

THE CLUSTER TOOL

Designing a clustering algorithm and graphical user interface for efficient clinical interpretation of Interictal Epileptiform discharges

Feline Louise Spijkerboer

November 2019

UNIVERSITY OF TWENTE.

The Cluster Tool: designing a clustering algorithm and graphical user interface for efficient clinical interpretation of Interictal Epileptiform Discharges

For the title of Master of Science in Technical Medicine

F.L. SPIJKERBOER Bsc.

November 4, 2019

Grudation committee

Chairman: Prof. dr. ir. M.J.A.M van Putten

Medical supervisor: dr. G.H. Visser

Technical supervisor: dr. ir. J. le Feber

Process supervisor: drs. B.J.C.C. Hessink - Sweep

External member: Prof. dr. ir. P.H. Veltink

UNIVERSITY OF TWENTE

Technical Medicine

Faculty of Science and Technology

PO BOX 217

7500 AE Enschede

The Netherlands

Abstract

Objective: Manual detection of interictal epileptiform activity in long-term EEG recordings is time-consuming and highly susceptible to individual interpretation. Automatic detection algorithms offer a faster, reproducible and more objective, and therefore more efficient method for EEG evaluation. These algorithms are able to reach human-like sensitivities. Nevertheless, they are rejected in the clinical practice due to their high false-positive rate and mostly impractical manner of displaying their results. Therefore, the objective of this research is to design a clustering algorithm and graphical user interface, that presents the automatically detected Interictal Epileptiform Discharges according to their morphology and localisation.

Methods: The clustering algorithm was based on the events found by the Persyst P13 spike detector. We divided single lead EEG segments into groups according to their localisation. The events within these groups were then clustered with the K-means algorithm. The Squared Euclidean distance and Dynamic Time Warping distance were considered as distance measures for the clustering. The combination of clustering algorithm and graphical user interface is referred to as the Cluster Tool.

The Cluster Tool was evaluated with usability and clinical performance tests. A total of 23 EEGs was used. The usability tests were performed by five EEG experts, through moderated testing. The clinical performance assessments were done by two test participants. Mutually agreed on clinical conclusions were compared to the clinical conclusion that was described in the EEG report, which was considered the gold standard.

Results: The usage of the tool resulted in remarkably similar clinical diagnoses in comparison to the EEG report. However, the clusters derived by the algorithm did not consistently meet the expectations of the neurologists. This decreased their trust in the performance of the tool and caused them to spend time on manually checking detections within clusters. The use of the Cluster Tool did not speed up the EEG evaluation. The Dynamic Time Warping distance showed a slightly better separation of the cluster results than the SE distance.

Discussion: The Cluster Tool shows the potential of a comprehensive visualisation of interictal epileptiform discharges for improving EEG evaluation. However, the clustering algorithm was too inaccurate to impact the clinical workflow positively. After the implementation of a sufficiently accurate clustering algorithm, the designed prototype promises a faster, reproducible and more objective method for EEG evaluation. The post processing of the output of automatic spike detection algorithms is a step which has been underestimated for too long. Focusing on the visualisation of these detections is an important step toward the clinical implementation of such algorithms.

Preface

When I started my graduation project, I knew little about clustering techniques. Through the scope of this project, I learned a lot about the wide variety of clustering algorithms and their applications. The exploratory nature of this project kept me curious and eager to look for solutions for every hurdle on the road. I enjoyed working together with the clinical field to ensure that we were working towards a product that would be clinically useful.

First of all, I would like to thank Gerhard for showing me the importance of designing clinical innovation in the clinical field. I have learned a lot from you about the field of epilepsy and the struggle of using new diagnostic tools within this field. Your relentless spirit for innovation and clinical implementation of new technologies were a great source of inspiration.

Joost, thank you for your technical support. Whenever I was lost in the world of cluster analysis, your advice and critical questions put me back on track. Although your door was on the other side of the country, it felt like I could always come by to knock on that door for advice.

I would also like to thank Michel for support and guidance. Our interesting discussion and your to-the-point comments helped me to focus on the important aspects of this project.

At the EMU of SEIN, I enjoyed participating in the clinical research team. Elise, thank you for teaching me the beginnings of EEG reading. I know now that there is still much more to learn. I also would like to thank Hannah for all the meetings and discussions. It was nice to have someone to brainstorm with.

To everyone from the clinical research team, thank you for all your intellectual input and of course, the time spent using the Cluster Tool. Without your help, I would have never come this far.

Bregje, thank you for being my process supervisor during the last two year. Your questions made me reflect and observe situations from different perspectives.

Thanks to all my friends for all the fun and warmth in my life. Elsa, Lennart and especially Rens, thank you for proofreading this thesis. Nikki, thanks for your fantastic help with the cover design. And of course, Casper, thank you for being there for me.

Last of all, I want to thank my family for supporting me and teaching me to do what feels right.

Feline Louise Spijkerboer

's-Gravenhage, November 4, 2019

List of abbreviations

AP	Affinity Propagation
BSS	Between cluster Sum of Squares
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
DBA	DTW Barycenter Averaging
DTW	Dynamic Time Warping
ECG	Electrocardiogram
EEG	Electroencephalogram
EMG	Electromyogram
EMU	Epilepsy Monitoring Unit
FPR	False Positive Rate
GUI	Graphical User Interface
IED	Interictal Epileptiform Discharge
LCM	Local Cost Matrix
MVP	Minimum Viable Product
PNES	Psychogenic non-epileptic seizure
POSTS	Positive Occipital Sharp Transients of Sleep
SE	Squared Euclidean
SEIN	Stichting Epilepsie Instellingen Nederland
SSE	Sum of Squared Error
t-SNE	t-Distributed Stochastic Neighbor Embedding

Contents

Preface	i
List of abbreviations	ii
1 Introduction	1
1.1 Motivation	1
1.2 Literature review	2
1.3 Outline of this thesis	3
2 Conceptual framework	5
2.1 Clinical setting and procedures	5
2.1.1 Future vision of the workflow	5
2.2 Persyst P13 spike detector	6
2.2.1 Performance of the P13 spike detector	7
2.3 Theoretical background	8
2.3.1 Interictal Epileptiform Discharges	8
2.3.2 Artefacts	9
2.3.3 Components of the clinical diagnosis epilepsy	10
2.4 Research objective	11
3 Development of the Clustering Algorithm	12
3.1 Introduction	12
3.2 Input and Output	13
3.2.1 Input data	13
3.2.2 Output data	15
3.3 Data Preparation	15
3.4 Distance measure	18
3.4.1 Categories of distance measures	18
3.4.2 Squared Euclidean Distance	19
3.4.3 Dynamic Time Warping distance	20
3.5 Clustering	21
3.5.1 K-means	21
4 Design of the graphical user interface	24
4.1 Introduction	24
4.2 Design	25

