarly required for the data basis of depressed patients, however the patients were presented with either a single stimulus or two stimuli, which consisted of pictures of objects, which is not the case for the data base of depressed patients.

4.1.2 Data of Depressed Patients

The main data was collected in the scope of the research by (Bučková et al. 2020). After eliminating data from subjects with distorted and not readable EEG recordings, the resulting dataset consisted of the data of 134 subjects. 93 of the subjects are women, and 41 are men with a mean age of 46 years. Every participant was recorded before and after the treatment which lasted for 4 weeks.

19 standard electrode positions were used for the recordings, namely, Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fz, Cz, and Pz. Patients were recorded for 10 minutes, laying on a bed with elevated upper body at 30-45 degrees (semirecumbent position). The recording was done in a room with dimmed light. The patients had their eyes closed but their alert state was controlled by using acoustic stimuli when signs of drowsiness appeared in EEG.

The data were downsampled to 250 Hz and the initial and last 30 seconds were removed. More details on the specific raw EEG pre-processing steps can be found in (Bučková et al. 2020).

The data was provided in different formats. The main dataset (consisting of EEG recordings) was provided in the form of a MATLAB file and had the structure presented in Table 4.1.

Table 4.1 – Structure of the depressed patients' dataset. Only first three rows are presented.

ID	Age	Sex	Visit	Response	Sampling rate	EEG data
"P001"	49	0	1	1	250	Structure data type of EEG data
"P001"	49	0	2	1	250	Structure data type of EEG data
"P002"	36	0	1	0	250	Structure data type of EEG data

The supplementary dataset to the presented dataset contains all information as in the MATLAB file except for EEG data. This dataset is in Excel format. Its value is that it contains information about the results of MADRS questionnaires.

4.2 Data Exploration

Data exploration is an important step in data analysis because it allows understanding the properties and characteristics of a dataset. This section outlines some properties and relationships in the data.

4.2.1 3-dimensional Trajectories of the EGG Data of Depressed Patients

Figure 4.1 visualizes EEG data of the first visit of subject 1 who responded to the treatment.

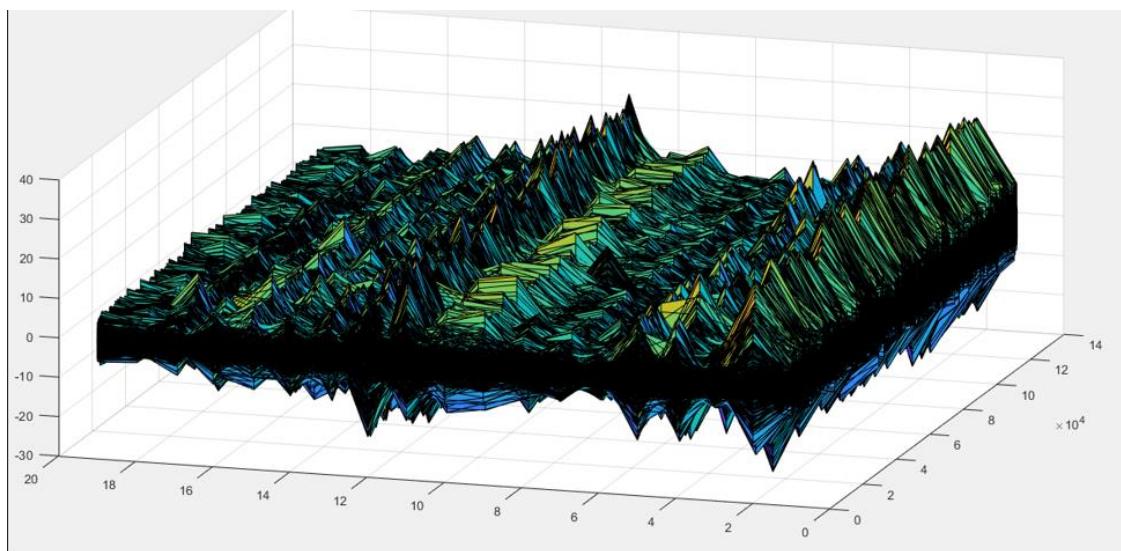


Figure 4.1 – EEG data of the first visit of the subject 1 who responded to the treatment.

Figure 4.2 visualizes EEG data of the second visit of the same subject.

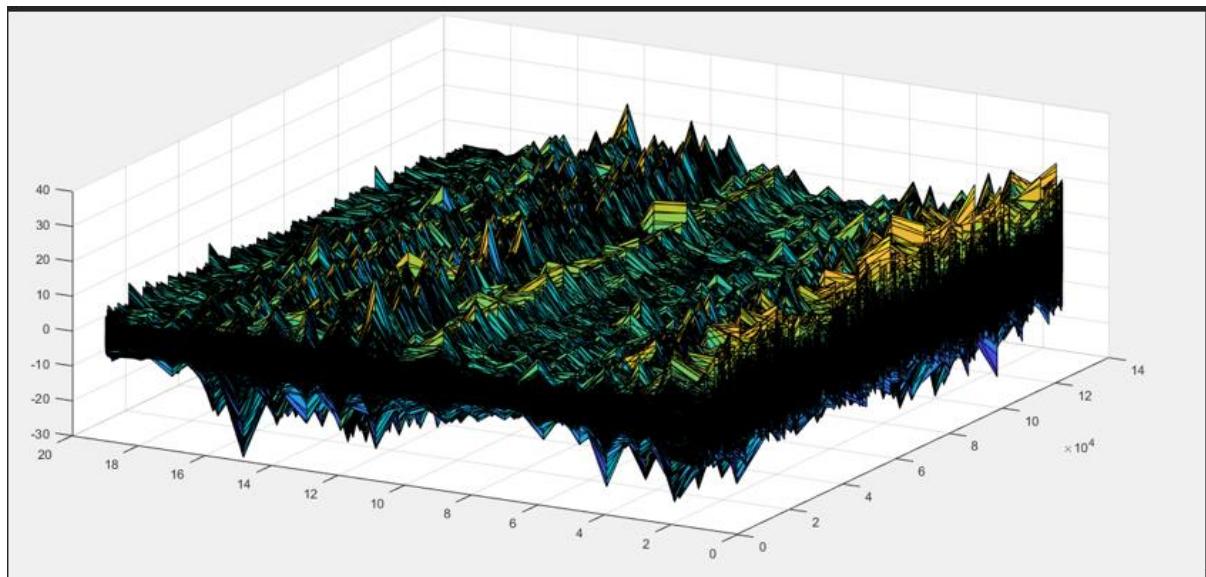


Figure 4.2 – EEG data of the second visit of the subject 1 who responded to the treatment.

In these two figures, the x-axis represents time points, the y-axis shows the number of dimensions, and the z-axis represents the values of electrodes. From the figures, it can be seen that the all electrodes of this patient have very similar trajectories. It might be because the data was collected during the rested state.

4.2.2 Comparison of Topomaps of Depression and Alcoholic EEG Data

In this subsection, we compared the representation between the EEG data of alcoholics-non-alcoholics and depressed patients. For this, we used topomaps.

Figures 4.3 and 4.4 show the snapshots from the topomap of the EEG data in different periods. This EEG data is from alcoholic and non-alcoholic people.

The difference between these two visualizations is quite noticeable. At the same time points, the activation of the subjects' brains happens differently, probably due to the presence of the stimuli.

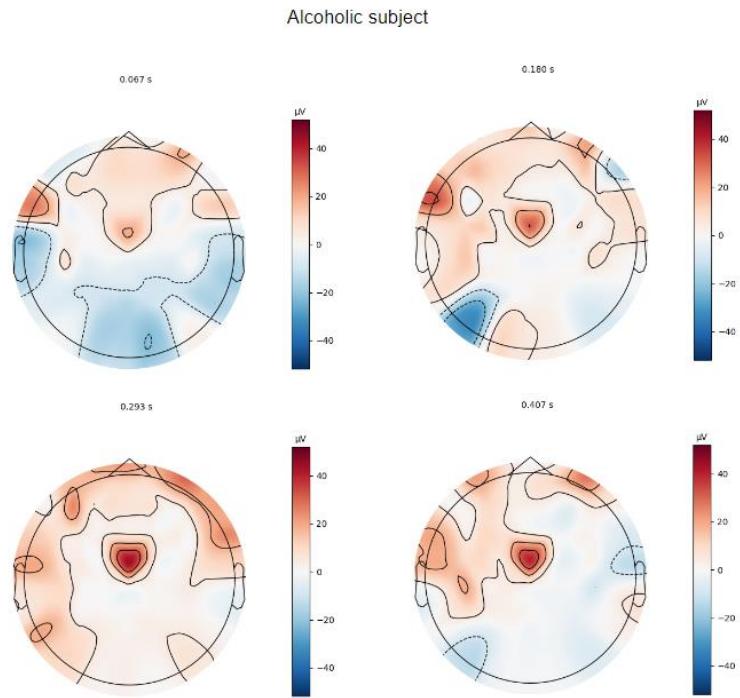


Figure 4.3 –Snapshots of a topomap of an alcoholic subject.

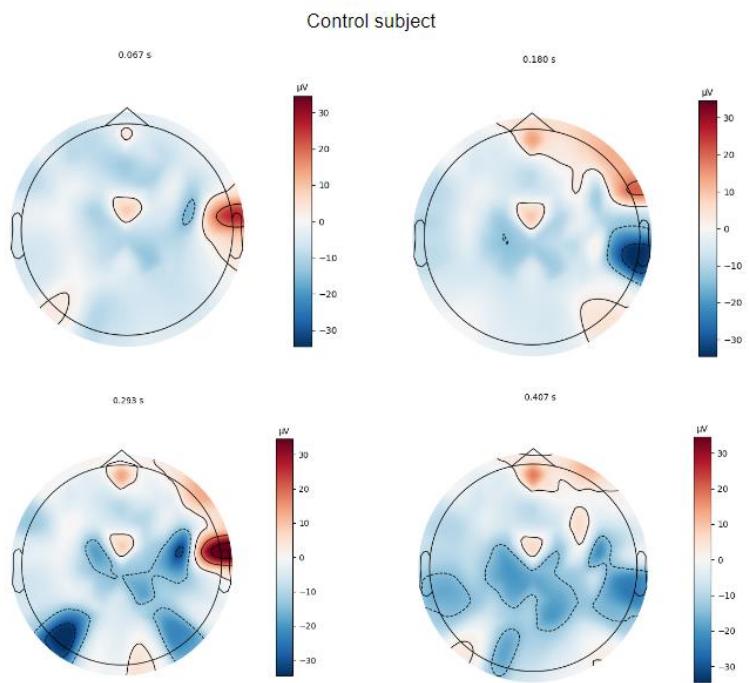


Figure 4.4 –Snapshots of a topomap of an alcoholic subject.

Figures 4.5 and 4.6 show the snapshots from the topomap of the EEG data in different periods of a depressed subject. These two figures represent the first and the second visit correspondingly. The subject is 25 years old, male, and responded to the treatment.

Here the difference between different snapshots at the same time is not very big compared to the example of alcoholic and non-alcoholic subjects.

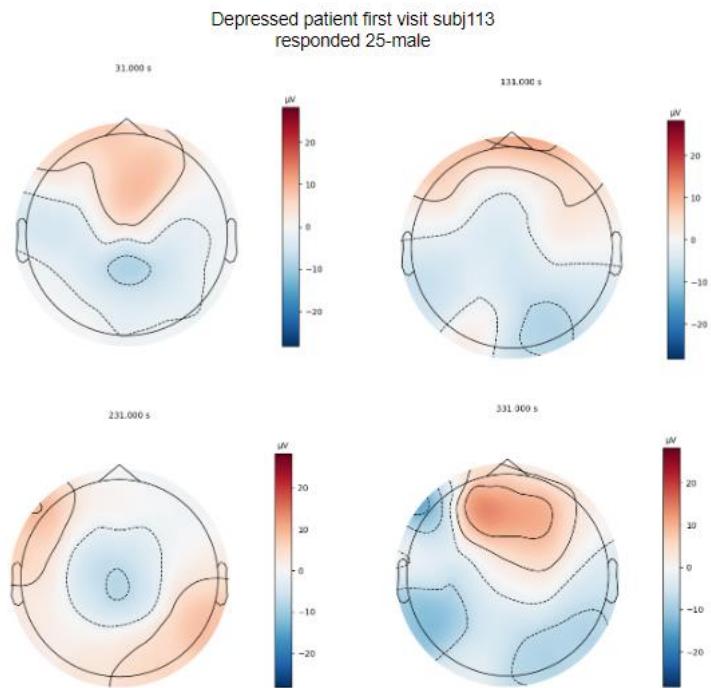


Figure 4.5 –Snapshots of a topomap of the first visit of a depressed subject.

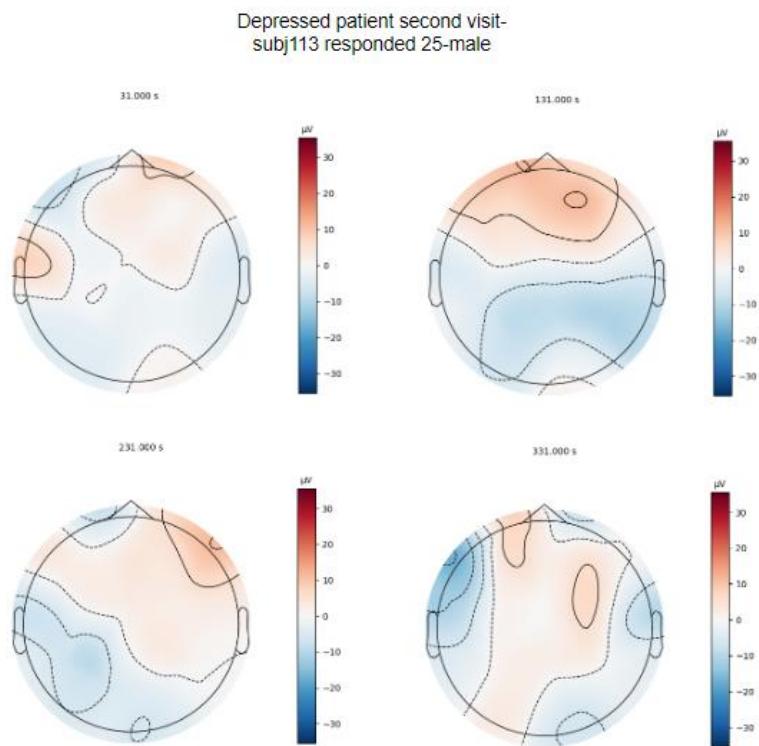


Figure 4.6 –Snapshots of a topomap of the second visit of a depressed subject.

It is noticeable that the depressed patient's topomaps exhibit greater similarity between their two visits when compared to the topomaps of both alcoholics and non-alcoholics, which display more noticeable differences. Additionally, the EEGs of alcoholics and non-alcoholics show more edge values than the EEG of a depressed patient.

4.2.3 Coefficients Correlation Matrices

Figure 4.7, relating to the depressed patients data, represents the correlation between cluster coefficients where clusters are ideal. The ideal clusters here are the two clusters that contain only responders and non-responders. The coefficients, therefore, were found for each cluster and Pearson's correlation was applied to them. The coefficients were found with the formula provided in Section “Interaction k-means”. The matrix in figure 4.7, shows the correlation matrix between the first visit data of non-responders and the second visit data of responders. The coefficients of the 10th to 15th dimensions have a stronger positive correlation with other dimensions.

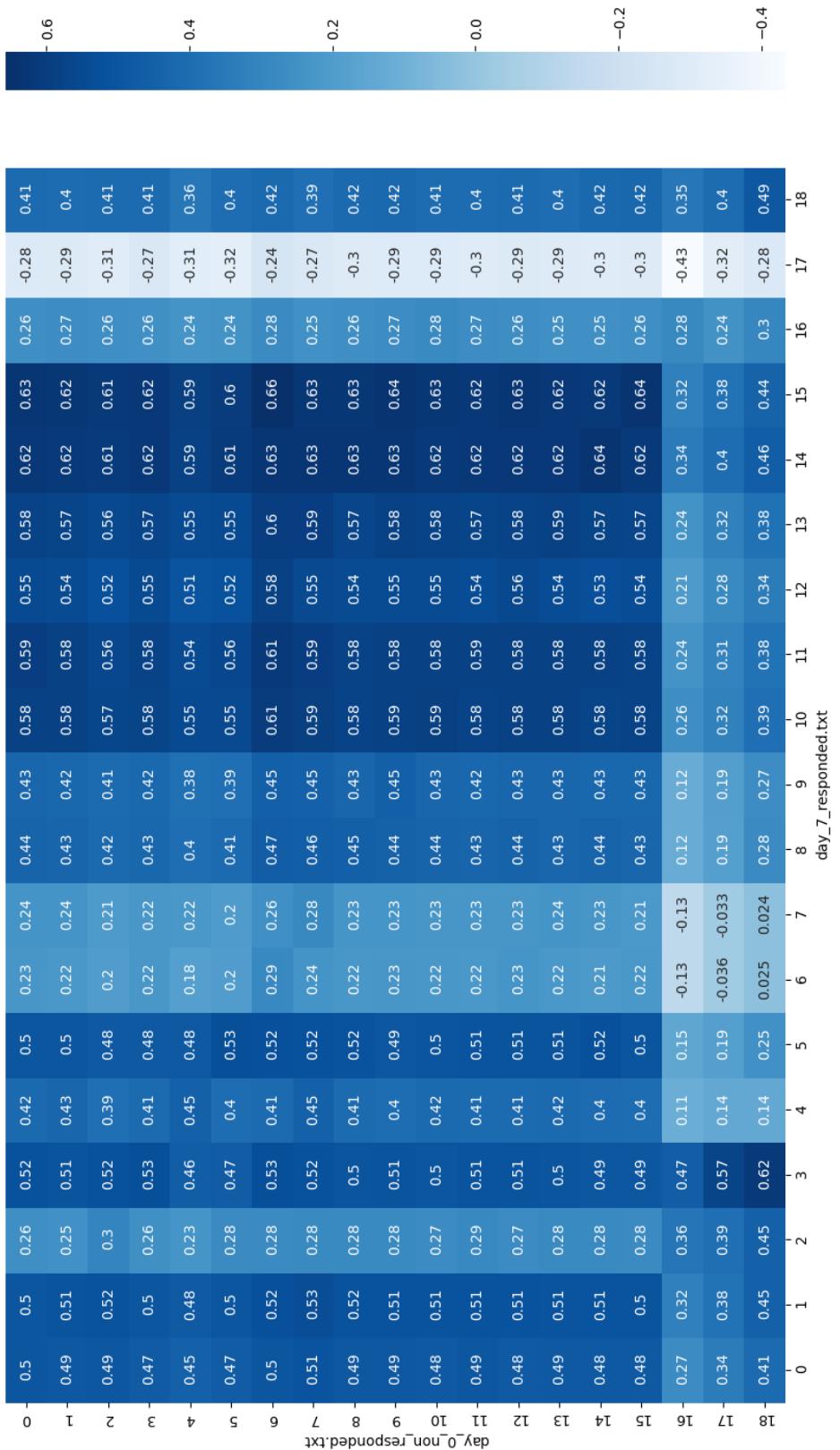


Figure 4.7 – Correlation matrix for the cluster coefficients of ideal clusters. The coefficients are calculated for the first visit non-responders and the second visit responders.