4.2.1	Data import, distance measure selection and clustering	26
4.2.2	Visualization of the clusters	28
5	Evaluation of the Cluster Tool	31
5.1	Introduction	31
5.2	Literature review	31
5.2.1	Scalar measurements	32
5.2.2	Case studies	32
5.2.3	Usability tests	33
5.3	Methods	33
5.3.1	Test data	34
5.3.2	Usability testing	34
5.3.3	Performance evaluation	35
5.4	Results	36
5.4.1	Included data	36
5.4.2	Performance evaluation	37
5.4.3	Usability evaluation	39
5.4.4	Distance measures	40
5.4.5	Case study	41
5.5	Discussion	42
6	General Discussion	46
6.1	The cluster algorithm	46
6.1.1	Data preparation	46
6.1.2	Distance measure	47
6.1.3	Clustering algorithm	48
6.1.4	Dimension reduction	49
6.2	The Graphical User Interface	51
6.2.1	Comparison with previous literature	51
6.2.2	Implications for future work	51
7	Conclusion	53
A	Appendix	57
A.1	Gap statistics	57
A.2	Flowchart data preparation	59
A.3	Internal validation	60
A.3.1	Calculation of the SSE and BSS	60
A.3.2	Interpretation of the results	60
A.4	Learn curve of the cluster tool	62
A.5	Table with results clinical conclusions	63

*"Success consists of going
from failure to failure
without loss of enthusiasm".*

WINSTON CHURCHILL

1 | Introduction

1.1 Motivation

Epilepsy is one of the most common neurological diseases worldwide, with around 50 million people diagnosed¹. Epilepsy is a disease of the brain that is characterised by at least one unprovoked epileptic seizure and a high risk of further seizures². The occurrence of seizures is unpredictable, and epilepsy can, therefore, lead to a sudden loss of autonomy. Besides of seizures, epilepsy can cause cognitive and psychological problems. Hence, the disease entails a major burden in seizure-related disability, comorbidities, and costs¹.

Accurately diagnosing epilepsy, including the specific seizure type and seizure onset area, can be challenging. Despite this difficulty, an accurate diagnosis is essential in epilepsy to ensure proper treatment and to avoid false diagnosis and thereby, ineffective treatment. Epilepsy can be diagnosed by examining the patients' history, where especially seizure semiology contains essential information. However, patient history is always subjective and often does not provide enough information to make a certain diagnosis.

The electroencephalogram (EEG) provides supplementary evidence of the clinical suspicion of epilepsy and is the most important technological device in the diagnosis and management of epilepsy. Once an epileptic seizure is recorded on EEG, the diagnosis 'epilepsy' can be confirmed. The EEG during a seizure is also referred to as ictal EEG. Epileptic seizures may occur daily in some patients, but in most cases, weeks, months or even years can pass without the occurrence of a seizure. Hence, it is often not possible to record the ictal EEG.

The interictal EEG is defined as the EEG between seizures. Epileptiform discharges can be seen in the interictal EEG. These Interictal Epileptiform Discharges (IEDs) are often referred to as 'spikes'. The presence of such IEDs in the EEG is a sign for an increased likelihood of seizures and therefore serves as a marker for epilepsy³. This stresses the importance of the EEG as a diagnostic tool.

To achieve an accurate diagnosis, it is very important that EEG evaluation is performed properly, by an experienced EEG reader and interpreted by an experienced physician, in the context of the clinical history⁴. It is common practice that ictal and interictal EEG characteristics are detected manually through visual assessment of the entire EEG recording, which is a very labour intensive process. Moreover, the manual detection of IEDs is inextricably linked to the issue of subjectivity. The inter-reader agreement of experienced EEG readers is remarkably low for interictal spike marking, with values ranging from 39-55%.⁵⁻⁹.

Automatic detection could lead to a more efficient interpretation of IEDs, by enabling faster, reproducible and more accurate evaluation of EEGs. The shift from manual evaluation to automatic detection will speed up the review time and assure that the outcome of the evaluation will always be the same, thereby complying with the reproducibility. It will also enable us to automatically quantify IEDs and gain additional insight into their role within epilepsy diagnosis and treatment⁶. Quantification of IEDs would make it easy to study e.g. whether the amount of interictal discharges is related to the effectiveness of treatment or if certain waveforms are related to specific syndromes. Until now, such studies are mainly performed by counting spikes and manually identifying interictal waveforms, which entails a huge workload¹⁰. Hence, the use of automatic detection will not only lead to more efficient EEG evaluation, but it will also open doors to further research for better understanding of the role of IEDs in epilepsy treatment and diagnosis.

1.2 Literature review

Various automatic detection algorithms for IEDs have been developed since the EEG became digitalised in the 1970s. Despite significant improvements in the field of these algorithms over time, they are not used frequently by clinicians. This is mainly caused by the general impression that automatic detection algorithms perform less than skilled EEG-readers^{5,8}. There is a substantial lack of agreement on the generally accepted determination of epileptic activity, which limits the definition of a proper gold-standard. Therefore, performance assessment of these algorithms remains challenging, and the major part of clinicians keeps questioning the reliability of automatic spike detection algorithms.

Despite the general lack of acceptance from the clinical field, several automatic spike detection software are commercially available, offering their added value as a tool for more efficient spike detection. Recently, a study by Scheuer et al. (2017) presented an algorithm, the Persyst P13 (©Persys Development Corporation 2016), which showed human-level performance for epileptogenic spike detection⁵. This shows that automatic detection software can be used as a more objective and efficient tool for EEG evaluation in clinical practice, without handing in on the quality of review performance. A clinical assessment by Halford et al. (2018) confirmed that the P13 algorithm has good sensitivity performance based on a pairwise comparison with 35 EEG readers. However, they also raised their concerns about the high rate of false-positives of the spike detector and state that the algorithm is not ready to use in a clinical setting¹¹.

The false-positive rate (FPR) of automatic spike detection algorithms has always been an obstacle for the clinical implementation of these algorithms. In 2002, Wilson recalled the fundamental issue, which was already described by Frost in 1985, that automated spike detection algorithms have high FPRs⁶. More than 30 years later, Halford et al. (2018) show that automatic spike detection algorithms still have FPRs, which are considered unacceptably high¹¹. It seems like researchers have been entangled in a battle-of-algorithms for the last 50 years, where everyone keeps searching for the perfect spike detection algorithm, but none seems to find it. The high sensitivity of the detection algorithm is crucial to detect all important events, but it must be considered what level of sensitivity is realistic and sufficient. Currently, the sensitivity of IED detection by humans, which can be considered the gold standard, varies between 39% and 70%⁵. Scheuer et al. (2017) showed that their P13 detection algorithm performed human-

like, with a sensitivity of 43.9%. However, this level of sensitivity comes with a false-positive rate (FPR) of 1.65 per minute⁵. Regarding the state-of-the-art of IED validation, the best possible sensitivity which can be achieved is the level of human-like-performance. Therefore, it is about time to stop focusing on improving detection algorithms and start searching for methods to deal with the high FPRs and look for ways to enable clinical implementation of those algorithms.

A possible solution to deal with false-positive detections is given by Wilson et al. in 1999, who states that it is sufficient when a user interface allows the neurologist “to quickly delete artefacts and determine whether there are multiple spike generators [...]”¹². Automatic detection algorithms are not designed to replace the work of a neurologist, but rather to assist in EEG evaluation as a decision support system. Therefore, an automatic detection algorithm must serve as a tool which enables the neurologist to review the IEDs in a quick and easy manner so that it can be decided which detections are truly epileptiform and which are false-positive detections.

Wilson et al. (1999) proposed to combine all detected events with nearly similar morphology and topology into one event through clustering¹². This way, the results of automatic detection are summarised and presented more comprehensively. They report an increase in reviewing speed and the opportunity of immediate identification of multiple detections at once when using the spike clusters. This shows the potential of cluster techniques to group the detected events according to their waveform, and thereby separating the false detections from the IEDs. Such a clustering method would enable us to create a comprehensive overview of the automatically detected spikes, and facilitate more efficient clinical interpretation of the results, without wasting time on false positive detections.

1.3 Outline of this thesis

This research aims to study if the clustering of automatically detected IEDs according to their localisation and morphology and their presentation in a comprehensive user interface, will enable more efficient clinical interpretation and facilitate the implementation of automatic detection algorithms in the clinical field.

Chapter 2 provides a detailed context analysis, which results in a clear problem definition and study aim. It describes the clinical setting for which the clustering algorithm and user interface were designed and introduces the automatic detection software used for this research; the Persyst P13 spike detector. The chapter also includes some background information about IEDs and common EEG artefacts. It concludes with a detailed problem description which focuses the problem stated in this introduction on a specific clinical setting while using the P13 spike detector.

Chapter 3 presents the clustering algorithm designed to partition the events detected by the P13 spike detector. It briefly introduces time series clustering and describes the different steps which are applied in the clustering algorithm, being: data preparation, calculation of a distance measure and clustering.

The design of the graphical user interface (GUI) is presented in **chapter 4**. The GUI facilitates the interaction between the cluster results and the user. The architecture of the

GUI and graphical layout are presented in this chapter.

The clustering algorithm and GUI together are referred to as the Cluster Tool. The performance and usability of the Cluster Tool are evaluated and discussed in **chapter 5**.

Chapter 6 provides a general discussion on the Cluster Tool, where the methods applied in this research are reviewed, and recommendations are suggested based on the results of the validation.

This thesis finalises with the conclusion presented in **chapter 7**.

2 | Conceptual framework

2.1 Clinical setting and procedures

Stichting Epilepsie Instellingen Nederland (SEIN) in the Netherlands, is a tertiary expertise centre for epilepsy and sleep medicine. At the Epilepsy Monitoring Unit (EMU) of SEIN, eight rooms are available for patient intake, where the long-term EEG recordings are performed under continuous monitoring of video, audio and co-registration of the Electrocardiogram (ECG) and on occasion Electromyography (EMG). Each room has four rotatable cameras, which are controlled and observed around the clock by nurses. These nurses also assist the patient and perform cognitive tests on the patient when a seizure occurs. This way, the EEG, ECG, video and audio of the patients are recorded during the entire intake. Two of the rooms are used for pre-surgical admissions. These patients come in on Monday and stay for five days. The other six rooms are used for 24- and 48-hour recordings. This means that during a full week over 800 hours of EEG recordings are registered.

Currently, all EEG recordings are analysed manually. Figure 2.1 shows a screenshot of one EEG page, which typically includes 15 seconds of EEG recording. The visual evaluation of the entire EEG is performed by the EEG technicians. They start with the evaluation of the background pattern and the diagnostic tests, of which they make a representing selection. The inspection of the rest of the registration is done by scrolling chronologically through the EEG. All pieces of EEG containing abnormal and suspicious events are marked. It can sometimes be difficult to distinguish abnormal activity from regular activity or artefacts. Many kinds of artefacts can occur in EEG recordings. The definition of these artefacts is explained more in detail in section 2.3.2. In cases when it is difficult to interpret EEG phenomena, the video recording provides additional information and is used by the technician to decide which parts to mark.

Once the entire EEG has been evaluated and annotated by the technician, the neurologist will look into the registration and review the representative selections and the annotated parts. Based on these selections, the neurologist will form a conclusion and recommendation according to the clinical question of the outpatient physician who had referred the patient to SEIN.

2.1.1 Future vision of the workflow

Instead of analysing the entire EEG recording manually, the experts at SEIN want to start using automatic spike detection in addition to the visual evaluation of a small selection of the EEG. Current spike detection algorithms are not capable of detecting IEDs with high specificity. Nevertheless, the experts at SEIN believe that it is about time that automatic

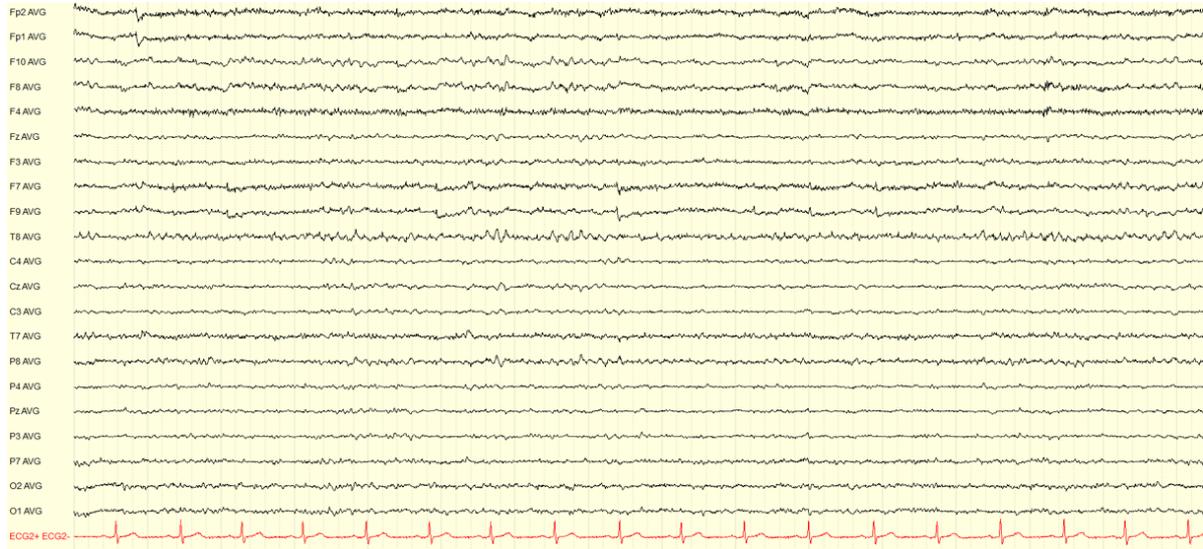


Figure 2.1: EEG recording shown for 15s in the average reference montage. The EEG was recorded according to the 10-20 system with additional F9 and F10 electrodes.

detection will be implemented within the clinical field. Their vision of future EEG evaluation is to review only one hour of the wake EEG, including the diagnostic test, the first hour of sleep and the first half-hour after waking. This should provide enough information to get a good impression of the background activity and general EEG of the patient. The rest of the EEG should be analysed by automatic detection software, as shown in Figure 2.2. Note that for a completely automated evaluation, a reliable seizure detection and trend analysis must also be used. However, the current study will only focus on the automatic detection of IEDs.