In Figure 4.8, on the other hand, slightly different behavior in the correlation can be observed. The correlation matrix was computed for the coefficients of the data of the first-day responders and the seventh-day responders. Here a very strong positive correlation between the 0th to 15th coefficients of responders can be observed from different visits as well as between coefficients of 16th to 18th.

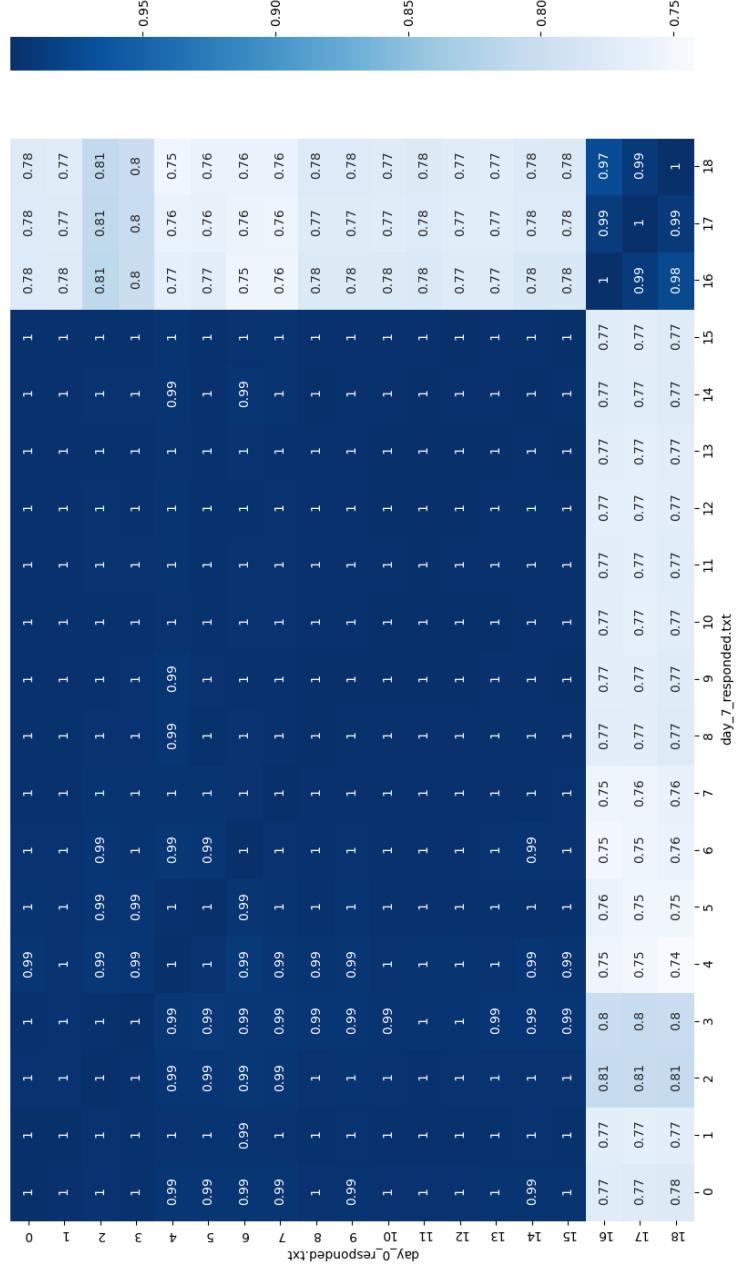


Figure 4.8 – Correlation matrix for the cluster coefficients of ideal clusters. The coefficients are calculated for the first visit responders and the second visit responders.

Figure 4.9 represents the correlation matrix for each object's coefficients. The objects were preallocated in the ideal clusters and only then the coefficients were calculated (the similar approach as for the cluster coefficients). The only difference in the calculation approach from

the previous figures is that the coefficients in this situation are calculated for each object separately and not for the whole cluster.

Overall, almost in all cases, the correlation between the coefficients of some objects can be observed in the middle of the set (as for object 34, for example). These middle-set coefficients of objects are strongly positively correlated with the coefficients of the other subjects. There is mostly no strong negative correlation.

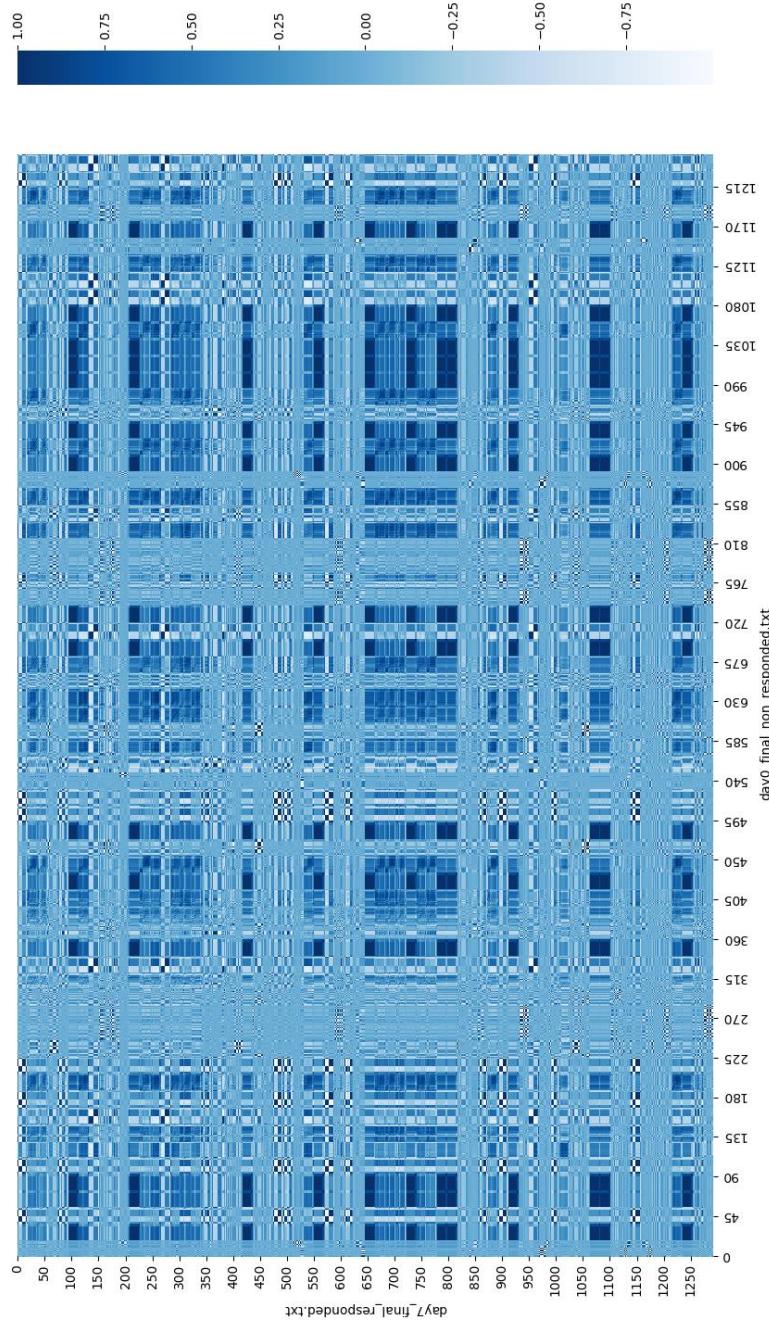


Figure 4.9 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the first visit non-responders and the second visit responders.

This section contains only some correlation matrices of ideal clusters. More correlation matrices can be found in Appendix A.

4.2.4 Distances Between Coefficients

Table 4.2 represents the Frobenius distance between coefficients of ideal clusters. The coefficients were obtained using the same method as in the previous examples of the coefficients correlation. The coefficients between non-responders and responders subjects have the largest difference. The difference between the responders of the first visit and the responders of the second visit is low. The same behavior is for the coefficients of non-responders from different visits.

Table 4.2 – Frobenius distance between coefficients of ideal clusters.

	1st-visit-non- responders	2nd-visit-non- responders	1st-visit- responders	2nd-visit- responders
<i>1st-visit-non- responders</i>	0	0.531339	6.423986	6.487957
<i>2nd-visit-non- responders</i>	0.531339	0	6.616068	6.67983
<i>1st-visit- responders</i>	6.423986	6.616068	0	0.162086
<i>2nd-visit- responders</i>	6.487957	6.67983	0.162086	0

Table 4.3 represents the Frobenius distance between the coefficients of each object allocated to ideal clusters. The results are slightly different from the results of the previous section. The lowest distance is still between the subjects of the first visit who responded and the subjects of the second visit who responded. However, the second lowest distance is between the subjects of the first visit who did not respond and the subjects of the first visit who responded. The biggest distance is between the first visit non-responders and the second visit non-responders.

Table 4.3 – Frobenius distance between the coefficients of each object allocated to ideal clusters.

	1st-visit-non-responders	1st-visit-responders	2nd-visit-non-responders	2nd-visit-responders
<i>1st-visit-non-responders</i>	0	1898.983	2275.569	1963.585
<i>1st-visit-responders</i>	1898.983	0	2132.343	1775.168
<i>2nd-visit-non-responders</i>	2275.569	2132.343	0	2107.632
<i>2nd-visit-responders</i>	1963.585	1775.168	2107.632	0

4.3 General Data Preprocessing

As was described in Section “Materials and Methods. Data Characteristics”, the data was provided in a MATLAB file. To make it acceptable by the algorithm, the data was transformed into text files, each corresponding to one subject. Columns in this text file represented the electrodes, rows – the time points. The name of each file encoded the information about the number of a dataset and the patient’s response to the treatment. To transform the data, the script written in MATLAB was used.

For the runs of the algorithm, the datasets were concatenated. Concatenating the two datasets (i.e., combining them into a single dataset) provides a more complete picture of the changes in the patient’s brain activity before and after treatment. It allows us to directly compare the EEG dynamics of the patient before and after treatment and identify any changes that may have occurred as a result of the treatment.

The data of the first and second visits of each subject was simply concatenated one after another and an algorithm has been run on it.

4.4 Used Tools

The authors of paper (Plant et al. 2013) provided the source code of the algorithm written in Java programming language. It was recoded in Python programming language as using an algorithm in this language allowed to use many supporting and useful libraries for data preprocessing and analysis. These are just few of them:

- Numpy (<https://numpy.org/>) for working with matrices
- Math (<https://docs.python.org/3/library/math.html>) for calculations
- Pandas (<https://pandas.pydata.org/>) for working with the Python data type dataframes
- SciPy (<https://scipy.org/>) for using the optimized algorithms or mathematical operations such as Hilbert transform.
- MNE (<https://mne.tools/stable/index.html>) for creating animated topographic maps.

5 Preprocessing Methods

In this section we will present the setting of clustering parameter k and all preprocessing methods we have applied to the data sets.

5.1 Alcoholic and Control EEG Data

As was mentioned before, the IKM algorithm was used for two datasets. The alcoholics-non-alcoholics dataset was used as a benchmark to repeat the initial experiment in the original paper describing the algorithm (Plant et al. 2013). After the run, the results were repeated except for the Information Criterion metric which was slightly different from the original one. They are represented in Table 5.1. One can see that the cluster purity is 100% for this data set.

Table 5.1 – Results of clustering on the alcoholic-non-alcoholic EEG dataset with 60 objects and 64 electrodes (dimensions).

Setting	Cluster Purity	Random Index	Information Criterion
Alcohol-control EEG data	100%	1	0.05

5.2 EEG Data of Depressed Patients

All following results relate only to the EEG dataset of the depressed patients with 134 subjects and 19 electrodes.

5.2.1 Determination of Optimal Number of Clusters

To check if the optimal number of clusters corresponds to the number of classes present in the data (patients who responded to the treatment and those who did not) the elbow method (Cui 2020) was used. The elbow method in cluster analysis is a way to determine the ideal number of clusters in a data set. It involves creating a plot that shows the explained variation versus the number of clusters and selecting the point at which the plot forms an "elbow" shape as the optimal number of clusters to use.

Figure 5.1 shows the error value for different numbers of clusters. One can see from it that after two clusters the error does not change very significantly. However, it still changes noticeably when the number of clusters is three. Despite that, it would be hard to calculate the metrics for three clusters as there is no clear identification of what those metrics should be based on. For two clusters, this indication is as mentioned before the patients who responded well and who did not respond to the treatment.

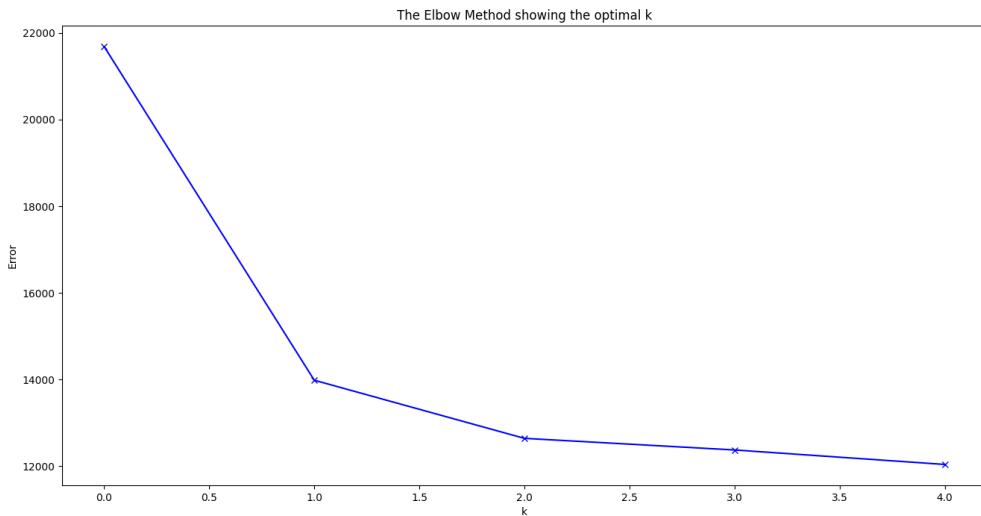


Figure 5.1 – The plot of the clusters error value with regard to a number of clusters.

Thus, in all subsequent experiments, k was chosen to be equal to two.

5.2.2 Untransformed Data

Table 5.2 shows the quality measures for results of the clustering on the untransformed data. This table will serve us for comparison whether the preprocessing and transformation methods applied to the EEG data improved the quality measures of the clustering.

Table 5.2 – Results of clustering on the not-preprocessed EEG data of depressed patients.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	50%	0.5	0.33
Combined	50%	0.5	0.33

One can see that untransformed data provided the results that could be obtained from separation data objects in two clusters by chance.

5.2.3 Sinus Non-linear Function

EEG data may have some characteristics that are related to the sinusoidal function, as both deal with oscillating patterns. The sinusoidal function is often used to model periodic phenomena, such as the alternating current of an electrical signal, and EEG data is characterized by various types of oscillations that occur in the brain. However, it is important to note that the relationship between sinusoidal functions and EEG data is not a simple one, and further research would be needed to fully explore any potential connections between the two.

We hypothesized that applying sinus function to the EEG data may increase the quality measures of clustering.

To use any non-linear function with the IKM algorithm new dimensions of the data must be created. The number of new dimensions must be identical to the number of initial dimensions. The non-linear function must be applied to each data point in the new dimensions. For the data of depressed patients, 19 new dimensions (as there were 19 initial dimensions) with applied sinus function were created.

Table 5.3 shows the results of the clustering of data with the applied sinus function

Table 5.3 – Results of clustering on the data with applied sinus function.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	50%	0.5	0.33

One can see that sinus transformation had only a small influence on the increase of CP and no influence on other metrics.

5.2.4 Downsampling

We hypothesized that downsampling of the EEG data can improve the results of the algorithm because the data of depressed people is more complex than of the healthy people (Torre Luque and Bornas 2017).

The data was downsampled in different ways:

- One-thirds of the dataset were left: the initial, middle, and last parts
- Every n^{th} element was removed from the data
- Every n^{th} element was left in the data

Tables 5.4-5.10 show the results of clustering on the data with the downsampling applied.

Table 5.4 – Results of clustering on the initial one-third of the data.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	50%	0.5	0.33

Table 5.5 – Results of clustering on the middle one-third of the data.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	50%	0.5	0.33

Table 5.6 – Results of clustering on the last one-third of the data.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	51.6%	0.5	0.33

Table 5.7 – Results of clustering on the data with every 5th element removed.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	50%	0.5	0.33

Table 5.8 – Results of clustering on the data with every 10th element removed.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	50%	0.5	0.33

Table 5.9 – Results of clustering on the data with every 5th element left.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	50%	0.5	0.33

Table 5.10 – Results of clustering on the data with every 10th element left.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	50%	0.5	0.33

One can see that downsampling showed almost the same results as for sinus transformation and it again had only a small influence on the increase of CP.

5.2.5 Dimensions selection

We hypothesized selection of only some channels of the EEG data can improve the results of the IKM algorithm. To test this, we calculated the distance between the mean across all time points and the mean across one dimension and got the results for each dimension. Then depending on the configuration, we either removed a dimension with the minimum or the maximum result.

Results from applying dimensions selection modification of the IKM algorithm (described in Section “Interaction k-means. Dimensions Selection”) are represented in the Table 5.11. For different configurations either 16 or 17 dimensions were selected out of 19.

Table 5.11 – Results from applying dimensions selection modification of the IKM algorithm.

Configuration	Setting	Cluster Purity	Random Index	Information Criterion
Maximum distance	First visit	50.8%	0.5	0.33
Maximum distance	Second visit	50.8%	0.5	0.33
Maximum distance	Combined	51.6%	0.5	0.33
Minimum distance	First visit	50.8%	0.5	0.33
Minimum distance	Second visit	51.6%	0.5	0.33
Minimum distance	Combined	50%	0.5	0.33

One can see that these dimension selection methods had only a small influence on the cluster purity. The best settings are using the maximum distance approach on the combined data and the minimum distance on the second visit data. These strategies provided 51.6% cluster purity.