The implementation of an automatic detection algorithm at SEIN would directly help to optimise the diagnostic process and thereby increase the quality of patient care. Experts at SEIN are willing to use semiautomatic spike detection software in the clinical practice and are actively testing available software. The current project is part of this research field.

2.2 Persyst P13 spike detector

The P13 spike detector, as is presented by Scheuer et al. (2017) uses EEG recordings in the common average reference montage to detect focal IEDs⁵. Generalised discharges are detected in another referential montage, which uses either the two frontopolar electrodes (Fp1 and Fp2), the temporal electrodes (T7 and T8) or the occipital electrodes (O1 and O2)⁵.

The morphology of the detection is described by dividing the waveform into six-half waves. The algorithm uses features of each half-wave, containing information about the amplitude, duration and curvature. The two waves in the middle represent the deflection of the spike. The two waves at the beginning describe the EEG activity preceding the spike, and the two waves at the end describe whether a slow component follows the spike. Whenever a similar detection is found around the same time, but on a different channel, the spike will be detected on the channel with the highest amplitude only⁵.

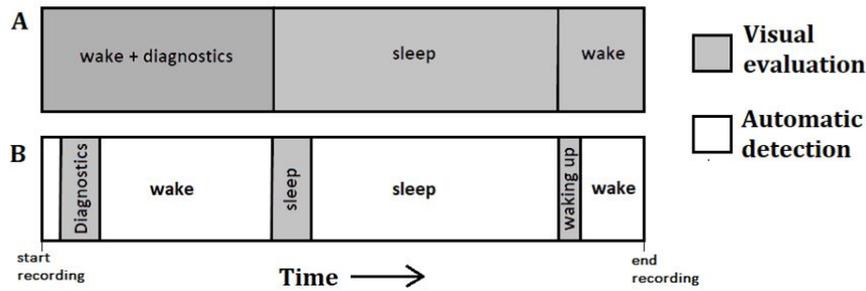


Figure 2.2: Schematic overview of the workflow of EEG evaluation. **a)** shows the current workflow of EEG evaluation used in SEIN, where the entire EEG is reviewed manually. **b)** presents an overview of the workflow of EEG evaluation SEIN wants to achieve in the future. Only one hour of wake, which includes the diagnostics tests, the first hour of sleep and the half hour after waking up are reviewed manually. The rest of the EEG is evaluated by automatic detection software.

All features describing the morphology, localisation and context of the detection are used in a set of neural networks to create a likelihood score for the event to be truly an IED. This results in every detection being assigned a perception value between 1 and 0. A value of 1 represents a very high likelihood, and a value of 0 represents that it is very unlikely for the event to be an IED. Whenever an event is uncertain, it is assigned a perception value near 0.5⁵.

2.2.1 Performance of the P13 spike detector

An internal study at SEIN compared three commercially available software packages for automatic spike detection and showed that the Persyst P13 outperformed the spike detection software of AIT Encevis and BESA Epilepsy 2.0¹³. The study revealed that the P13 indeed performed equal to the human reviewers, as is also stated by other studies^{5,11}. The performance of the Persys P13 was evaluated by comparing the clinical conclusion based on the software, with the clinical conclusion as described in the EEG report. Each event was categorised based on its importance as either high, medium or low. Events with high importance had a direct impact on the clinical diagnosis, medium important event supported a diagnosis and events with low importance only gave vague information about waveforms present in the EEG, without influencing the clinical diagnosis. Figure 2.3 shows the results of this comparison and it can be observed that by using Persyst a large part of the events were detected.

Although the performance of Persyst P13 was similar to human performance, the way the results were presented was experienced as limited. The current user interface of the Persyst spike review presents the detections to the user based on electrode location, as can be seen in Figure 2.4. This poor way of presenting the detections caused that certain waveforms got lost in the list of detections and were overlooked. This had a negative impact on the clinical conclusion and is the reason that not all events of high and medium importance were found in the internal study at SEIN¹³. It was also noted that a relatively large number of the detected spikes were false. This high FPR was also described by Halford et al. (2018) who found mean

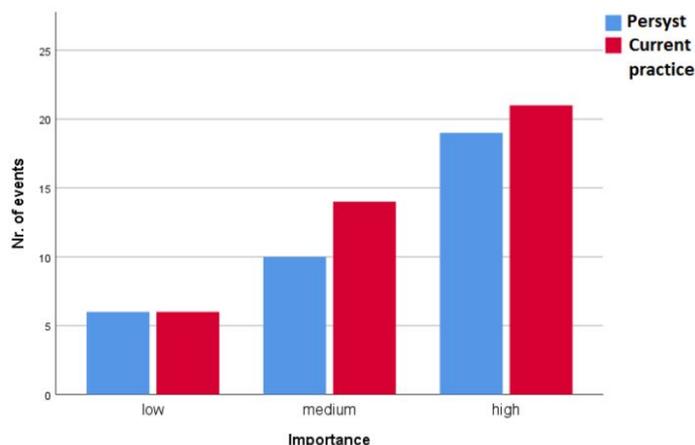


Figure 2.3: Number of events detected when using Persyst P13, compared to the current practice. The events are divided by importance. Events with high importance had a direct impact on the clinical diagnosis, medium important event supported a diagnosis and events with low importance only gave vague information about waveforms present in the EEG, without influencing the clinical diagnosis. Persyst P13 performed similar to the current practice on all levels of importance. Figure adapted from ‘*A practical comparison of automatic detection software for interictal spikes in long-term EEG recordings at SEIN*’ by Spijkerboer, F. L. (2018).¹³

pairwise false positive rate of 1.2 per minute, when applying the 0.9 perception threshold¹¹. The fear of overlooking important events resulted in a workflow where all detected events were inspected and classified manually. Hence, the workload of long-term EEG review was not found to be mitigated when using Persyst P13. It was therefore concluded that the software is not ready for implementation in the clinical workflow yet, which agrees with the conclusion of Halford et al. (2018)¹¹.

2.3 Theoretical background

2.3.1 Interictal Epileptiform Discharges

Identifying Interictal Epileptiform Discharges (IEDs) and differentiating them from normal variant can be difficult. In the revised glossary of terms most commonly used by clinical electroencephalographers, Kane et al. (2017) defined IEDs as transient which is distinguishable from the background activity with characteristic morphology¹⁴. They contain a sharp or spiky aspect and a wave duration which is either shorter or longer than the ongoing background activity. The transient disrupts the background activity surrounding the epileptiform discharge and can be followed by a slow wave. Different kind of IEDs can be distinguished, as presented in Table 2.1. It may seem like clear definitions for IEDs exists, but in reality, it can be difficult to distinguish different IED morphologies. Interictal discharges can vary significantly between patients, even if both waveforms would be classified as the same type of IED. Within a patient though, IEDs tend to be morphologically very similar¹⁵.

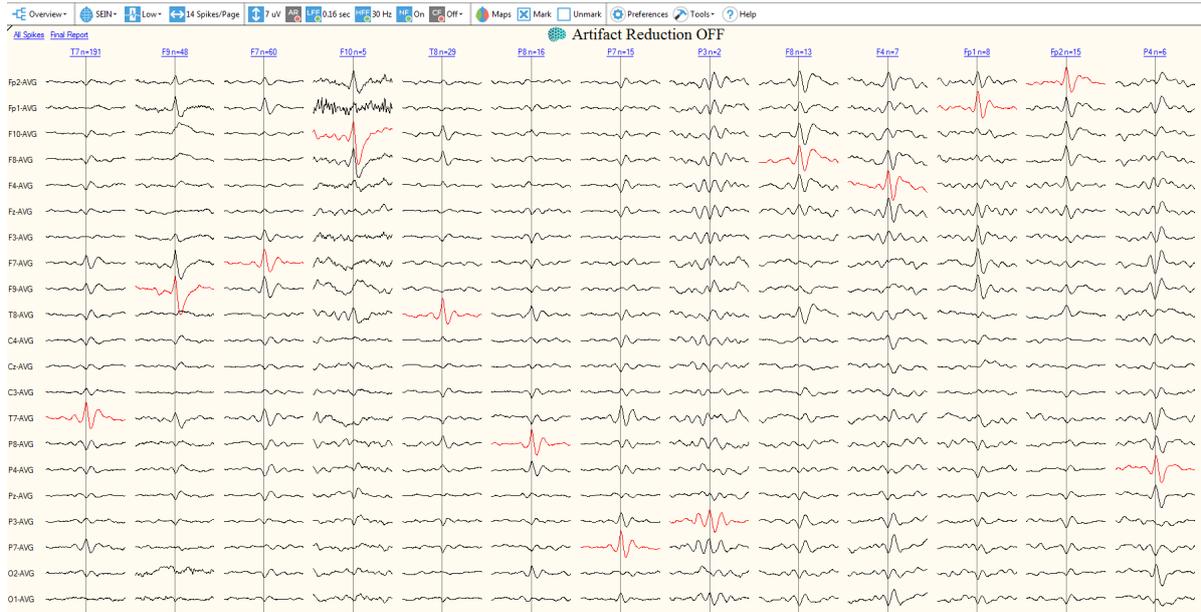


Figure 2.4: Overview of the user interface of Persyst P13 spike review. The events are grouped per electrode location, which are shown in the top row above the EEG segments. Below each electrode group a very short segment of the average waveform of all detections present in group are presented in the average reference montage.

It is important to realise that many epileptiform like patterns exist, which do not support the diagnosis epilepsy. In patients with non-epileptic disorders, such as psychogenic non-epileptic spells (PNES) and syncope, misreading epileptiform like patters have caused incorrect diagnosis many times. Studies have shown that approximately 30% of adult patients which are referred for intractable epilepsy have non-epileptic events¹⁶. Distinctive physiological waveforms like vertex waves, lambda waves, positive occipital sharp transients of sleep (POSTS), or sharp transients which are poorly distinguished from background activity, such as 6Hz spike-and-slow-waves, are not considered epileptiform. Generalised paroxysmal fast activity or wicket spikes are also examples of epileptiform like patterns that are frequently confused with IEDs¹⁶. This shows the difficulty of accurately detecting IEDs.

2.3.2 Artefacts

Many kinds of artefacts can occur in EEG recordings, some of which might be mistaken for sharp epileptic activity. When these artefacts originate from electrical activity from other body parts, they are called biological artefacts. Eye blinks produce high amplitude signals over the frontal electrodes, and lateral eye movements produce sharp positive signals on the left or right frontal electrodes, depending on the direction of the eye movement. Muscle tension, originating from chewing, tongue movement, or swallowing result in spike trains, which shape and amplitude depend on the degree of the muscle contraction. Other artefacts can result from, for example, cardiac activity, poor electrode contact, the 50 Hz transmission line, or physical movement of the patient. The morphology of these artefacts can be mistaken for

Table 2.1: Definition of different types of IEDs. The definitions adapted from ‘*A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings. Revision 2017.*’ By Kane et al. (2017). In *Clinical Neurophysiology Practice*.¹⁴

Polyspike complex	A sequence of two or more spikes. This waveform can be epileptiform but can also be confused with generalised paroxysmal fast activity or wicket spikes.
Polyspike-and-slow-wave complex	An epileptiform pattern consisting of two or more spikes associated with one or more slow waves
Sharp wave	An epileptiform transient clearly distinguished from the background activity, although amplitude varies. A pointed peak at a conventional time scale and duration of 70–200 ms, usually with a steeper ascending phase when compared to the descending phase. The main component is generally negative relative to other areas and may be followed by a slow wave of the same polarity. Comment: Sharp waves should be differentiated from spikes, i.e. transients having similar characteristics but shorter duration. However, it should be kept in mind that this distinction is largely arbitrary and primarily serves descriptive purposes.
Sharp-and-slow-wave complex	An epileptiform pattern consisting of a sharp wave and an associated following slow-wave, clearly distinguished from background activity. May be single or multiple.
Spike	A transient, clearly distinguished from background activity, with a pointed peak at a conventional time scale and duration from 20 to less than 70 ms. Amplitude varies but typically >50 μ V. The main component is generally negative relative to other areas. Comments: 1. term should be restricted to epileptiform discharges. EEG spikes should be differentiated from sharp waves, i.e. transients having similar characteristics but longer durations. However, it should be kept in mind that this distinction is largely arbitrary and primarily serves descriptive purposes. 2. EEG spikes should be clearly distinguished from the brief unit spikes recorded from single cells with microelectrode techniques
Spike-and-slow-wave complex	An epileptiform pattern consisting of a spike and an associated following slow-wave, clearly distinguished from background activity. May be single or multiple

epileptiform activity and lead to false interpretation. Due to the wide variety of morphologies of IEDs, the similarity of IEDs with physiological epileptiform like patterns, and the presence of artifact within the EEG, the detection of IEDs is difficult¹⁷. This not only entails difficulties for the diagnosis of epilepsy but makes it also difficult to design and validate algorithms for automatic detection of IEDs.

2.3.3 Components of the clinical diagnosis epilepsy

To get to a comprehensive and clinically usable overview of all events which are detected by Persyst P13, it must be considered carefully what information should be presented in the

comprehensive overview. Three crucial components can be distinguished, by which epileptiform events are described and interpreted in the current clinical practice; wave morphology, temporal occurrence, and localization. A neurologist would write a conclusion in the clinical report formulated like ” *The EEG showed occasional poly spikes with maximum right fronto- temporal*” or ” *The EEG is often interrupted by clusters of high-amplitude, bioccipital, sharp and slow waves*”⁴. Therefore, the comprehensive overview should provide information on the wave morphology, temporal occurrence and localisation of the detected events.

2.4 Research objective

This research was executed at the EMU of SEIN, and the IEDs were detected by the Persyst P13 spike detector. The general objective is to design a clustering algorithm to group automatically detected IEDs according to their localisation and morphology and present those clusters in a comprehensive overview, to enable efficient clinical interpretation of long-term EEG recordings. This is realized through the following specific research objectives:

- Develop a cluster algorithm which groups all events detected by Persyst P13 according to their morphology and localization
- Design a Graphical User Interface (GUI) which presents the results of the clustering by their morphology, localization and temporal occurrence
- Evaluation of the clustering and GUI to assess the impact of using a comprehensive overview, on the clinical interpretation of long-term EEG recordings.