5.2.6 Box-Cox Transformation

Box-Cox transformation is a type of power transformation that is used to create a monotonic transformation of data. A monotonic transformation is a mathematical function that preserves the relative order of the elements in a dataset.

Box-Cox transformation can be used to stabilize the variance of a dataset and make it more normally distributed. Thus, we hypothesized that the stabilization of the variance by applying Box-Cox can improve the validity of the IKM algorithm applied to the data.

The one-parameter Box–Cox transformation is defined as (Box and Cox 1964):

$$y_i^\lambda = \begin{cases} \frac{y_i^\lambda - 1}{\lambda} & \text{if } \lambda \neq 0, \\ \ln y_i & \text{if } \lambda = 0 \end{cases}$$

where λ is the power parameter.

Before applying the Box-Cox transformation, firstly the data was transformed to have only positive numbers. To do it, the shift value must be calculated. The formula was used is the next:

$$\text{Shift value} = |\text{Minimum number in the dataset}| + 1$$

Then this number was added to each number in the dataset.

The Table 5.12 shows the results of clustering with applied Box-Cox transformation

Table 5.12 – Results of clustering on the data with applied Box-Cox transformation.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	51.6%	0.50	0.33
Second visit	50.8%	0.50	0.33
Combined	50.8%	0.50	0.31

Again, one can see that Box-Cox transformation had only a small influence on the CP of the first visit and almost no influence on the second and combined visits.

5.2.7 Discrete Wavelet Transformation

A discrete wavelet transform (DWT) is a wavelet transform in which the wavelets are discretely sampled. A wavelet is a type of wave that oscillates and has a varying amplitude that starts at zero, increases or decreases, and then returns to zero one or more times.

DWT is used to analyze signals and functions in order to extract both frequency and location information. Unlike Fourier transforms, which only provide frequency information, DWTs are able to provide both frequency and temporal resolution, making them useful for analyzing signals that vary over time.

In this thesis, the Daubechies 4 (db4) wavelet function was used to decompose EEG signals into five levels. The wavelet waveform is presented in Figure 5.2. The db4 wavelet function was chosen for its good time-frequency localization properties, which are important for accurately analyzing signals that vary over time (Murugappan, Ramachandran, and Sazali 2010).

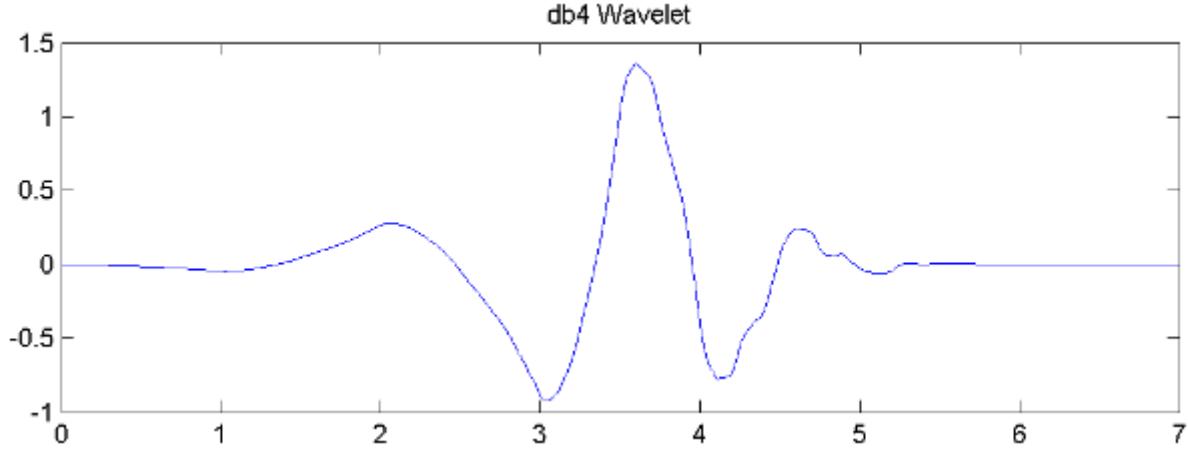


Figure 5.2 -- The Wavelet Daubechies 4 (db4) waveform (Illustration from Belkhou, Jbari, and Belarbi 2017).

Table 5.13 shows the results of clustering with applied Discrete Wavelet transformation.

Table 5.13 – Results of clustering on the data with applied Discrete Wavelet transformation.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	51.6%	0.5	0.33
Second visit	52.3%	0.5	0.33
Combined	51.6%	0.5	0.33

One can see that db4 transformation increased the CP more significantly than the previous transformations for all visits, in particular for the second visit.

5.2.8 z-normalization

EEG data is known for its high complexity, high dimensionality, non-stationary oscillations, and low signal-to-noise ratio. When collecting multi-trial EEG data, it is difficult to ensure that all the trials have the same phase or that the EEGs are recorded for the same length of time. This is due to the unpredictability and uncontrolled nature of brain activity during EEG recording sessions. Additionally, multi-trial EEG data may be incomplete, as some trials may be shorter or longer than others, or may be lost altogether. Finally, EEG data is often contaminated by various types of noise and artifacts, such as those resulting from environmental factors or unspecific brain activity, like eye blinks. These characteristics make it challenging to analyze and interpret multi-trial EEG data (Dai et al. 2018).

To address this issue and make the EEG data more amenable to clustering, it is common to apply z-normalization (Goldin and Kanellakis 1995) as a preprocessing step. This helps to make the data shift- and complexity-invariant, improving the performance of distance-based clustering methods. It helps to eliminate distortions in the data and reduce the impact of outliers on the similarities between the EEG trials (Dai et al. 2018).

The z-normalization is defined as following:

$$e' = \frac{e - \mu(e)}{\sigma(e)},$$

Where $\mu(e)$, $\sigma(e)$ denote the mean and standard deviation of the data e , respectively. The data e is usually normalized such that $\mu(e) = 0$ and $\sigma(e) = 1$.

Table 5.14 shows the results of clustering with applied z-normalization.

Table 5.14 – Results of clustering on the data with applied z-normalization.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	55.1%	0.5	0.33
Second visit	52.2%	0.5	0.33
Combined	51.4%	0.5	0.33

One can see that z-normalization increased CP value more significantly than for all previous methods for the first visit, for the second visit and gave the CP comparable to db4 transformation for combined data.

5.2.9 z-score

The z-score is a measure of how far a raw score (an observed value or data point) is from the mean in terms of the number of standard deviations. A z-score is calculated by the next formula (Kreyszig 1979):

$$z = \frac{x - \mu}{\sigma},$$

where:

- μ is the mean of the population,
- σ is the standard deviation of the population,
- x is the raw score.

The raw scores below the mean are negative, above – positive. We hypothesized that this kind of normalization can better separate abnormal EEG activity.

Table 5.15 shows the results of clustering with applied z-score measure.

Table 5.15 – Results of clustering on the data with applied z-score measure.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	50%	0.5	0.33

The z-score transformation did not improve the results of the clustering as much as the z-normalization. The only slight improvement with regard to the random allocation of data objects was for the second visit data.

5.2.10 Z-transform

The Z-transform converts a sequence of real or complex numbers in the time domain, into a complex frequency-domain representation on the z-plane (Mandal 2020).

The formula of Z-transform is the next (Sullivan 1996):

$$Z[x(t)] = X(z) = \sum_{n=-\infty}^{\infty} x(n)z^{-n}$$

This equation is commonly known as the bilateral transform because it is defined for both positive and negative values of n.

The value of $1j = 1 \cdot \sqrt{-1}$ is a complex number represented as the imaginary unit in the complex number system. It is used as the variable z in the Z-transform to represent a rotation of 90 degrees around the origin in the complex plane. This makes it a suitable choice as it allows Z-transform to capture the full range of frequencies in the signal.

For this type of preprocessing, the magnitude which represents the amplitude of the signal at each frequency, and the phase which represents the phase shift at each frequency were calculated. For the magnitude the next formula was used:

$$\text{Magnitude} = |Z(x(t))|$$

For the phase, the counter-clockwise angle from the positive real axis on the complex plane in the interval $(-\pi, \pi]$ was calculated.

Table 5.16 shows the results of clustering with extracted magnitude applying Z-transform while Table 5.17 shows the results of clustering with extracted phase applying Z-transform.

Table 5.16 – Results of clustering on the data with extracted magnitude applying Z-transform.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	53.3%	0.5	0.33
Second visit	50.6%	0.5	0.33
Combined	55.4%	0.5	0.33

Table .517 – Results of clustering on the data with extracted phase applying Z-transform.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	51.4%	0.5	0.33
Second visit	51.7%	0.5	0.33
Combined	50%	0.5	0.33

One can see that clustering on the extracted magnitude applying Z-transform showed improved cluster purity for the combined and the first visit. For the extracted phase applying Z-transform, the results did not show any improvement in CP.

5.2.11 Exponential Smoothing

As it was already mentioned, EEG data tends to have non-stationary oscillations and high complexity. We decided to apply exponential smoothing to the EEG data to filter out noise irregular variations in the signal. In other words, the goal of this transformation is to make a smooth function out of the EEG signal.

Exponential smoothing is a technique used to forecast the future values of a univariate time series based on past data. Exponential smoothing methods are used to create forecasts by taking a weighted average of past observations, where the weight of each observation decreases exponentially as it becomes older. This means that more recent observations are given more weight in the average (Hyndman, 2021).

The simplest form for the simple exponential smoothing with the sequence that begins at $t = 0$ is defined as (Guthrie 2020):

$$s_0 = x_0$$

$$s_t = \alpha x_t + (1 - \alpha)s_{t-1}, t > 0,$$

where:

- α is the smoothing factor and $0 < \alpha < 1$
- the output of the exponential smoothing algorithm is commonly written as s_t .

Double exponential smoothing is a method that involves applying an exponential smoothing filter recursively twice. It is called "double" because it involves two applications of the exponential filter.

The purpose of double exponential smoothing is to account for trends in a series of data by introducing a term that is updated using exponential smoothing. This helps to improve the accuracy of forecasts by considering the possibility that the data may not be stationary, but rather may have some form of trend.

Double exponential smoothing has the next formulas (Guthrie 2020):

$$s_0 = x_0$$

$$b_0 = x_1 - x_0$$

And for $t > 0$ by

$$s_t = \alpha x_t + (1 - \alpha)(s_{t-1} + b_{t-1}),$$

$$b_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1},$$

where:

- $\alpha(0 \leq \alpha \leq 1)$ is the data smoothing factor
- $\beta(0 \leq \beta \leq 1)$ is the trend smoothing factor
- b_t is the best estimate of trend at time t

Triple exponential smoothing uses exponential smoothing three times. This type of smoothing involves calculating a trend line for a series of data, as well as seasonal indices that are used to adjust the values in the trend line based on the position of each time point within a repeating cycle of length L .

For the exponential smoothing applied to the data, only the additive seasonality was considered.

The next formulas are used in the triple exponential smoothing with additive seasonality (Guthrie 2020):

$$s_0 = x_0$$

$$s_t = \alpha(x_t - c_{t-L}) + (1 - \alpha)(s_{t-1} + b_{t-1})$$

$$b_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1}$$

$$c_t = \gamma(x_t - s_{t-1} - b_{t-1}) + (1 - \gamma)c_{t-L},$$

where:

- $\alpha(0 \leq \alpha \leq 1)$ is the data smoothing factor,
- $\beta(0 \leq \beta \leq 1)$ is the trend smoothing factor,
- $\gamma(0 \leq \gamma \leq 1)$ is the seasonal change smoothing factor
- L is the length of the cycle of seasonal change.

The formula for the initial trend estimate b_0 is the next:

$$b_0 = \frac{1}{L} \left(\frac{x_{L+1} - x_1}{L} + \frac{x_{L+2} - x_2}{L} + \cdots + \frac{x_{L+L} - x_L}{L} \right)$$

A low value for alpha results in a line that changes smoothly, similar to a moving average with a large number of periods. On the other hand, a high value for alpha gives more weight to recent data, resulting in a line that more closely tracks the data. Taking in the consideration that the data consists of many data points, the $\alpha = 0.05$ was taken.

Types of trend and seasonal components for double and triple exponential smoothing were chosen as additive. Additionally, for the triple exponential smoothing 12 seasonal periods were chosen.

Tables 5.18-5.20 show the results of clustering with different applied exponential smoothing methods.

Table 5.18 – Results of clustering on the data that was simple exponentially smoothed.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50.8%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	50%	0.5	0.33

Table 5.19 – Results of clustering on the data that was double exponentially smoothed.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	51.6%	0.5	0.33
Second visit	51.6%	0.5	0.33
Combined	50%	0.5	0.33

Table 5.20 – Results of clustering on the data that was triple exponentially smoothed.

Setting	Cluster Purity	Random Index	Information Criterion
First visit	50%	0.5	0.33
Second visit	50%	0.5	0.33
Combined	50%	0.5	0.33

All types of exponential smoothing did not bring any improvements for the clustering results.

5.2.12 EEG Bands Extraction

As was already mentioned in the section “Electroencephalography (EEG)”, different brain activities are associated with different frequency ranges, which are generally classified as delta, theta, alpha, beta, and gamma. To extract these bands, the Butterworth filter was applied.

The Butterworth filter is a type of filter used in signal processing that is designed to have a frequency response that is as flat as possible over a specified range of frequencies, called the passband (Acharya et al. 2016).

The square function of the Butterworth filter defines it. The equation of it with applied bilinear transformation (Roberts and Roberts 1978) is

$$|H_B(j\omega)|^2 = \left[1 + \frac{\tan\left(\frac{\omega T}{2}\right)}{\tan\left(\frac{\omega_0 T}{2}\right)} \right]^{-2n},$$

where:

- T is the sampling interval (Stearns and Hush 1990)
- ω is the cutoff frequency
- $j = (-1)^{\frac{1}{2}}$
- n is the number of order of the filter.

As suggested by (Jenkins and Nayeri 1986), a decomposed filter structure consisting of parallel or cascaded second-order sections is a superior one. A cascade of second-order sections is a set of cascaded second-order filters that can be combined to form a filter with any desired order. Thus, this type of structure was used when extracting EEG bands.

Each second-order section consists of a single biquadratic section, which is a type of two-pole, a two-zero filter that can be implemented using a simple difference equation. The number of biquadratic sections was chosen to be 5, thus the resulting order of the final filter is $2*5 = 10$. Typically, a fifth-order Butterworth filter is used for EEG bands extraction because it provides a good balance between attenuation in the stopband and preserving the signal in the passband. This order provides a steep roll-off while avoiding excessive signal distortion, which can occur with higher-order filters. Moreover, the frequency range of interest for EEG bands extraction is usually below 100 Hz, and a fifth-order Butterworth filter can provide sufficient attenuation at higher frequencies while preserving the signal in the frequency range of interest.

Figure 5.3 represents the filtered bands of the EEG data of subject’s second visit. The first graph shows the original time-series of the first 1000 time points. All other graphs visualize delta, theta, alpha, beta and gamma bands in the mentioned order. We added this figure to show the output of the filtering, thus only the data for the second visit was chosen.

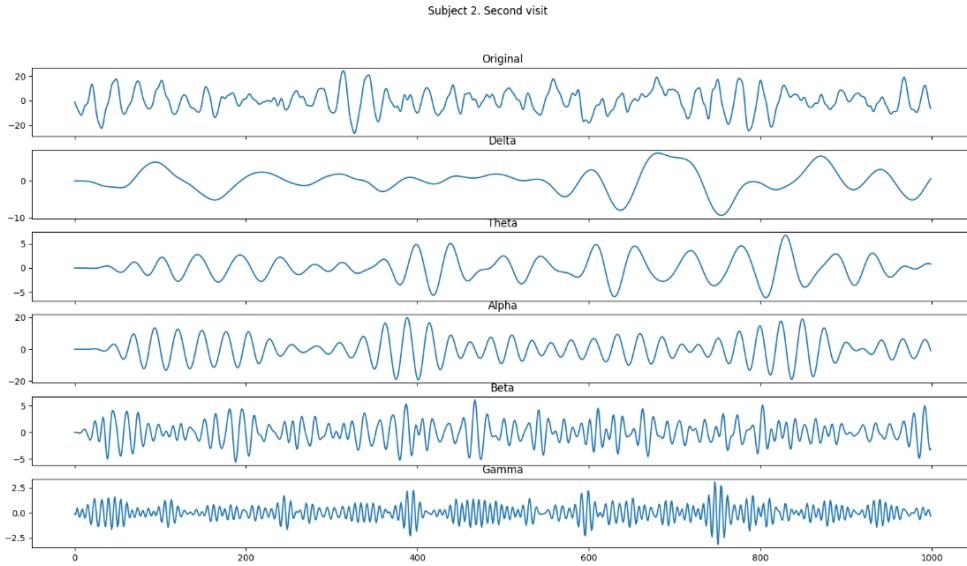


Figure 5.3 – Visualization of different bands of the depressed patient for the 2nd visit

5.2.13 Hilbert Transform

This preprocessing strategy was used to extract the instantaneous phase and amplitude of a signal. To do so, first the so-called analytical signal was found. An analytic signal is a type of complex-valued function that is defined in such a way that it has no negative frequency components (Smith 2007). The real and imaginary parts of an analytic signal are two real-valued functions that are related to each other in a specific way.

The Hilbert transform is a mathematical operation that takes a function of a real variable, $x(t)$, and produces another function of a real variable, $H(x)(t)$. It is a specific type of linear operator, which means that it follows certain rules for combining functions in a predictable way.

The analytical signal using Hilbert transform is computed using the next formula (Scipy signal hilbert):

$$x_a = F^{-1}(F(x)2U) = x + iy,$$

where:

- F is the Fourier transform,
- U is the unit step function
- y is the Hilbert transform of x .

The amplitude is then obtained by getting the absolute value of the analytical signal:

$$\text{Amplitude} = |x_a|$$

The phase of a signal can be wrapped or unwrapped. Wrapped phase is a phase representation that wraps the phase angle to a limited range, such as $[-\pi, \pi]$ or $[0, 2\pi]$, depending on the convention used. On the other hand, unwrapped phase is a phase representation that removes the wrapping and provides a continuous phase angle that can be easily interpreted.

The phase is obtained by calculating the unwrapped phase of the analytical signal.

The Tables 5.21-5.22 show the results of clustering on the amplitude and phase of bands extracted from Hilbert transform operation. This clustering was done only on the combined EEG data.

Table 5.21 – Results of clustering on the bands phase obtained from Hilbert transformation.