3 | Development of the Clustering Algorithm

3.1 Introduction

This chapter describes the algorithm, developed to cluster the events which are detected by the Persyst P13 spike detector, according to their morphology and localisation. Clustering is a form of unsupervised classification where groups are created in a way that objects within a cluster are similar, and objects belonging to different clusters are not similar. It is not known in advance what the groups will look like and no label is assigned to the groups or clusters. The process of clustering can be divided into four steps:

1. Data preparation

The preparation step determines the structure of the clusters. This may include the data size, data selection and pre-processing steps. When using a feature-based approach, the selection of the features is also included in this step.

2. Definition of the distance measure

This is often considered as the most important step of the entire clustering process¹⁸. The distance measure quantifies the degree of dissimilarity between two or more time-series, in a way that it can be used as a criterion for creating clusters. Care should be taken when choosing a distance measure because a proper criterion for dissimilarity is based on the characteristics of the time-series, the representation method of the data, and the objective of the clustering¹⁹.

3. Clustering

The clustering algorithm uses the set of distance measures as input to create clusters based on the characteristics of the algorithm. Many different types of clustering exist, and they can serve in many different applications. The choice for a clustering algorithm depends on the application, the type of clustering desired and the type of input data.

4. Validation of the clustering

Cluster evaluation is not a well-developed, though an important part of cluster analysis²⁰. Due to its very nature, the definition of good clustering can be troublesome, and different type of cluster algorithms require different kinds of evaluation measures. The selection of a validation method should always be made in the context of the data type and objective of the clustering.

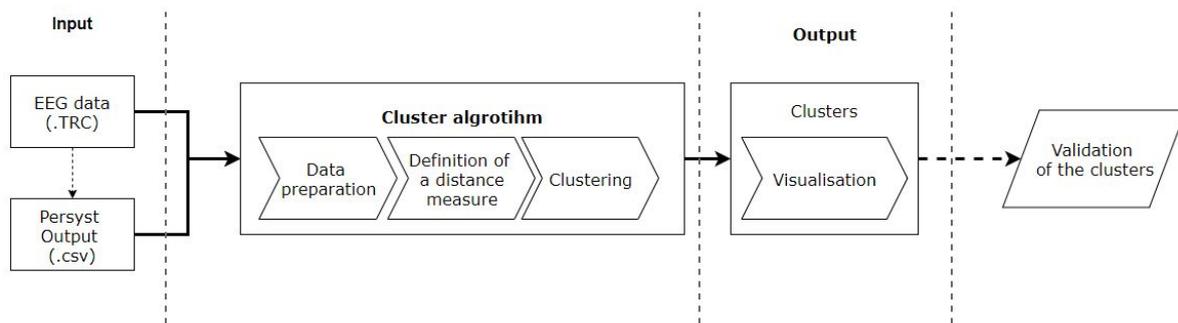


Figure 3.1: Flowchart of the clustering process. It starts on the left with the input data, which is applied to the clustering algorithm. The input data consists of an EEG file, and a file containing the output of the Persyst P13 spike detector. Both input files are patient-specific. The algorithm consists of three steps: the data preparation, the definition of a distance measure and clustering. The output consists of clusters, which require a visualisation step to be inspected. The visualised clusters can then be validated, as the last step of the clustering process.

The first three steps of the clustering process define the performance of a clustering algorithm. The following sections will present the algorithm developed in this project. First, the in- and output data are presented. Subsequently, the methods applied for data preparation, calculation of the distance measure and clustering are described. The last step of the clustering process is the validation of the results. This will be discussed in chapter 5.

3.2 Input and Output

3.2.1 Input data

A flowchart of the in- and output data is shown in Figure 3.1. The clustering algorithm depended on two input files, which were patient-specific. These were the EEG recording and the output of the spike detection software Persyst P13. The latter was used to select segments in the EEG recording which contained a spike. These segments were selected based on the time where Persyst P13 marked a detection. This section analyses the types of input data which is important because it is essential to choose proper methods for data representation, the calculation of the distance measure and the clustering algorithm.

EEG data

The algorithm was based on EEG files which were stored as .TRC file. EEG recordings are a type of time series data. Time-series clustering is a special type of clustering because its feature values change as a function of time. Time series data is stored with multiple entries per second and is therefore naturally high dimensional and often large in data size¹⁹. Dimensionality in this context is defined by the number of samples and is represented by the length of the time series.

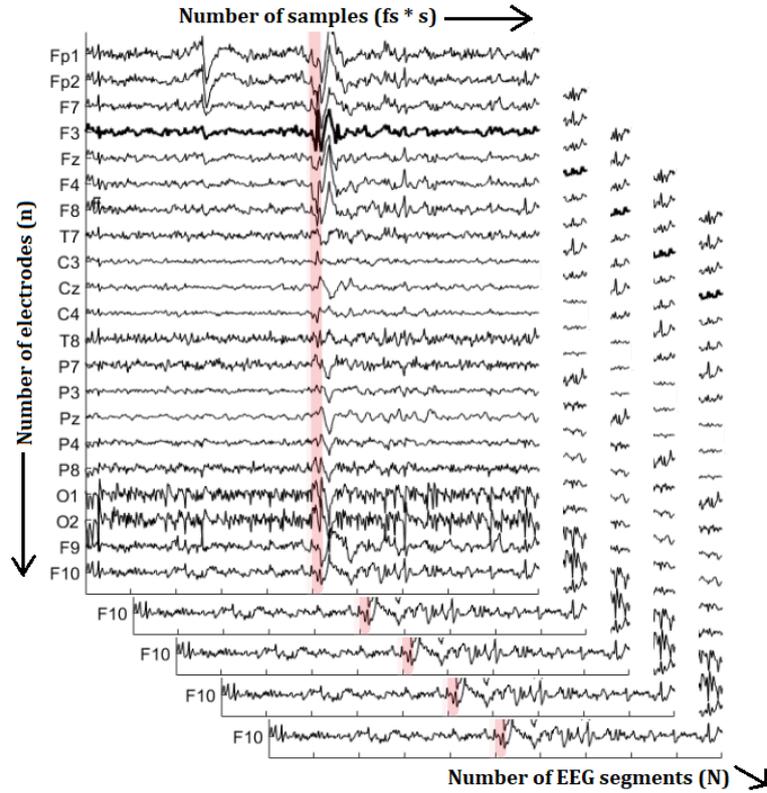


Figure 3.2: Visualisation of the dimensionality of EEG data. The horizontal axis represents the number of samples, which consists of the sample frequency fs times the duration s in seconds. The vertical axis shows the number of electrodes n , which is 21 electrodes for the EEG shown. In depth, the number of EEG segments N are presented.