Band	Configuration	Cluster Purity	Random Index	Information Criterion
Delta	Untransformed	52%	0.5	0.33
Delta	Z-score	52%	0.5	0.33
Delta	Box-Cox and Z-score	51%	0.5	0.33
Theta	Untransformed	52%	0.5	0.33
Theta	Z-score	52%	0.5	0.33
Theta	Box-Cox and Z-score	52%	0.5	0.31
Alpha	Untransformed	52%	0.5	0.32
Alpha	Z-score	52%	0.5	0.32
Alpha	Box-Cox and Z-score	54%	0.5	0.32
Beta	Untransformed	57%	0.51	0.32
Beta	Z-score	56%	0.5	0.33
Beta	Box-Cox and Z-score	53%	0.5	0.32
Gamma	Untransformed	54%	0.5	0.31
Gamma	Z-score	55%	0.5	0.31
Gamma	Box-Cox and Z-score	52%	0.5	0.33

One can observe that the best CP was obtained by clustering on the untransformed Beta band phase obtained from Hilbert transformation of the data. This could be because individuals with anxiety type depression had significantly higher levels of beta 1 and beta 2 power compared to normal controls, specifically over the parietal and occipital regions for beta 1 and the frontal region for beta 2 as was shown in (Yamada et al. 1995).

Table 5.22 – Results of clustering on the bands amplitude obtained from Hilbert transformation.

Band	Configuration	Cluster Purity	Random Index	Information Criterion
Delta	Untransformed	53%	0.5	0.33
Delta	Z-score	51%	0.5	0.33
Delta	Box-Cox and Z-score	53%	0.5	0.33
Theta	Untransformed	55%	0.5	0.32
Theta	Z-score	56%	0.5	0.32
Theta	Box-Cox and Z-score	52%	0.5	0.33
Alpha	Untransformed	52%	0.5	0.32
Alpha	Z-score	54%	0.5	0.32
Alpha	Box-Cox and Z-score	52%	0.5	0.33
Beta	Untransformed	54%	0.5	0.32
Beta	Z-score	50%	0.5	0.3
Beta	Box-Cox and Z-score	51%	0.5	0.33
Gamma	Untransformed	52%	0.5	0.3
Gamma	Z-score	50%	0.5	0.31
Gamma	Box-Cox and Z-score	51%	0.5	0.33

The best result using this transformation techniques was clustering on the z-score transformed Hilbert amplitude of the theta band.

5.2.14 Selected Electrodes

Another hypothesis we postulated was to use only specific combination of electrodes for improving the clustering results. Thus, to test the hypothesis, the time-series of the chosen electrodes was left in the dataset and the time-series of other electrodes was removed. We have done these selections after consulting with the project partner who is a medical doctor in psychiatry.

Table 5.23 represents the results of clustering on the left and right located electrodes as well as front and back located electrodes. The data for this clustering was not transformed in any way.

Table 5.23 – Results of clustering on the specific electrodes on the non-transformed data.

Location of Setting electrodes		Cluster Purity	Random Index	Information Criterion
Back	First visit	52.1%	0.50	0.33
Back	Second visit	50.7%	0.50	0.33
Back	Combined	52.1%	0.50	0.33
Front	First visit	51%	0.5	0.33
Front	Second visit	50.9%	0.5	0.33
Front	Combined	50%	0.5	0.31
Left	First visit	50%	0.50	0.33
Left	Second visit	51.4%	0.50	0.33
Left	Combined	53.6%	0.50	0.33
Right	First visit	50%	0.50	0.33
Right	Second visit	50.9%	0.50	0.32
Right	Combined	51.4%	0.50	0.33

These locations include the next electrodes:

- Back: T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6 O1, O2
- Front: T3, C3, Cz, C4, T4, F7, F3, Fz, F4, F8, Fp1, Fp2
- Left: Fz, Cz, Pz, Fp1, F3, F7, C3, T3, P3, T5, O1
- Right: Fz, Cz, Pz, Fp2, F4, F8, C4, T4, P4, T6, O2

5.2.15 Clustering on Coefficients

The definition of coefficients is presented in Section “Interaction k-means. General Definitions”. After coefficients extraction the k-means, DBSCAN and hierarchical agglomerative clustering algorithms were applied.

To use these clustering methods on the data of multiple objects which is presented in the form of multi-dimensional matrices, the matrices were “flattened” such that all coefficients of one object represented a row in the dataset. First, the coefficients for the first and second visit were identified. Then, they were concatenated. Clustering was done on the resulting dataset.

Table 5.24 represents the results of the clustering using different configurations.

It can be seen, that in all configurations, the best run of the DBSCAN algorithm (based on the epsilon and minimum number of samples) identified only one cluster marking all points that were outside of this cluster as noise. Thus, the metrics of this algorithm were not considered in the table as they are not representative.

Other algorithms such as agglomerative clustering and simple k-means did not show any significant results. The maximum cluster purity was 52% for several configurations.

Table 5.24 – Results of clustering on the coefficients obtained from the different parts of the data.

Location of electrodes	Algorithm	Cluster Purity	Random Index	Information Criterion	Comment
Back	DBSCAN (Epsilon = 6.5, minimum samples = 6)	--	--	--	Only one cluster and noise
	Agglomerative	52.4%	0.5	0.32	
	k-means	51.6%	0.5	0.32	
Front	DBSCAN (Epsilon = 7, minimum samples = 6)	--	--	--	Only one cluster and noise
	Agglomerative	50.1%	0.5	0.33	
	k-means	50.1%	0.5	0.33	
Left	DBSCAN (Epsilon = 43, minimum samples = 7)	--	--	--	Only one cluster and noise
	Agglomerative	52.4%	0.5	0.32	
	k-means	51.6%	0.5	0.32	
Right	DBSCAN (Epsilon = 47.5, minimum samples = 6)	--	--	--	Only one cluster and noise
	Agglomerative	52.4%	0.5	0.32	
	k-means	51.6%	0.5	0.32	

Another configuration of the clustering data included choosing 3 electrodes from the left hemisphere (Fp1, F7, F3) and applying this reduction to the matrices of coefficients. Thus, the resulting matrices got size 18×3 . The results of the clustering are in Table 5.25.

Table 5.25 – Results of clustering on the reduced coefficients matrices.

Algorithm	Cluster Purity	Random Index	Information Criterion	Comment
DBSCAN (Epsilon = 49.5, minimum samples = 6)	--	--	--	Only one cluster and noise
Agglomerative	50%	0.5	0.3	
k-means	50%	0.5	0.3	

The further approach was to first reduce the dataset and leave only dimensions corresponding to the electrodes Fp1, F7, and F3. The next step was to calculate coefficients for the reduced dataset using the least squares method. The resulting matrix for each subject got size 2×3 . The results of the clustering are in Table 5.26.

Table 5.26 – Results of clustering on the coefficients matrices of the reduced dataset.

Algorithm	Cluster Purity	Random Index	Information Criterion	Comment
DBSCAN (Epsilon = 2, minimum samples = 3)	--	--	--	Only one cluster
Agglomerative	50%	0.5	0.31	
k-means	50%	0.5	0.33	

The last two approaches did not show any improvements in the results.

5.2.16 Different Distance Error Types

As described in Section “Interaction k-means. Distance Errors Used”, different error types were used for clustering. The results are presented in Tables 5.27- 5.28. The combined dataset was used. Clustering using Hamming error was not possible due to the errors that this approach brought for calculations.

Table 5.27 – Results of clustering on the combined dataset with the different distance errors.

Error Type	Clusters Purity	Rand Index	Information Criterion
Total	51.6%	0.50	0.33
Max	53%	0.50	0.33
Jaccard	57.8%	0.51	0.32

Table 5.28 – Results of clustering on the combined dataset with the different distance errors considering only right located electrodes.

Error Type	Clusters Purity	Rand Index	Information Criterion
Total	51.6%	0.5	0.33
Max	53%	0.5	0.33
Jaccard	57.8%	0.51	0.32

One can see that clustering using Jaccard distance error brought the best CP results in both cases.

5.2.17 Mixed Transformations

This approach included mixing different transformation techniques such as, Box-Cox and z-score.

The detailed results are presented in Table C-1 in Appendix C. Some combination of settings was not feasible to be used for calculations because, for example, some of the matrices were singular as an outcome of the specific transformations on them. All clustering was done on the combined data.

The best results were for the next setups:

- Left located + Cz, Delta band, Euclidean error type with the cluster purity **60.3%**
- Left located + Cz, Hilbert amplitude, delta band, Euclidian error type with the cluster purity **60.3%**
- Left located + Cz, Box-Cox, z-score, Euclidian error type with the cluster purity **60.5%**

Figure 5.4 shows the electrodes chosen for the clustering to obtain the best clusters purity results.

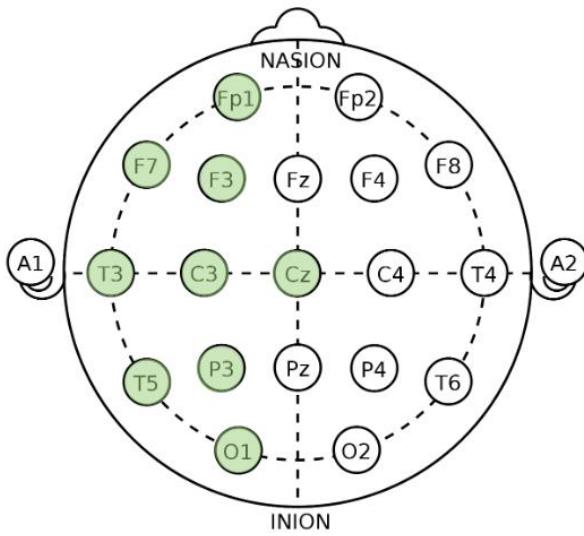


Figure 5.4 – Marked in green are electrodes chosen for the clustering to obtain the best clusters purity results.

Such improvement in the clusters purity might be supported by the theory that the left hemisphere is broadly associated with the emotions that are positively valenced, while the right hemisphere is broadly associated with the emotions that are negatively valenced (Gibson et al. 2022).

5.3 Interpretation of the Results

To interpret the results, the algorithm described in Section “Interaction k-means. Model Finding” was used. I interpreted the results for the “ideal” clustering and the best clustering setup, namely Left located + Cz electrodes, Box-Cox, z-score and Euclidean error type.

“Ideal” clustering means running the algorithm on the artificially ideally separated clusters. In each of these clusters, only subjects of one type are present. For example, cluster one would have only subjects that did not respond to the treatment and cluster two would have subjects that responded to the treatment.

The first step in the interpretation was to extract coefficients for these two setups. For the “ideal” clustering, we found coefficients by applying the least squares method to the clustering. For the best clustering setting, the coefficients used in the best clustering were taken.

The next step is to identify the most significant electrodes. To find out these electrodes, the interpretation algorithm shall be used as mentioned in Section “Interpretation of the Results”. The outcome of the algorithm is the list of errors for each dimension (electrode). This list shall be sorted ascendingly to identify the best discriminating models (the ones that are on the top of the sorted list).

Taking several of the most significant electrodes, the investigation of what is their relationship to other electrodes can be done. To do so, the extracted coefficients mentioned in the first step shall be transformed to have their absolute value. Then, the strongest relationships are the ones with the highest value of coefficients.

As can be seen in Figure 5.5, the Euclidian errors of the electrodes of the “ideal” clusters all have positive values. It means that despite ideally separating subjects based on the response to the treatment criteria, many of the subjects should not belong to one cluster with other subjects according to the algorithm. The error of the electrodes in this case is bigger for the clusters to which the subjects were initially assigned and not for the opposite cluster. Thus, the interpretation of the relationship between electrodes in this situation can be hard and we are not able to make any conclusive statements.

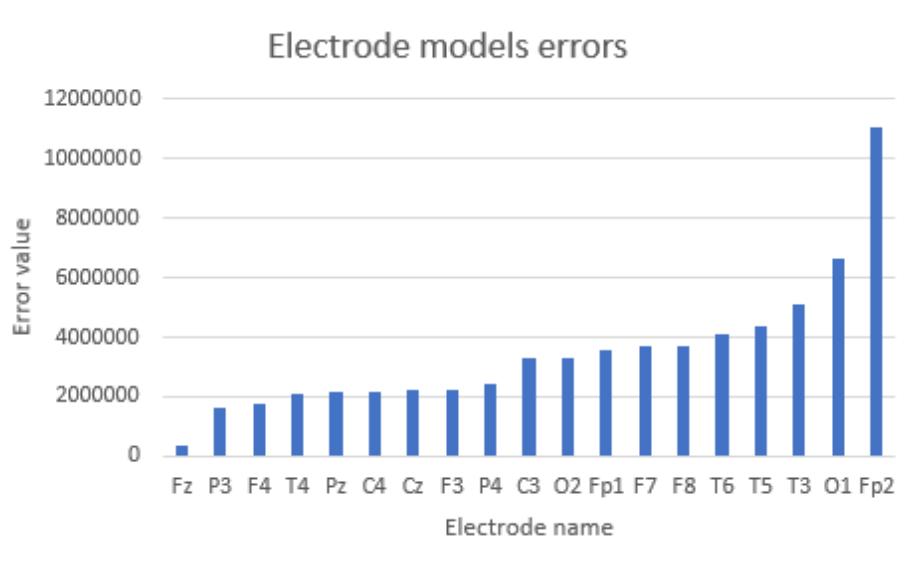


Figure 5.5 – Error values of electrodes models of “ideal” clusters. Positive values of errors here mean that most subjects were allocated in the wrong cluster.

On the other hand, for the clustering on the left located + Cz electrodes applying Box-Cox and z-score transformations, all the Euclidian errors have negative values. Figure 5.6 represents these error values. From the graph, models of electrodes P3 and T5 distinguish the best among the defined clusters.

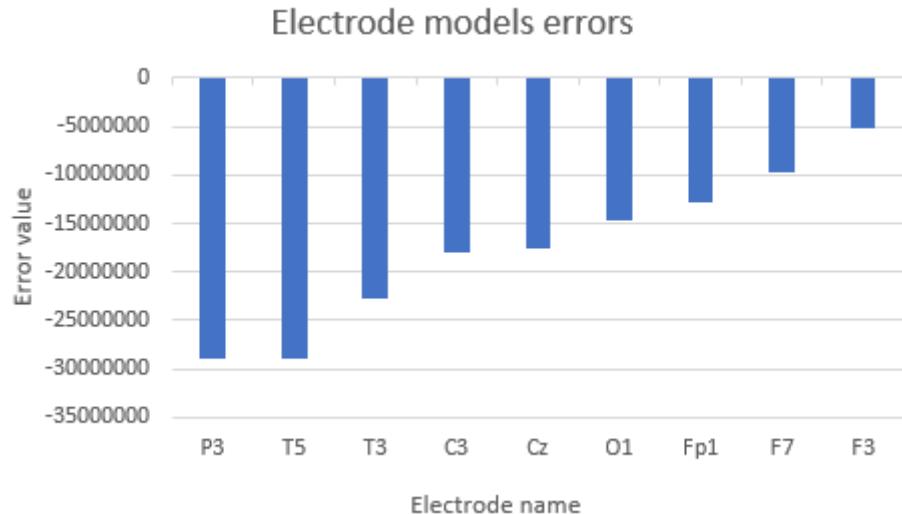


Figure 5.6 – Error values of Cz, Fp1, F3, F7, C3, T3, P3, T5, O1 models applying Box-Cox and z-score transformations for clustering. The smaller the error the better the model of the electrode discriminate among the clusters.

Cluster 1 from the best clustering contained 41 non-responders and 26 responders. Cluster 2 contained 27 non-responders and 40 responders. Thus, we called Cluster 1 “non-responders” cluster and Cluster 2 “responders” cluster.

Figure 5.7 shows the relationship of the most discriminative electrodes to other electrodes in “non-responders” cluster. The electrodes express values corresponding to the scale. The values that are represented on the scale are absolute values of the coefficients. The darker the color the higher the value. It can be seen that P3 (green in the left figure) and T5 (green in the right figure) electrodes have the strongest relation to the Fp1 and F3 electrodes respectively (shown by arrows). The absolute values of coefficients of the Fp1 and F3 electrodes are 1.05 and 0.76, respectively. We can argue that P3 and Fp1, and T5 and F3 electrodes represent some interaction patterns for depressed patients that did not respond to the treatment.

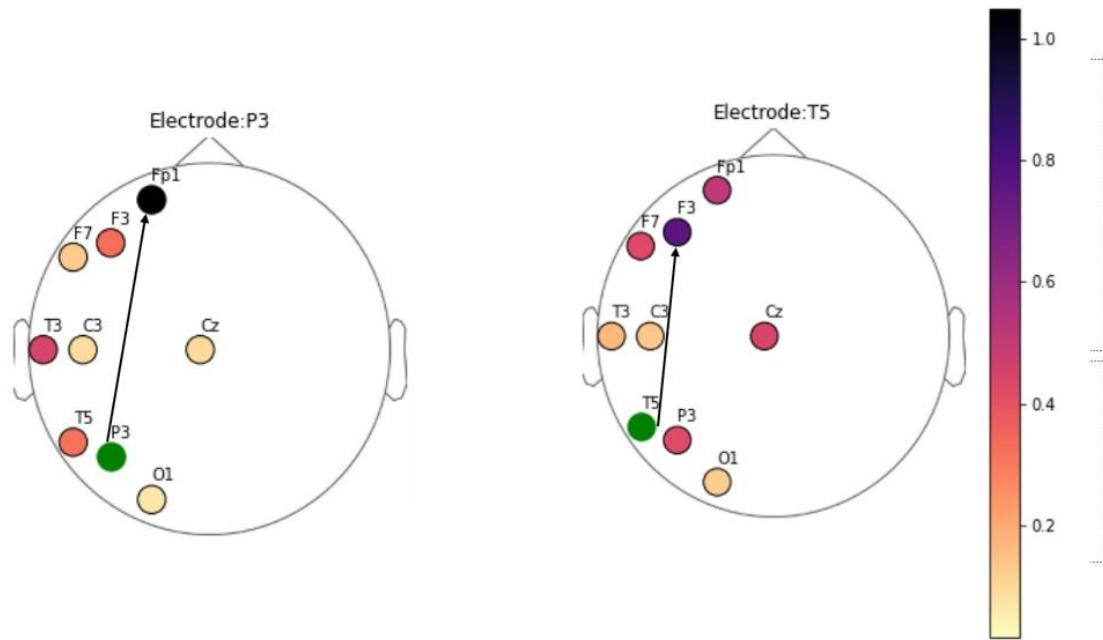


Figure 5.7 – The electrodes for “non-responders” cluster that discriminate the most (i.e., having the highest error value) among two clusters of the best clustering setting and their relationship to other electrodes.

Figure 5.8 shows the relationship of the most discriminative electrodes to other electrodes in “responders” cluster. Here electrode P3 has the strongest relationship to electrode F3 and somehow a strong relationship to electrodes Fp1, F7 and Cz. The absolute values of coefficients of the F3, Fp1, F7 and Cz electrodes are 0.81, 0.65, 0.63, 0.6 respectively. On the other hand, electrode T5 has the strongest relationship to electrode O1. The value of the coefficient of the O1 electrode is 0.61. Here as well, we can argue that P3 and F3, Fp1, F7, Cz, and T5 and O1 electrodes represent some interaction patterns for depressed patients that responded to the treatment.

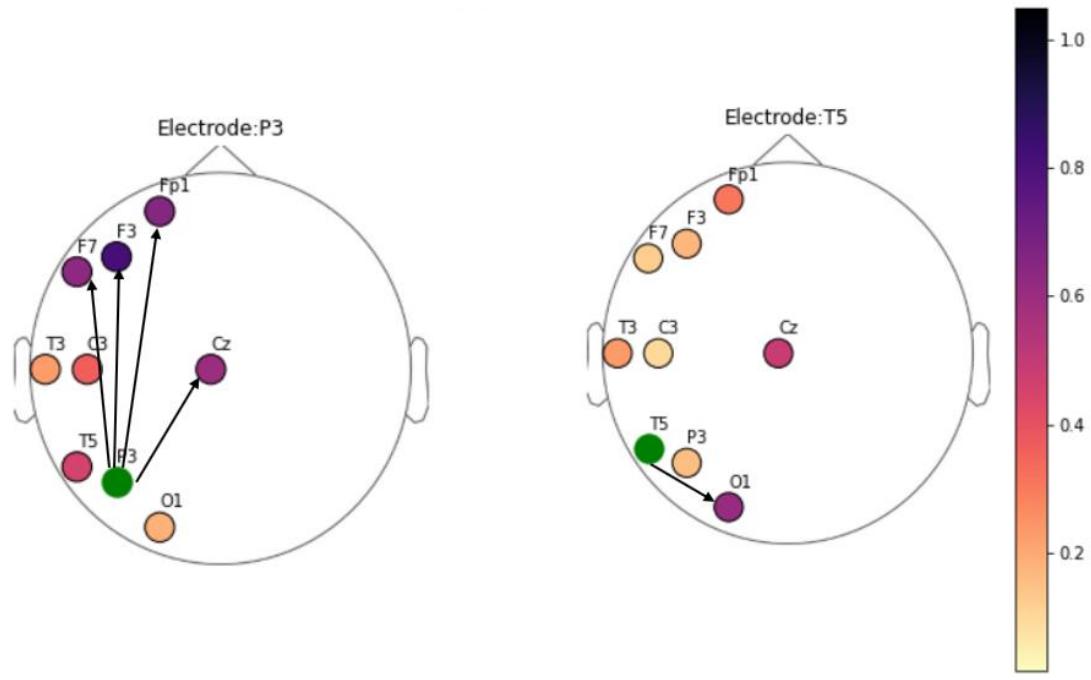


Figure 5.8 – The electrodes for “responders” cluster that discriminate the most among two clusters and their relationship to other electrodes.

Summing up these observations, one can see that the most discriminative electrodes for the clustering relate to different electrodes in their clusters. The P3 electrode in the “non-responders” cluster is strongly related to the Fp1 electrode whereas in the “responders” cluster it strongly related not only to Fp1 but also to the F7, F3 and Cz where the relation to F3 is the strongest. On the other hand, in the “non-responders” cluster, the electrode T5 is strongly related to F3 electrode whereas in the “responders” cluster it has the strongest relation to the O1 electrode.

It can be hypothesized that the different activities of the mentioned electrodes influence the allocating of a subject in one or another cluster.

6 Conclusion and Outlook

Clinical depression, also known as Major Depressive Disorder (MDD), is a mental disorder that is characterized by a persistent feeling of sadness, low self-esteem, and a loss of interest in activities that once brought pleasure. Modern treatments for MDD only have a success rate of up to 65%, and the effects of treatment can take several weeks to be observed. To better understand the brain regions and possible patterns in them of those who respond to treatment and those who do not, and to improve the choice of effective therapies, we have used clustering algorithms to analyze multi-trial EEG data of MDD patients in this work.

In this work, the Interaction K-means (IKM) clustering algorithm was applied to the EEG data. We used this algorithm to analyze multi-variate time-series data and this analysis allowed us to identify interaction patterns between brain regions. Various preprocessing strategies were used to improve the accuracy of the clustering results. The interaction between electrodes was also explored. Additionally, popular clustering methods such as k-means, DBSCAN, hierarchical clustering were applied to parameters derived from the data using the least squares method.

The best clustering results were obtained using a combination of the Box-Cox and z-score transformation methods, and considering only the Cz, Fp1, F3, F7, C3, T3, P3, T5, and O1 electrodes. The cluster purity (CP) for this setting was 60.5%. Similar results were obtained for clustering only the Delta band from the data of the same set of electrodes, with a cluster purity of 60.3%. The same cluster purity was observed for the derived Hilbert amplitude from the Delta band from the same set of electrodes.

Despite we could improve the CP, the Rand index (RI) and information criterion (IC) were not improved. Several potential reasons for this can be considered, such as the EEG data lacking discriminatory features for clustering, the high complexity of EEG data, limitations of the algorithm used, or the data set being too small to yield significant improvements. It should be noted that the EEG data was obtained from depressed patients, and the underlying neurological differences in these patients may have influenced the observed clustering results. Further research is necessary to gain a better understanding of the underlying factors that affect clustering results in this context.

Two electrodes, P3 and T5, were identified as the most discriminative among the clusters. In the cluster of non-responders, P3 and T5 were strongly related to the Fp1 and F3 electrodes, respectively. In the cluster of responders, P3 had strong relationships to the F3, Fp1, F7, and Cz electrodes, while T5 was strongly related to the O1 electrode.

Based on the identified electrodes and their relationships within and between clusters, it is possible to make some observations and hypotheses about the underlying neurological differences between responders and non-responders in the context of depression. We can hypothesise that the observed differences in electrode relationships between responders and non-responders may be related to underlying differences in neural connectivity or functional networks. Furthermore, it is possible that the identified electrodes and their relationships may have clinical implications for predicting treatment response in depressed patients. For example, future studies could investigate whether measuring activity in these specific electrode locations could be used as a biomarker for predicting treatment response in depressed patients.

We are aware that our results are the first attempt in exploratory analysis of the depressive patient's data set. The fact, that trajectories of time series of depressive patients are relatively uniform in the comparison to the epilepsy patients having seizures, makes our analysis a hard task, however worth of the further effort.

Further work on the improvement of the results might include using the phase models with sine or cosine. Additionally, the bigger dataset can be used for metrics improvement. Last but not least, the dataset with some stimuli might also improve the outcome of the IKM.

Bibliography

- Acharya, J. N., Hani, A., Cheek, J., Thirumala, P., & Tsuchida, T. N. (2016). American Clinical Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature. *Journal of Clinical Neurophysiology*, 33(4), 308–311. <https://doi.org/10.1097/WNP.0000000000000316>.
- Balazia, M., Hlavackova-Schindler K., Sojka P., and Plant C.. ‘Interpretable Gait Recognition by Granger Causality’. arXiv, 15 June 2022. <https://doi.org/10.48550/arXiv.2206.06714>.
- Belkhou A., Jbari A., and Belarbi L., ‘A Continuous Wavelet Based Technique for the Analysis of Electromyography Signals’, in 2017 International Conference on Electrical and Information Technologies (ICEIT) (2017 International Conference on Electrical and Information Technologies (ICEIT), Rabat: IEEE, 2017), 1–5, <https://doi.org/10.1109/EITech.2017.8255232>.
- Böhm, C., Läer L., Plant C., and Zherdin A.. ‘Model-Based Classification of Data with Time Series-Valued Attributes.’, 287–96, 2009.
- Box, G. E. P., & Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211–243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>.
- Bučková, Barbora, Brunovský M., Bareš M., and Hlinka J.. ‘Predicting Sex From EEG: Validity and Generalizability of Deep-Learning-Based Interpretable Classifier’. *Frontiers in Neuroscience* 14 (2020). <https://www.frontiersin.org/articles/10.3389/fnins.2020.589303>.
- Butterworth, S.. "On the theory of filter amplifiers." *Wireless Engineer* 7, no. 6 (1930): 536–541.
- Chen Y., Dong G., Han J., Pei J., Wah Benjamin W., & Wang J.. (2006). Regression Cubes with Lossless Compression and Aggregation. *IEEE Transactions on Knowledge and Data Engineering*, 18(12), 1585–1599. <https://doi.org/10.1109/TKDE.2006.196>.
- Cui, M. (2020). Introduction to the K-Means Clustering Algorithm Based on the Elbow Method. *Accounting, Auditing and Finance*, 1(1), 5–8. <https://doi.org/10.23977/accaf.2020.010102>.
- Dai, C., Pi, D., Cui, L., & Zhu, Y. (2018). MTEEGC: A novel approach for multi-trial EEG clustering. *Applied Soft Computing*, 71, 255–267. <https://doi.org/10.1016/j.asoc.2018.07.006>.
- Diagnostic and Statistical Manual of Mental Disorders: DSM-5™, 5th Ed, Diagnostic and Statistical Manual of Mental Disorders: DSM-5™, 5th Ed (Arlington, VA, US: American Psychiatric Publishing, Inc., 2013), <https://doi.org/10.1176/appi.books.9780890425596>.
- DiFrancesco, P.-M., Bonneau, D., & Hutchinson, D. J. (2020). The Implications of M3C2 Projection Diameter on 3D Semi-Automated Rockfall Extraction from Sequential Terrestrial Laser Scanning Point Clouds. *Remote Sensing*, 12(11), 1885. <https://doi.org/10.3390/rs12111885>.
- Dom B. E., ‘An Information-Theoretic External Cluster-Validity Measure’, 2013, <https://doi.org/10.48550/ARXIV.1301.0565>.

El Bouchefry, K., & de Souza, R. S. (2020). Chapter 12 - Learning in Big Data: Introduction to Machine Learning. In P. Škoda & F. Adam (Eds.), Knowledge Discovery in Big Data from Astronomy and Earth Observation (pp. 225–249). Elsevier. <https://doi.org/10.1016/B978-0-12-819154-5.00023-0>.

Ester, M., Kriegel, H.-P., & Xu, X. (n.d.). A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.

Fournier, J. C., DeRubeis, R. J., Hollon, S. D., Dimidjian, S., Amsterdam, J. D., Shelton, R. C., & Fawcett, J. (2010). Antidepressant Drug effects and Depression Severity: A Patient-Level Meta-Analysis. *JAMA : The Journal of the American Medical Association*, 303(1), 47–53. <https://doi.org/10.1001/jama.2009.1943>.

Gibson, B. C., Vakhtin, A., Clark, V. P., Abbott, C. C., & Quinn, D. K. (2022). Revisiting Hemispheric Asymmetry in Mood Regulation: Implications for rTMS for Major Depressive Disorder. *Brain Sciences*, 12(1), 112. <https://doi.org/10.3390/brainsci12010112>

Goldin, D. Q., & Kanellakis, P. C. (1995). On similarity queries for time-series data: Constraint specification and implementation. In U. Montanari & F. Rossi (Eds.), *Principles and Practice of Constraint Programming—CP '95* (Vol. 976, pp. 137–153). Springer Berlin Heidelberg. https://doi.org/10.1007/3-540-60299-2_9.

Guthrie, W. F. (2020). NIST/SEMATECH e-Handbook of Statistical Methods (NIST Handbook 151) [Data set]. National Institute of Standards and Technology. <https://doi.org/10.18434/M32189>.

Halkidi, M. (2009). Hierachial Clustering. In: LIU, L., ÖZSU, M.T. (eds) *Encyclopedia of Database Systems*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-39940-9_604

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems*, 17(2), 107–145. <https://doi.org/10.1023/A:1012801612483>.

Han, J., & Kamber, M. (2012). *Data mining: Concepts and techniques* (3rd ed). Elsevier.

Harezlak (2022, October 5). Machine Learning for Biostatistics. 2 K-means clustering. https://bookdown.org/tproto_home/Unsupervised-learning/k-means-clustering.html. Accessed 10 February 2023

Herwig, U., Satrapi, P., & Schönfeldt-Lecuona, C. (2003). Using the International 10-20 EEG System for Positioning of Transcranial Magnetic Stimulation. *Brain Topography*, 16(2), 95–99. <https://doi.org/10.1023/B:BRAT.0000006333.93597.9d>.

Hyndman, R.J., & Athanasopoulos, G. (2021) *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on 23.02.2023.

Jenkins, W., & Nayeri, M. (1986). Adaptive filters realized with second order sections. ICASSP '86. IEEE International Conference on Acoustics, Speech, and Signal Processing, 11, 2103–2106. <https://doi.org/10.1109/ICASSP.1986.1168633>.

- Kessler, R. C., & Bromet, E. J. (2013). The epidemiology of depression across cultures. *Annual Review of Public Health*, 34, 119–138. <https://doi.org/10.1146/annurev-publhealth-031912-114409>.
- Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62(6), 593–602. <https://doi.org/10.1001/archpsyc.62.6.593>.
- Kirmizi-Alsan, E., Bayraktaroglu, Z., Gurvit, H., Keskin, Y. H., Emre, M., & Demiralp, T. (2006). Comparative analysis of event-related potentials during Go/NoGo and CPT: Decomposition of electrophysiological markers of response inhibition and sustained attention. *Brain Research*, 1104(1), 114–128. <https://doi.org/10.1016/j.brainres.2006.03.010>.
- Kreyszig, E. (1979). Advanced engineering mathematics (4th ed). Wiley.
- Kruk, K. A., Aravich, P. F., Deaver, S. P., & deBeus, R. (2014). Comparison of Brain Activity During Drawing and Clay Sculpting: A Preliminary qEEG Study. *Art Therapy*, 31(2), 52–60. <https://doi.org/10.1080/07421656.2014.903826>.
- Larose D. T., Data Mining Methods and Models. Hoboken, NJ, USA: Wiley, 2006.
- Larose, D. T. (2015). *Data mining methods and models*. Hoboken, NJ: Wiley-Interscience.
- Luque A. T. and Bornas X., ‘Complexity and Irregularity in the Brain Oscillations of Depressive Patients: A Systematic Review’, *Neuropsychiatry* 07, no. 05 (2017), <https://doi.org/10.4172/Neuropsychiatry.1000238>.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, 5.1, 281–298.
- Mandal, J. K. (2020). Z-Transform-Based Reversible Encoding. In J. K. Mandal, Reversible Steganography and Authentication via Transform Encoding (Vol. 901, pp. 157–195). Springer Singapore. https://doi.org/10.1007/978-981-15-4397-5_7.
- Murugappan, M., Ramachandran, N., & Sazali, Y. (2010). Classification of human emotion from EEG using discrete wavelet transform. *Journal of Biomedical Science and Engineering*, 03(04), 390–396. <https://doi.org/10.4236/jbise.2010.34054>.
- Neurophysiology Society Guideline 2: Guidelines for Standard Electrode Position Nomenclature’, *Journal of Clinical Neurophysiology* 33, no. 4 (August 2016): 308–11, <https://doi.org/10.1097/WNP.0000000000000316>.
- Otte, C., Gold, S. M., Penninx, B. W., Pariante, C. M., Etkin, A., Fava, M., Mohr, D. C., & Schatzberg, A. F. (2016). Major depressive disorder. *Nature Reviews Disease Primers*, 2(1), Article 1. <https://doi.org/10.1038/nrdp.2016.65>.
- Patton, L. L., Glick, M., & American Dental Association (Eds.). (2016). The ADA practical guide to patients with medical conditions (Second edition). John Wiley & Sons, Inc.

- Plant, C., Zherdin, A., Sorg, C., Meyer-Baese, A., & Wohlschläger, A. M. (2013). Mining Interaction Patterns among Brain Regions by Clustering. *IEEE Transactions on Knowledge and Data Engineering*, 26(9), 2237–2249. <https://doi.org/10.1109/TKDE.2013.61>.
- Roberts, J., & Roberts, T. D. (1978). Use of the Butterworth low-pass filter for oceanographic data. *Journal of Geophysical Research*, 83(C11), 5510. <https://doi.org/10.1029/JC083iC11p05510>.
- Rot, M. aan het, Mathew, S. J., & Charney, D. S. (2009). Neurobiological mechanisms in major depressive disorder. *CMAJ*, 180(3), 305–313. <https://doi.org/10.1503/cmaj.080697>.
- Saby, J. N., & Marshall, P. J. (2012). The Utility of EEG Band Power Analysis in the Study of Infancy and Early Childhood. *Developmental Neuropsychology*, 37(3), 253–273. <https://doi.org/10.1080/87565641.2011.614663>.
- Scipy signal hilbert. `scipy.signal.hilbert` - SciPy v1.10.0 Manual. <https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.hilbert.html#r37d8c8a6fd16-1>. Accessed 10 February 2023
- Smith, J. O. (2007). Mathematics of the discrete Fourier transform (DFT): With audio applications (2. ed). BookSurge.
- Stearns, S. D., & Hush, D. R. (1990). Digital signal analysis. Prentice Hall.
- Sullivan D.M., ‘Z-Transform Theory and the FDTD Method’, *IEEE Transactions on Antennas and Propagation* 44, no. 1 (January 1996): 28–34, <https://doi.org/10.1109/8.477525>.
- Teixeira-Pinto A. & Harezlak J., 2 K-Means Clustering | Machine Learning for Biostatistics, accessed 24 December 2022, https://bookdown.org/tpinto_home/Unsupervised-learning/k-means-clustering.html.
- UCI Machine Learning Repository: Eeg database data set. <https://archive.ics.uci.edu/ml/datasets/eeg+database>. Accessed 10 February 2023
- Wang, X., Wirth, A., & Wang, L. (2007). Structure-Based Statistical Features and Multivariate Time Series Clustering. *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, 351–360. <https://doi.org/10.1109/ICDM.2007.103>.
- Yamada, M., Kimura, M., Mori, T., & Endo, S. (1995). [EEG power and coherence in presenile and senile depression. Characteristic findings related to differences between anxiety type and retardation type]. *Nihon Ika Daigaku zasshi*, 62(2), 176–185. <https://doi.org/10.1272/jnms1923.62.176>.
- Zherdin, A., ‘Efficient Data Mining Algorithms for Time Series and Complex Medical Data’, application/pdf (Ludwig-Maximilians-Universität München, 2016), <https://doi.org/10.5282/EDOC.19406>.