EEG data is not only multidimensional, but it also consists of several recordings on the same time scale, recorded by multiple electrodes. This makes the data multivariate. When we use a sample frequency fs and select EEG segments with a duration s , we get a time series with a length of $fs \times s$ samples. Considering an EEG recording on n different electrodes, one EEG segment would already consist of a $n \times fs \times s$ matrix. Figure 3.2 shows an example of the input data with N EEG segments. The high dimensionality of the multivariate EEG data limits the choice of clustering algorithms, and a large data size slows down the computational time of the algorithm.

Persyst P13 output

Automatic spike detection was done for each EEG. The Persyst P13 spike detection software performed the automatic detection. An overview of this software is presented in Section 2.2. The output of the detection algorithm was extracted to a comma-separated value (csv) file, which stored three important variables: the exact timestamp of all N detections, the

Detection Time	Perception Value	Channel
d1 14:14:49.355	0.60	F7-Av12
d1 14:14:59.640	0.55	F9-Av12
d1 14:20:35.362	0.44	Fp2-Av12
d1 14:24:49.702	0.62	F10-Av12
d1 16:06:07.617	0.97	F7-Av12
d1 16:06:56.672	0.57	F9-Av12
d1 16:07:49.072	0.41	F9-Av12
d1 21:40:25.972	0.90	T7-Av12
d1 21:55:14.997	0.42	F9-Av12
d1 22:11:16.772	0.44	O2-Fp12
d2 00:29:07.653	0.53	C3-Av12

Table 3.1: Overview of the data present in the output file of the Persyst P13 spike detection algorithm. The first column includes the exact time stamp when the IED was detected, the second column shows the perception value of the detection, and the third column presents the electrode channel where the highest amplitude was detected as well as to which reference electrode the IED was detected.

electrodeposition at which each detection had the highest amplitude, and the perception value of each detection. The perception value is a measure, introduced by Persyst, to indicate the likelihood of a detection to truly be epileptiform, where a higher value represents a higher likelihood. An example of an output file from the P13 spike detector is presented in Table 3.1. The output file of the spike detection and the EEG file were both stored under the same name so that for a certain patient *p001* an EEG file *p001.TRC* and corresponding spike detection output file *p001.csv* existed.

3.2.2 Output data

The output of the algorithm consisted of several clusters that divide the set of EEG segments into groups, based on the morphology and localisation of the EEG waveform. The visualisation of these clusters is discussed in the next chapter. Figure 3.1 shows the in- and output data in regards to the steps of the clustering process.

3.3 Data Preparation

To deal with the multivariate high dimensional input data, we applied several selection steps. An important aspect to consider when clustering time series, is the possible presence of noise, shifts, artefacts, discontinuities and temporal drift. The data from the Persyst output file was used to select short segments of the EEG around the exact detection time. The duration of these EEG segments was four seconds, ranging from two seconds before the detection time until two seconds after, as shown in Figure 3.3. We chose a range of four seconds because it is important for clinical evaluation of an IED, to see the surrounding EEG and get an impression of the background activity.

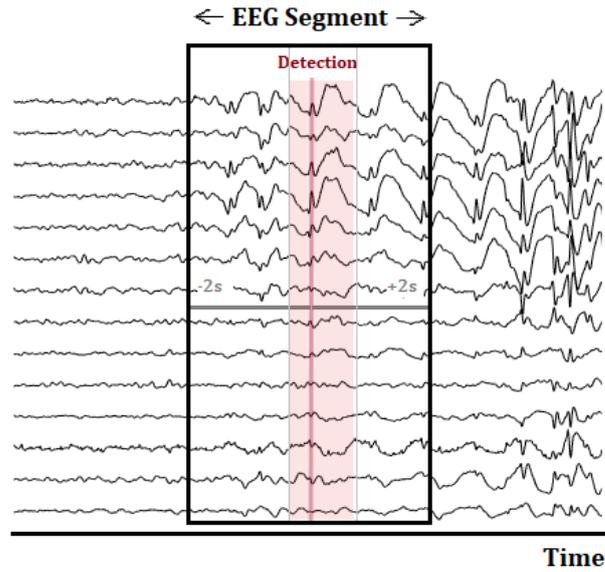


Figure 3.3: Selection of an EEG segment with a total duration of four seconds, used for visualisation. The zero-time, marked in the figure by the red line, corresponds to the exact timestamp which was detected by Persyst P13. The lighter red area represents the segment which ranges from 200ms before the detection time to 500ms after the detection time. This smaller EEG segment is used for calculation of the distance measure and clustering.

The EEG segments were loaded into MATLAB (R2019, MathWorks Inc.). The Fieldtrip Toolbox was used for preprocessing and visualisation of the EEG segments. Fieldtrip is a free MATLAB toolbox for EEG analysis. All EEG segments were re-referenced to the average reference montage. A highpass filter, with a cutoff frequency of 2 Hz and a Hanning window was applied on the EEG segments to get rid of low-frequency drifts and to taper off the EEG segments towards the ends. The hereby created EEG segments with a duration of four seconds were used for the visualisation of the surrounding EEG. For clustering, the EEG segments were further narrowed to an interval of 200ms before the exact detection time and 500ms after (see Figure 3.3). By narrowing the EEG segment, we decreased the possible amount of background activity present in the signal and thereby ensured that the clustering was done mainly based on the EEG waveform, and less on the background activity.

Subsequently, we divided the EEG segments further into groups based on the perception value. All EEG segments which corresponded to a perception value of 0.9 or higher were put in one group and the EEG segments corresponding to a perception value of 0.4 or higher in another group. The EEG segments with a perception value lower than 0.4 were ignored since preliminary studies revealed that these events did not contain events significant for the clinical diagnosis. All steps of this selection process are shown in the flowchart in Figure A.2 in Appendix A.2. Note that the group with the medium perception value (0.4 threshold) also contained all EEG segments which were also included in the high perception value group (0.9 thresholds).

Brain region	Included Channels
Frontal Left	Fp1, F3
Frontal Right	Fp2, F4
Frontolateral Left	F7, F9, T7
Frontolateral Right	F8, F10, T8
Centroparietal Left	C3, P3
Centroparietal Right	C4, P4
Parieto-occipital Left	P7, O1
Parieto-occipital Right	P8, O2
Midline	Fz, Cz, Pz
Other groups	
Generalised spikes	*
Residuals	**

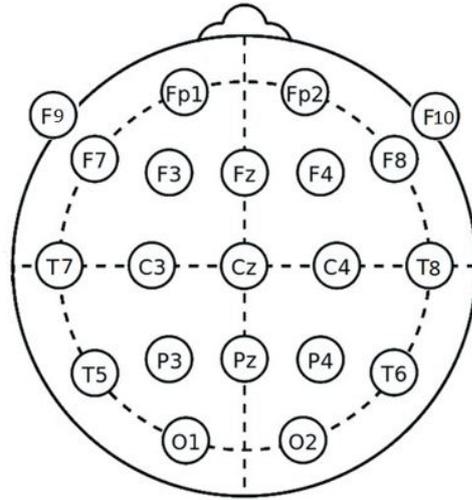


Figure 3.4: Definition of the Brain Regions used in the cluster algorithm **Left:** Overview of the Brain Regions and the electrodes that are included in each of them. **Right:** Electrode placement according to the 10-20 system with additional F9 and F10 electrodes.