Appendices

Appendix A. Cluster Coefficients Correlation Matrices

In Section ‘‘Data Sets and Methods. Data Exploration’’ we presented some correlation matrices for clusters coefficients of ideal clusters. These correlation matrices we created for different combinations of depression patients visits and their response to the treatment. This Appendix contains the rest of those matrices.

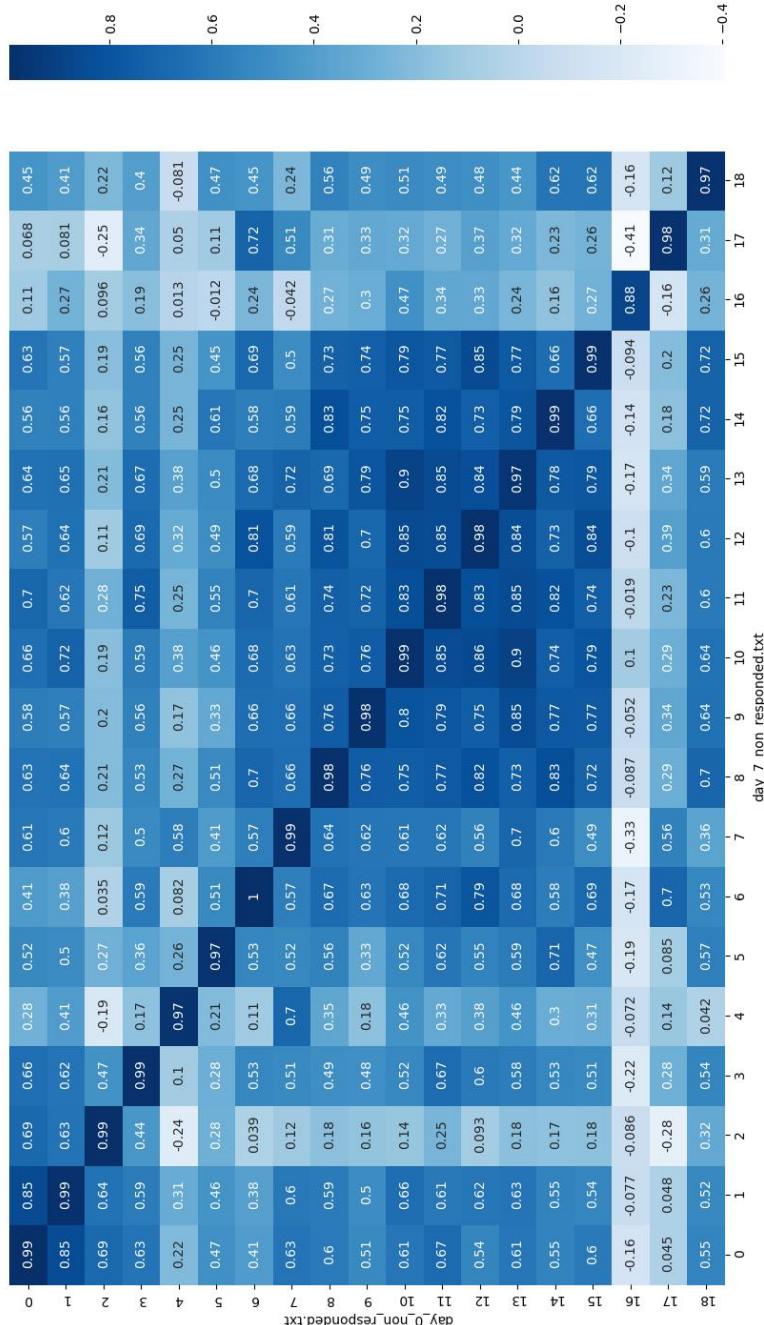


Figure A-1 – Correlation matrix for cluster coefficients of ideal clusters. The coefficients are calculated for the first visit non-responders and the second visit non-responders.

In Figure A-1, one can see that coefficients of 8th to 15th dimensions of one visit have stronger positive correlation with the similar dimensions from another visit.

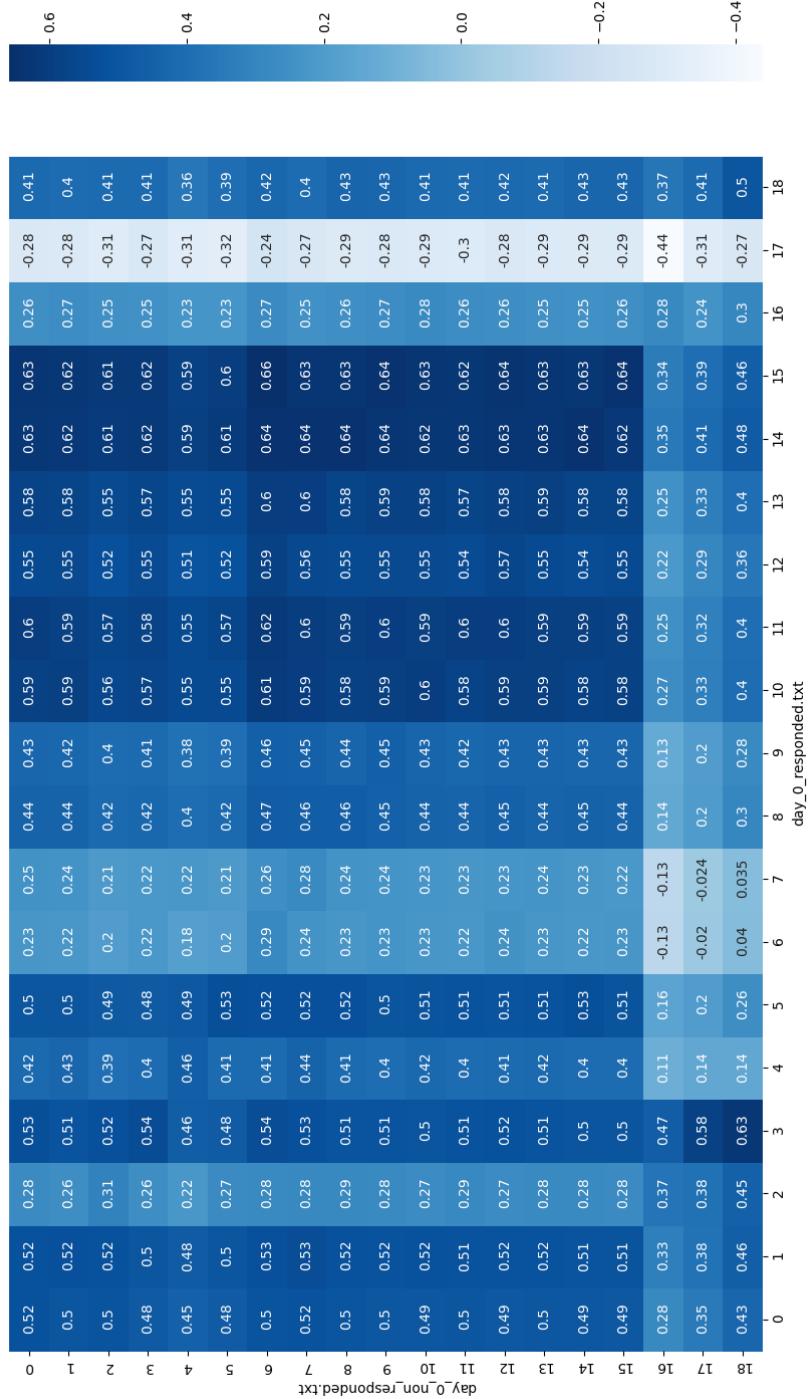


Figure A-2 – Correlation matrix for cluster coefficients of ideal clusters. The coefficients are calculated for the first visit non-responders and the first visit responders.

In Figure A-2, one can see that there is a stronger positive correlation between 14th and 15th coefficients of responders with coefficients 0th to 15th of non-responders.

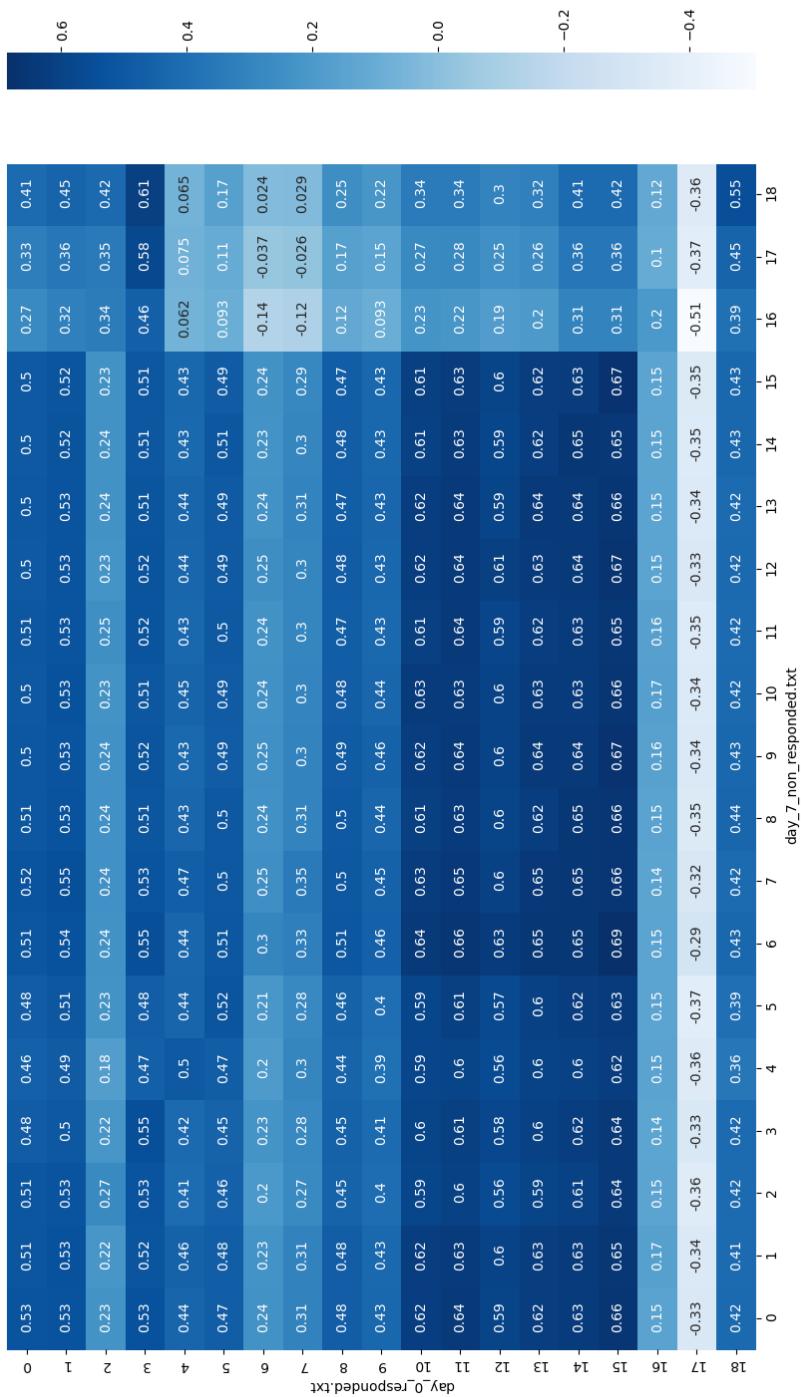


Figure A-3 – Correlation matrix for cluster coefficients of ideal clusters. The coefficients are calculated for the first visit responders and the second visit non-responders.

In Figure A-3, one can see that there is a stronger positive correlation between coefficients of 0th to 15th dimensions of non-responders from the second visit with 10th to 15th dimensions of the first visit.

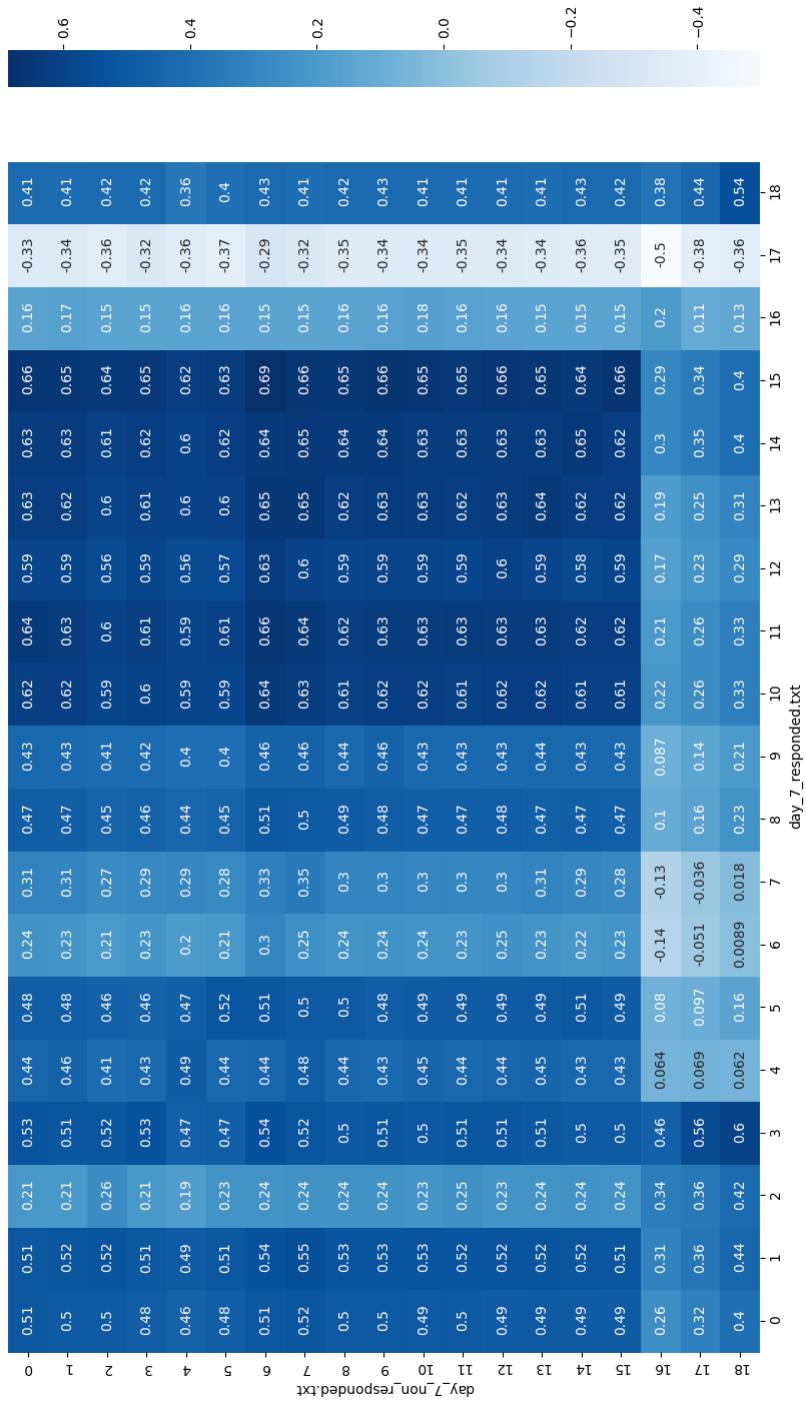


Figure A-4 – Correlation matrix for cluster coefficients of ideal clusters. The coefficients are calculated for the second visit non-responders and the second visit responders.

In Figure A-4, one can see that there is a stronger positive correlation between coefficients of 10th to 15th dimensions of responders from the second visit with 0th to 15th dimensions of responders from the same visit.

Appendix B. Each Object Coefficients Correlation Matrices

In Section “Data Sets and Methods. Data Exploration” we presented some correlation matrices for coefficients of each object allocated in ideal clusters. These correlation matrices we created for different combinations of depression patients visits and their response to the treatment. This Appendix contains the rest of those matrices.

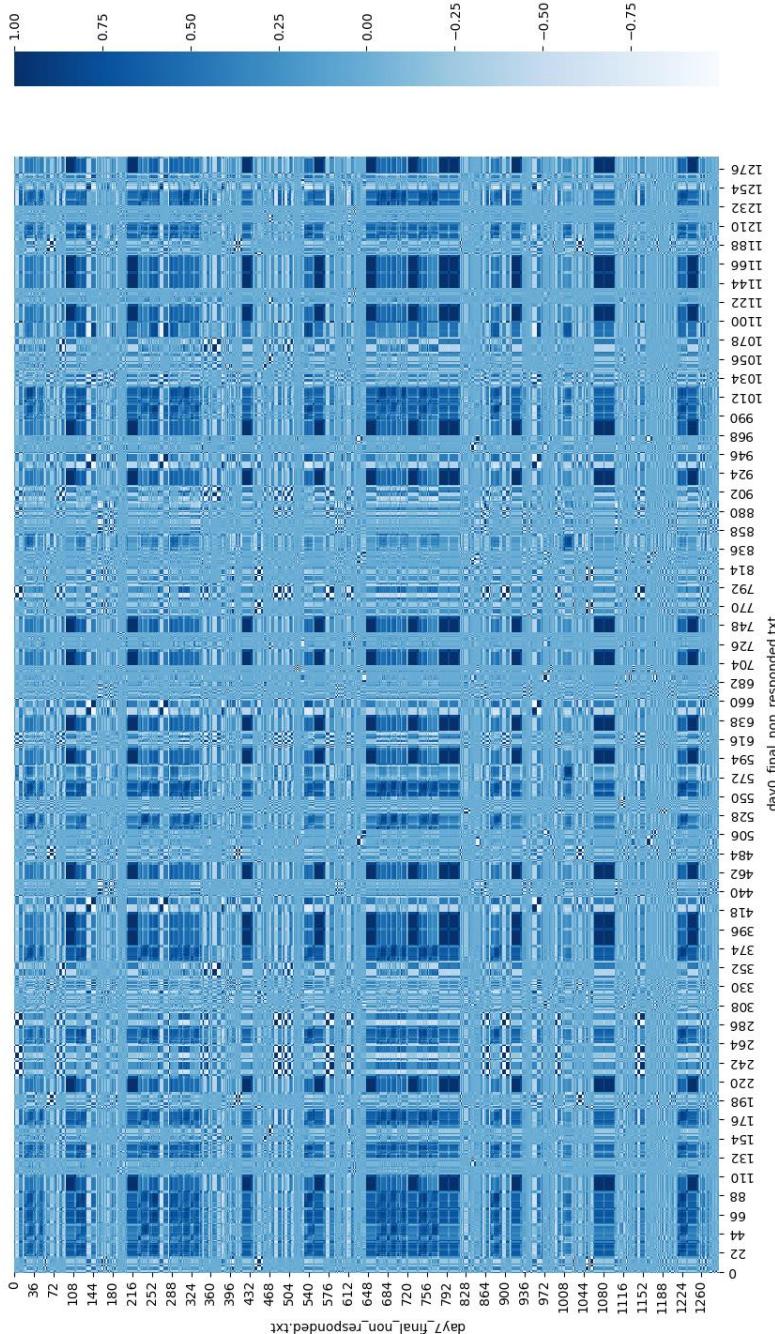


Figure B-1 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the first visit non-responders and the second visit non-responders.

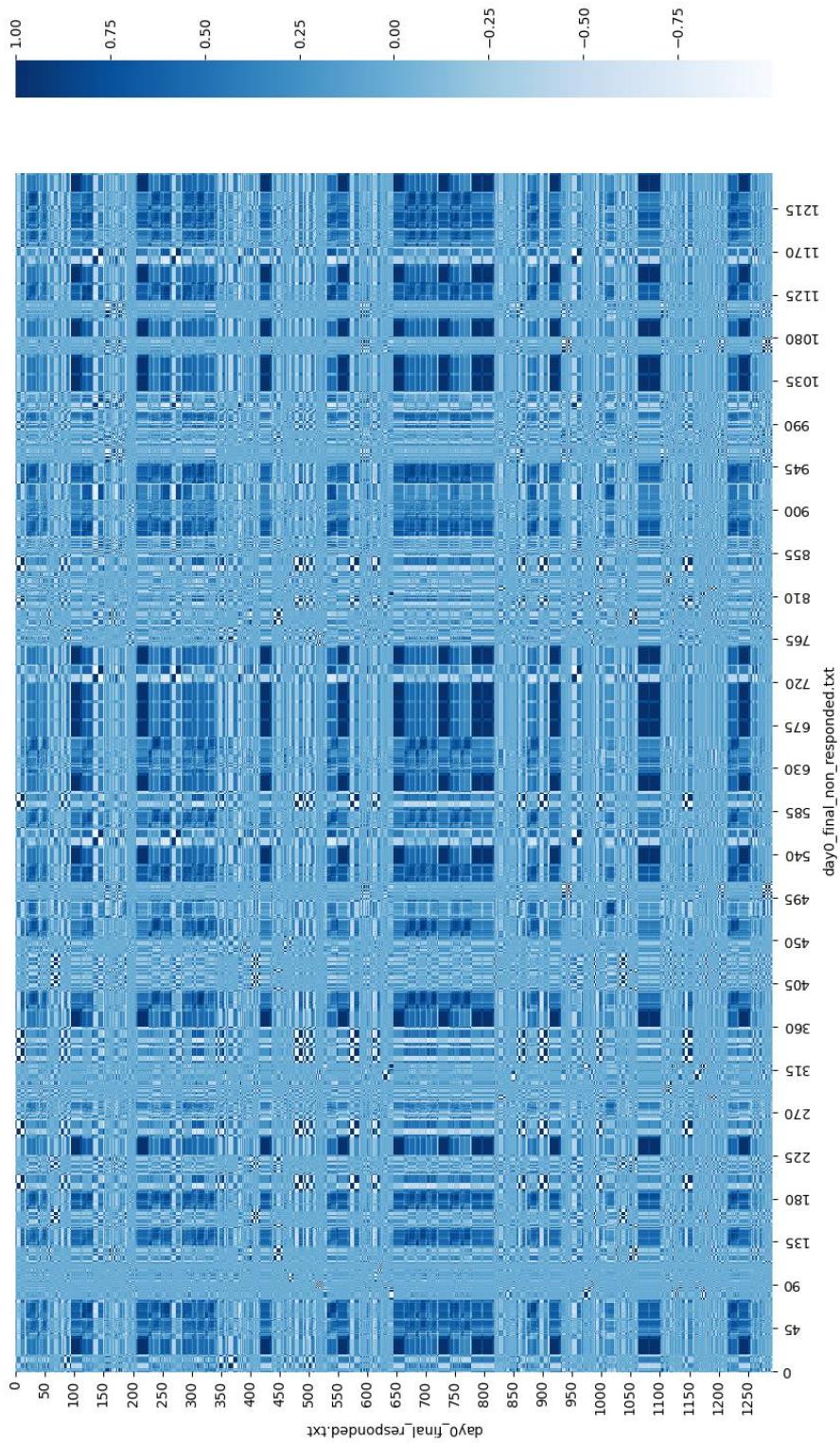


Figure B-2 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the first visit non-responders and the first visit responders.

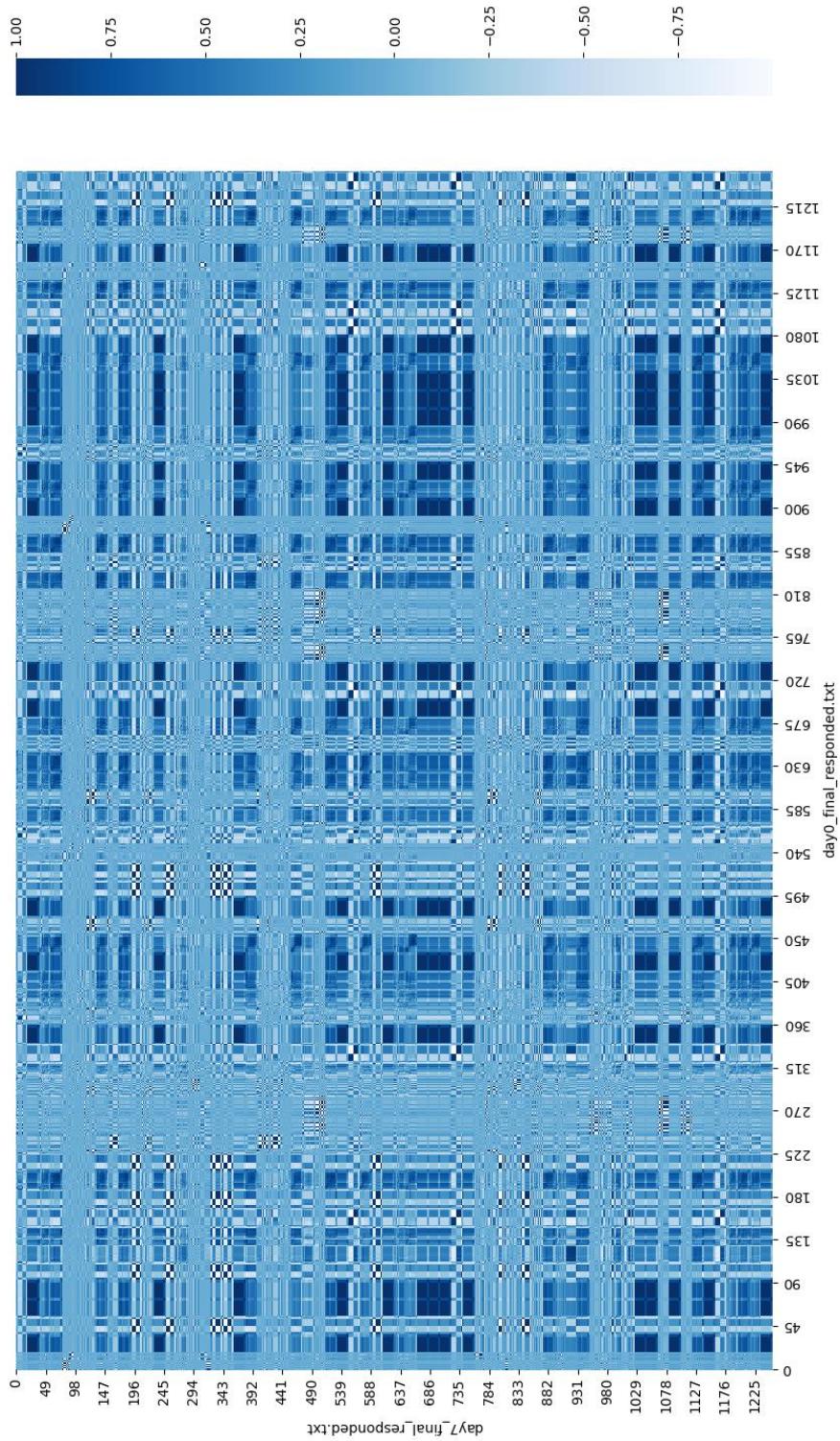


Figure B-3 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the first visit responders and the second visit responders.

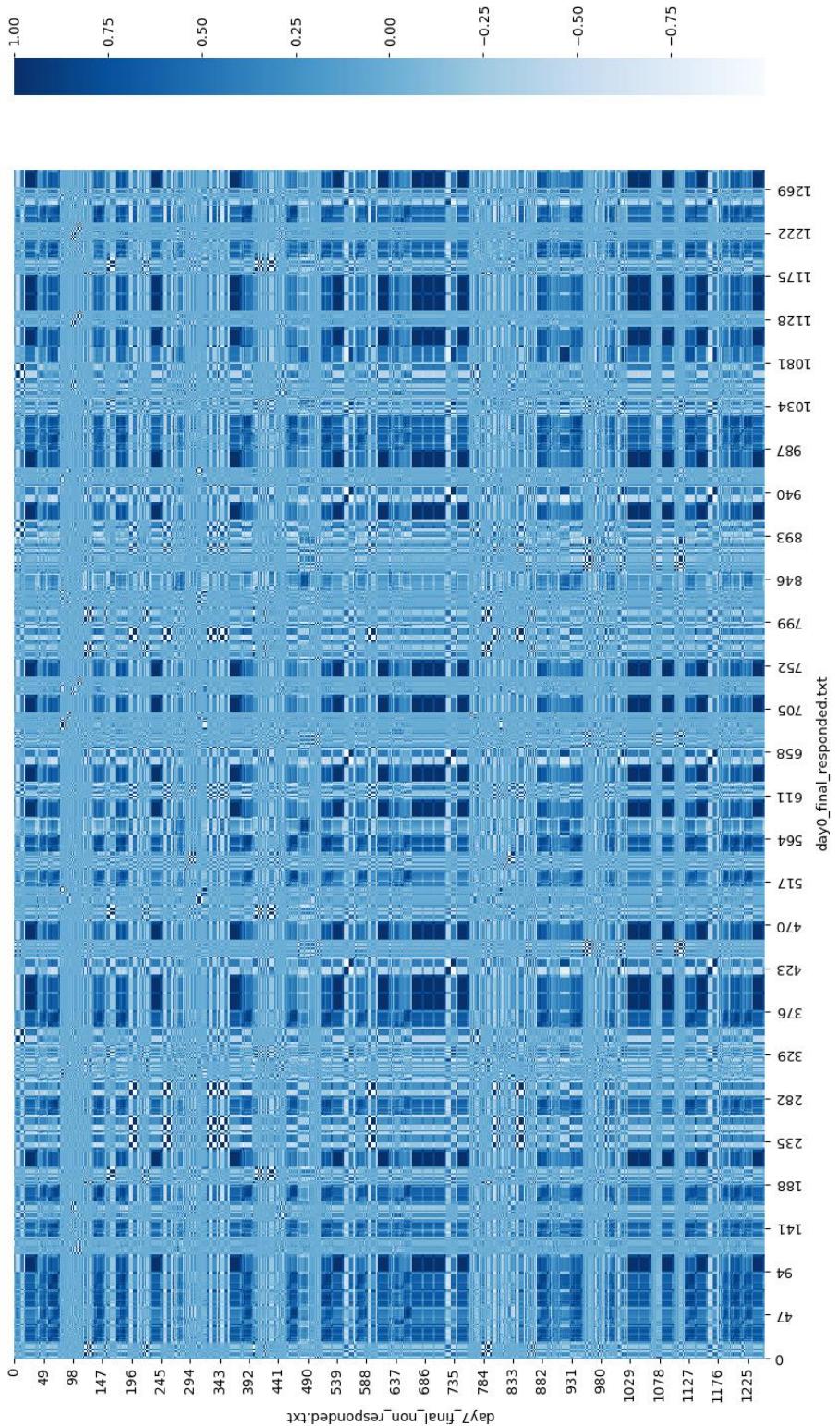


Figure B-4 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the first visit responders and the second visit non-responders.

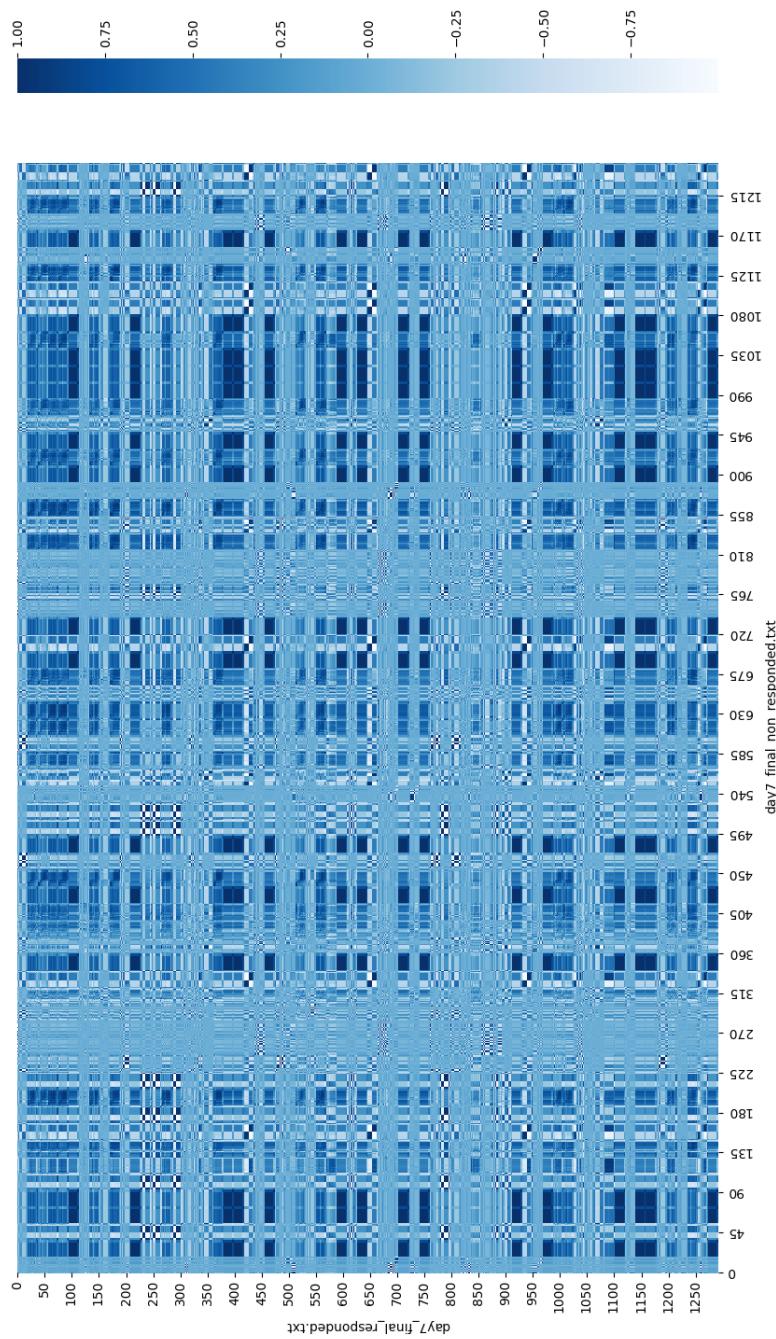


Figure B-5 – Correlation matrix for coefficients of each object allocated in ideal clusters. The coefficients are calculated for the second visit non-responders and the second visit responders.

In all correlation heatmaps presented in Figures B1-B5, one can see the strong positive correlation between some coefficients of some subjects. Almost in all cases this correlation can be observed in the middle of the set (such as for an object 34, for example). These middle-set coefficients of subjects are strongly positively correlated with the coefficients of most of the other subjects. No strong negative correlation in all settings can be observed.

Appendix C. Clustering Results on the Different Parts of Electrodes Using Mixed Transformations

This Appendix contains the results of clustering on the different parts of electrodes and using mixed transformations. Table C-1 shows these results. The most significant results are for the combinations:

- Left located + Cz, Delta band, Euclidean error type with the cluster purity **60.3%**
- Left located + Cz, Hilbert amplitude, delta band, Euclidian error type with the cluster purity **60.3%**
- Left located + Cz, Box-Cox, z-score, Euclidian error type with the cluster purity **60.5%**

Table C-1 – Results of clustering on the different parts of electrodes and using mixed transformations.

Electrodes locations	Transformation	Clusters Purity	Rand Index	Information Criterion	Error Type
Left located + Cz	Z-transform magnitude	50.8%	0.50	0.30	Euclidean
Left located + Cz	Hilbert phase	51.5%	0.50	0.33	Euclidian
Left located + Cz	Hilbert amplitude	55.3%	0.50	0.33	Euclidian
Left located + Cz	Delta band	60.3%	0.52	0.32	Euclidian
Left located + Cz	Hilbert phase, delta band	54.3%	0.50	0.33	Euclidian
Left located + Cz	Hilbert amplitude, delta band	60.3%	0.52	0.32	Euclidian
Left located + Cz	Theta band	56.6%	0.51	0.33	Euclidian
Left located + Cz	Hilbert phase, theta band	54.3%	0.50	0.33	Euclidian
Left located + Cz	Hilbert amplitude, theta band	56.6%	0.51	0.33	Euclidian
Left located + Cz	Beta band	55.7%	0.50	0.32	Euclidian
Left located + Cz	Hilbert phase, beta band	50.0%	0.50	0.33	Euclidian
Left located + Cz	Hilbert amplitude, beta band	55.7%	0.50	0.32	Euclidian

Electrodes locations	Transformation	Clusters Purity	Rand Index	Information Criterion	Error Type
Left located + Cz	Gamma band	50.0%	0.50	0.33	Euclidian
Left located + Cz	Hilbert phase, gamma band	51.7%	0.50	0.33	Euclidian
Left located + Cz	Hilbert amplitude, gamma band	50.0%	0.50	0.33	Euclidian
Left located + Cz	Alpha band	55.1%	0.50	0.33	Euclidian
Left located + Cz	Hilbert phase, alpha band	52.8%	0.50	0.33	Euclidian
Left located + Cz	Hilbert amplitude, alpha band	55.1%	0.50	0.33	Euclidian
Left located + Cz	Box-Cox, z-score	60.5%	0.52	0.32	Euclidian
Left located + Cz	Box-Cox, z-score	50.0%	0.50	0.33	Jaccard
Left located + Cz	Box-Cox, z-score	56.9%	0.51	0.33	Total
Left located + Cz	Box-Cox, z-score	56.1%	0.50	0.33	Max
Left located + Cz	Box-Cox, z-score	50.0%	0.50	0.33	Jaccard
Left located + Cz	Box-Cox, z-score	50.0%	0.50	0.32	Hamming
Left located + Cz	Box-Cox, z-normalization	50.7%	0.50	0.33	Euclidian
Left located	Box -Cox	50.8%	0.50	0.33	Euclidian
Left located	z-score	54.4%	0.50	0.33	Euclidian
Left located	Box -Cox, z-score	54.5%	0.50	0.33	Euclidian
Right located	Box -Cox	50.8%	0.50	0.31	Euclidian
Right located	z-score	53.0%	0.50	0.33	Euclidian
Right located	Box -Cox, z-score	53.1%	0.50	0.33	Euclidian
Front located	Box -Cox	51.6%	0.50	0.31	Euclidian
Front located	z-score	54.5%	0.50	0.33	Euclidian

Electrodes locations	Transformation	Clusters Purity	Rand Index	Information Criterion	Error Type
Front located	Box -Cox, z-score	55.3%	0.50	0.33	Euclidian
Back located	Box -Cox	50.8%	0.50	0.33	Euclidian
Back located	z-score	51.3%	0.50	0.33	Euclidian
Back located	Box -Cox, z-score	54.5%	0.50	0.33	Euclidian
Left located	Box-Cox, Hilbert phase	54.5%	0.50	0.33	Euclidian
Left located	Box-Cox, Hilbert amplitude	50.8%	0.50	0.33	Euclidian
Right located	Box-Cox, Hilbert phase	52.3%	0.50	0.33	Euclidian
Right located	Box-Cox, Hilbert amplitude	50.8%	0.50	0.31	Euclidian
Front located	Box-Cox, Hilbert phase	51.7%	0.50	0.32	Euclidian
Front located	Box-Cox, Hilbert amplitude	51.6%	0.50	0.31	Euclidian
Back located	Box-Cox, Hilbert phase	55.0%	0.50	0.33	Euclidian
Back located	Box-Cox, Hilbert amplitude	50.8%	0.50	0.33	Euclidian
Right located	Z-transform magnitude	52.7%	0.50	0.33	Euclidian
Right located	Box-Cox, Z-transform magnitude	50.8%	0.50	0.31	Euclidian
Right located	Z-transform phase	50.9%	0.50	0.33	Euclidian

Appendix D. Results Interpretation Details

This Appendix contains the visuals of the relationships between electrodes in “ideal clusters” as well as in clusters of the best clustering setting which is applying Box-Cox and z-score transformation to the left hemisphere electrodes + Cz electrode. The first cluster is the cluster with non-responders, the second is the cluster with responders in both cases.

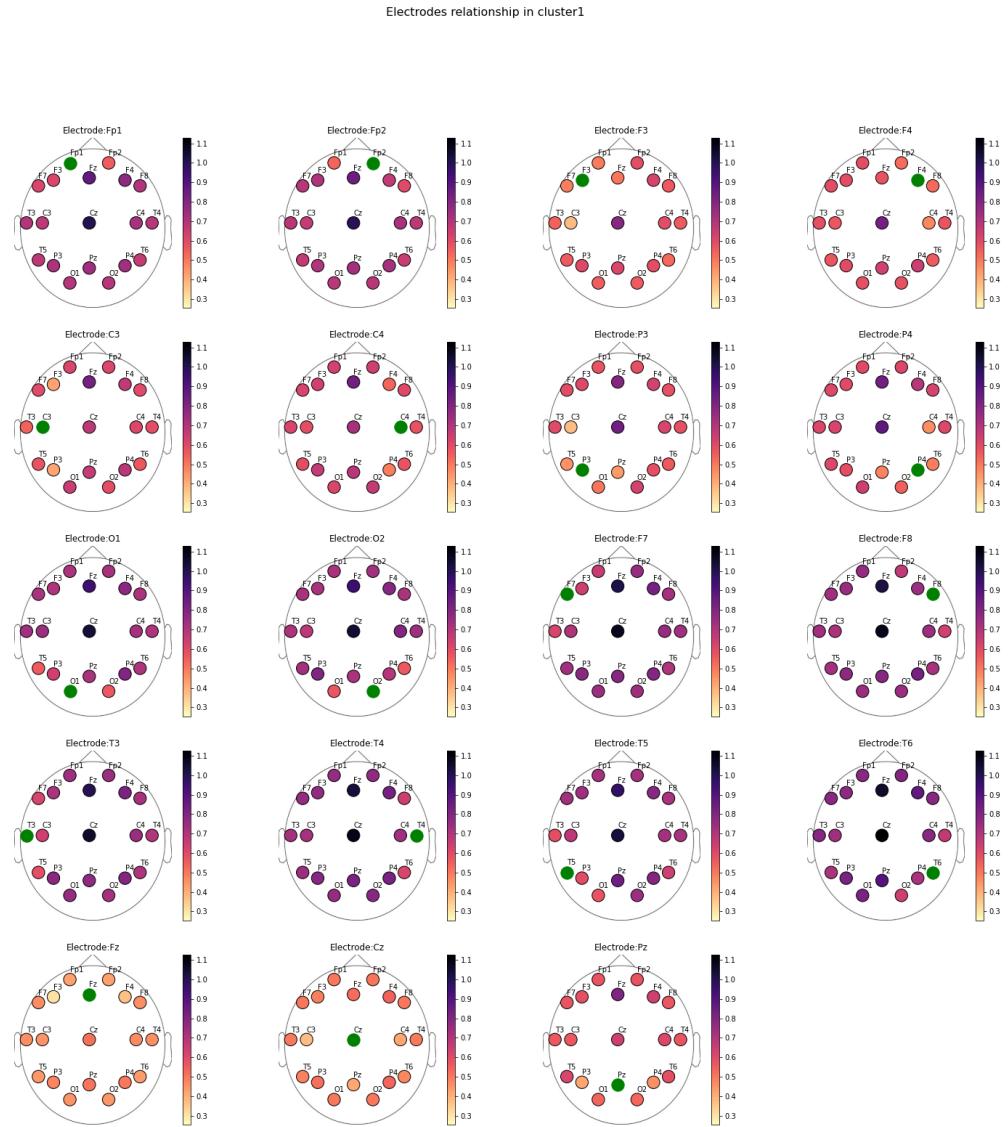


Figure D-1 – Relationships between electrodes of the first cluster with non-responders of “ideal” clusters.

From Figure D-1, one can see that electrodes O2, F7, F8, T3, T4, T5 and finally T6 have strong relationship almost to all other electrodes. On the other hand, the electrodes Fz, Cz and Pz have not very strong relation to all other electrodes.

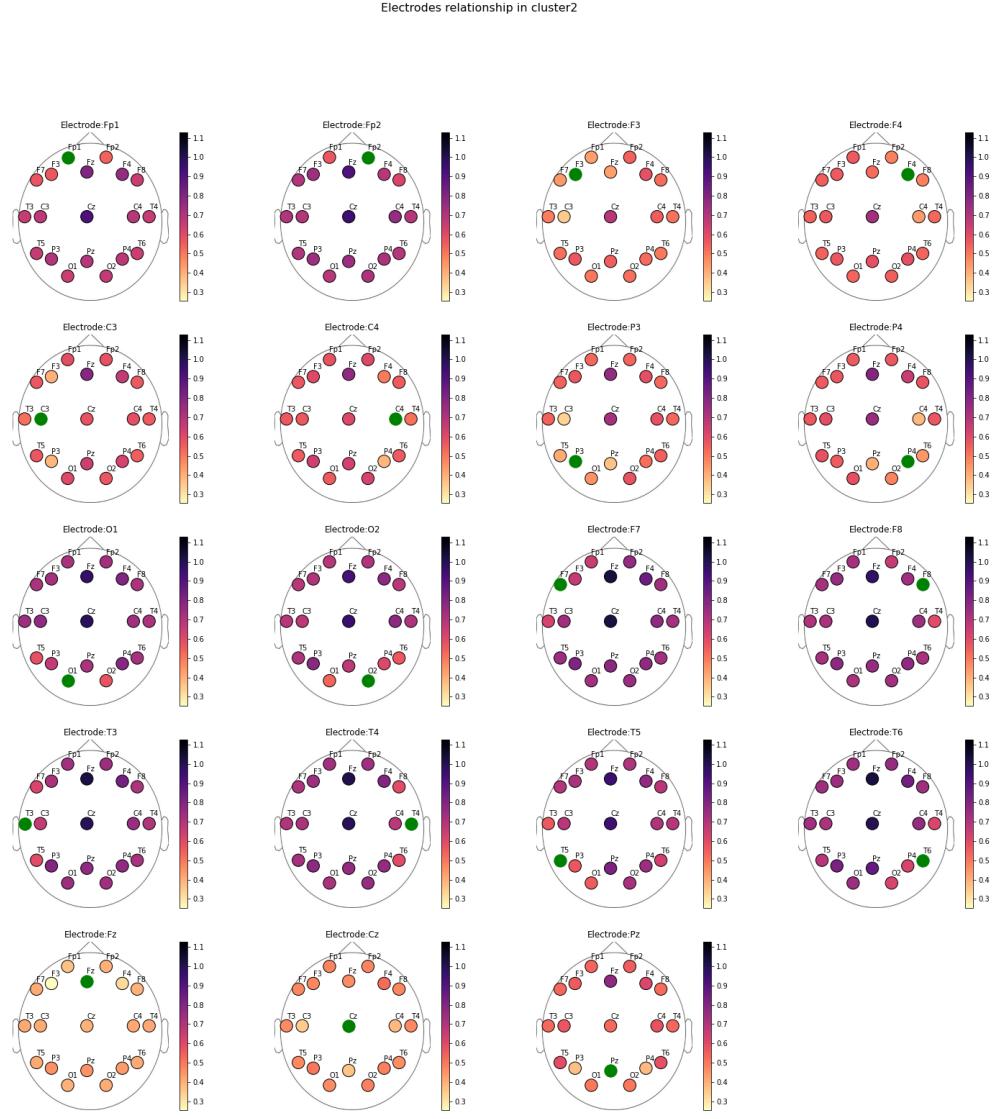


Figure D-2 – Relationships between electrodes of the second cluster with responders of “ideal” clusters.

From Figure D-2, one again can see that electrodes O2, F7, F8, T3, T4, T5 and finally T6 have strong relationship almost to all other electrodes. However, the number of those other electrodes is lower than it was for the “non-responders” cluster. The electrodes Fz, Cz and Pz have not a strong relation to all other electrodes but the electrode Pz has the slightly stronger relation to other electrodes comparing to the “non-responders” cluster.

Electrodes relationship in cluster1

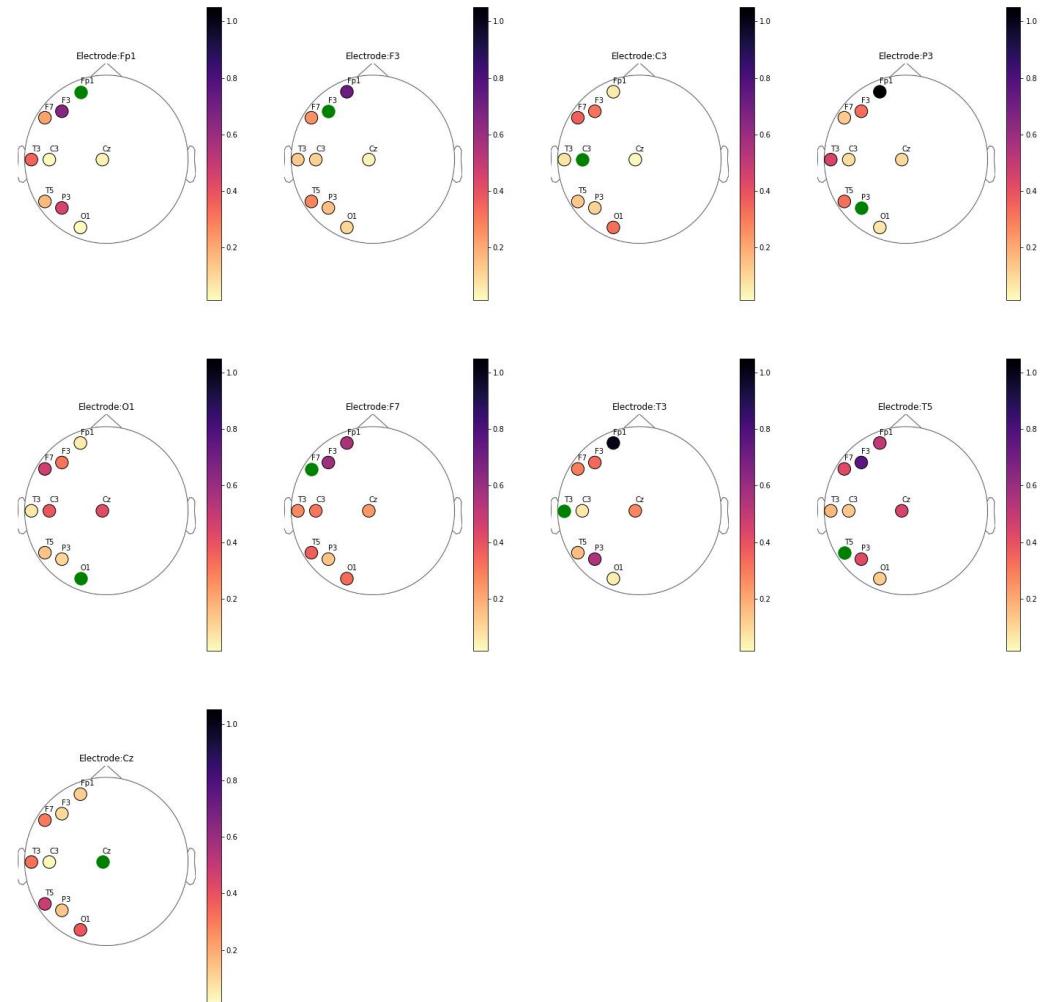


Figure D-3 – Relationships between electrodes of the first cluster obtained with the clustering setting of applying Box-Cox and z-score transformations on the left hemisphere electrodes + Cz.

In Figure D-3, one can see that F3, P3 and T3 have very strong relation to the electrode Fp1. On the other hand, C3 electrode does not show a very strong relation to other electrodes.

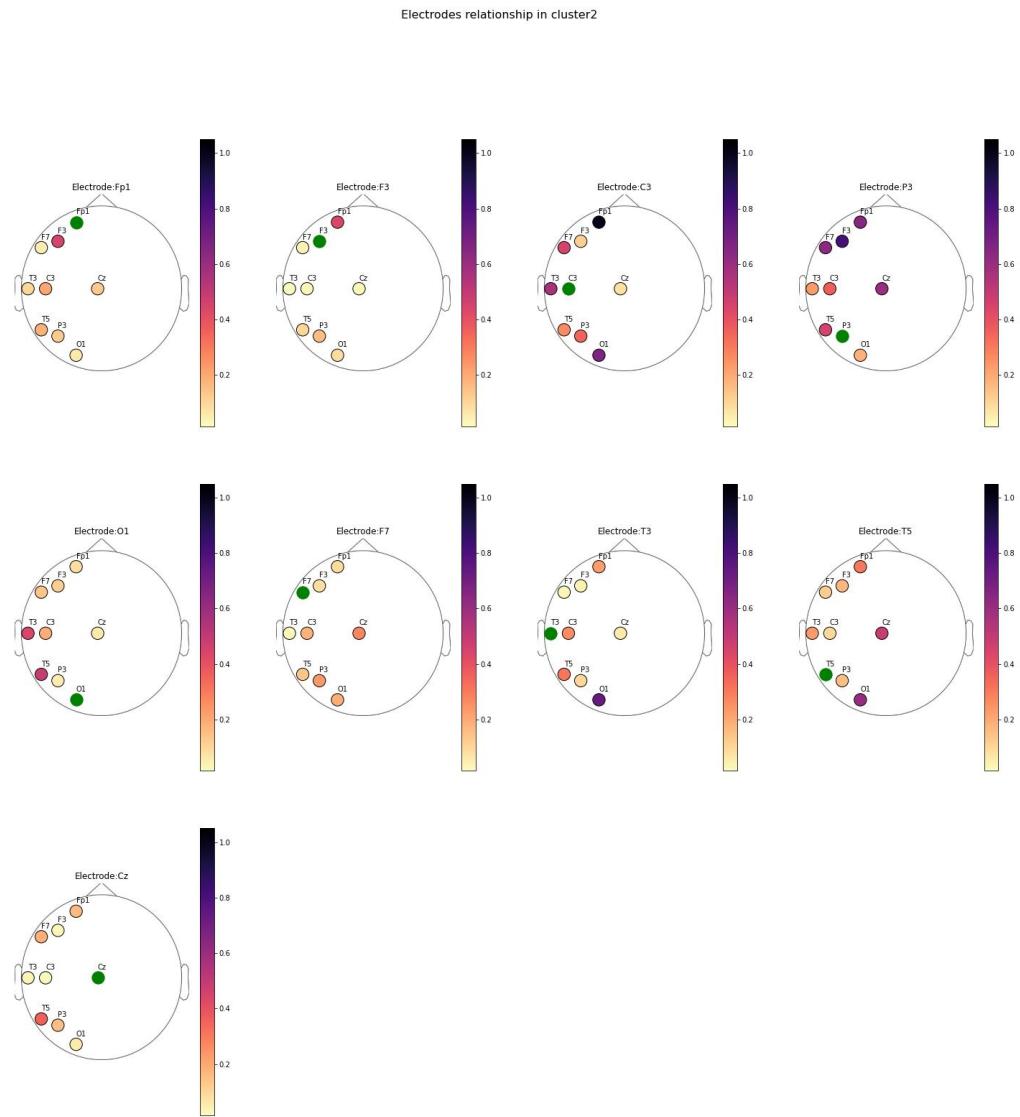


Figure D-4 – Relationships between electrodes of the second cluster obtained with the clustering setting of applying Box-Cox and z-score transformations on the left hemisphere electrodes + Cz.

In Figure D-4, one can see that C3 electrode has a very strong relationship to the Fp1 electrode. On the other hand, Fp1, F7 and Cz do not show a very strong relation to other electrodes.

Overall, we can see that the relationships between electrodes is different among clusters. Further research is needed to find what are the causes of these differences and how they can be explained.