* The EEG segments included in this group are detected by Persyst P13 as generalized and are therefore not found on a specific electrode.

** The EEG segments included in the residuals can results from all electrodes

The pre-selected IEDs were then further divided according to their localisation. We defined a total of nine brain regions to describe the localisation of the IED; Frontal, Frontotemporal, Centroparietal and Parieto-occipital, all separated in the right and left hemisphere, and the midline. A detailed overview of the definition of the brain regions and the corresponding electrodes is presented in Figure 3.4. This figure also shows the scalp position of the electrodes. Not all events were assigned to one of the brain regions. Some events were marked as generalised by Persyst, meaning that they did not have a specific source but arose from activity all over the brain. Therefore, these events could not be assigned to a certain brain region and were therefore assigned to the ‘generalised’ group.

Only the channel where Persyst detected the spike was selected, in order to deal with the multivariate data. This resulted in EEG segments which consisted of a $1 \times fs \times s$ array, thereby reducing the number of simultaneously recorded samples to one (see Figure 3.5). The events included in the ‘generalised’ group were not found on a single electrode and removing leads would delete important information of the generalised IED. Hence, the entries of the generalised group were not clustered.

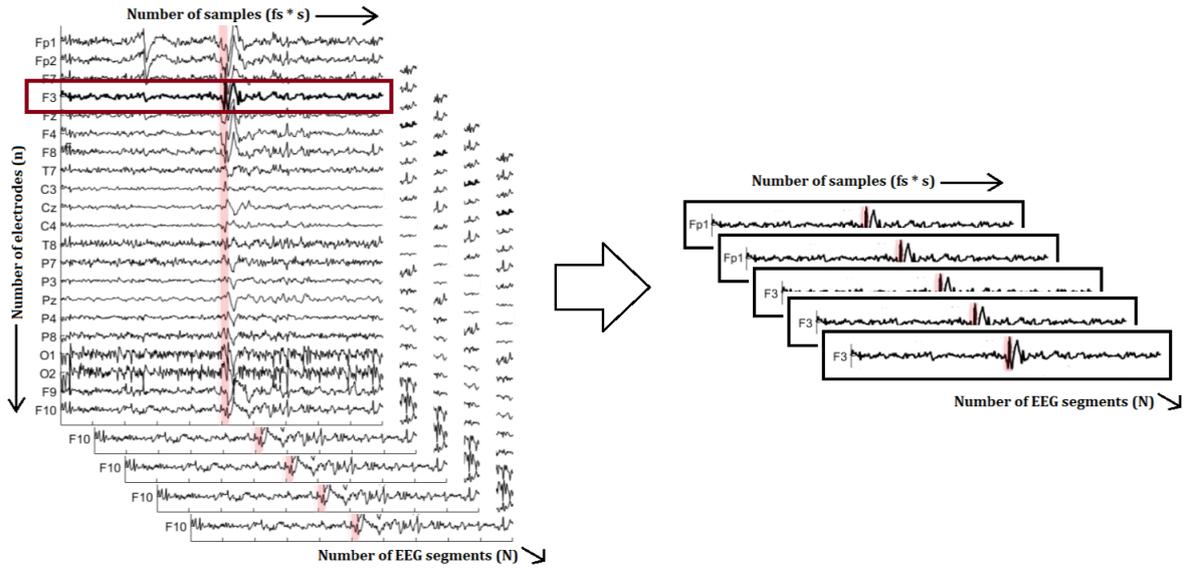


Figure 3.5: Dimension reduction of the EEG data by selecting only the channel where the spike was detected by Persyst. The new EEG segment is no longer multivariate, but exists of a $1 \times fs \times s$ time series.

3.4 Distance measure

To determine whether time series were similar, we had to define a function to measure similarity. This so-called distance measure could then be used to quantify the degree of dissimilarity between two or more time-series. Note that similarity and distance are inverse concepts. Finding a proper distance measure is one of the most important steps of the clustering process since it directly influences the shape of the clusters²¹. Humans are very good at visually recognising patterns and determining similarity, but programming an algorithm to perform the same is a difficult problem²². Moreover, time series can be noisy, contain outliers and shifts, and suffer from discontinuities and temporal drifts¹⁹. Therefore, the choice for a distance measure should be well considered.

Notation

We use the notation $D(X_i, Y_j)$ to represent the distance between two EEG segments $X = (x_1, x_2, \dots, x_i)$ and $Y = (y_1, y_2, \dots, y_j)$, where $X \in \mathcal{R}$ and $Y \in \mathcal{R}$. Note that i and j represent the length of the EEG segments $t = fs \times s$, with t the number of samples, fs the sample frequency and s the duration of the time series in seconds.

3.4.1 Categories of distance measures

Aghabozorgi et al. (2015) reviewed that three different ways of time series clustering can be defined; shape-based, feature-based and model-based. Feature-based measures require a selection of features from the data that describe the actual time series. They are often applied to obtain a reduction in both dimensionality and noise. The model-based methods first fit

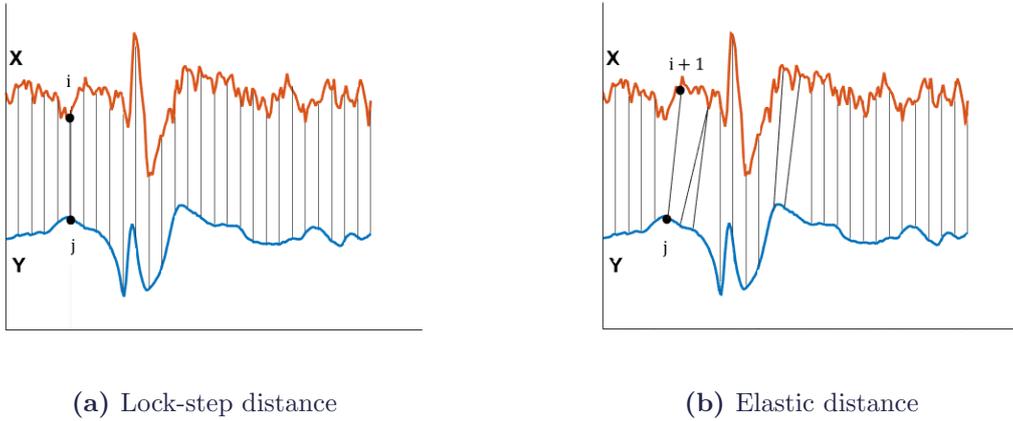


Figure 3.6: Comparison of two time series made with lock-step and elastic measures, respectively. **a.** Example of a lock step measure, where sample i will always be compared with sample $j = i$. **b.** Elastic measure, where sample i can be compared with sample $j = i + x$.

a model to the time series and subsequently compare the parameters of the hereby created models. Shape-based distance measures compare a pair of time series directly, based on their raw data. Literature study reveals that feature and shape-based methods are most common in time series clustering^{18,19}. Esling and Agon (2012) state that shape-based methods are most appropriate when the time series are relatively short and visual evaluation can be used for interpretation of the results²³. The EEG segments at hand correspond to short time series. Therefore, a shape-based approach was considered most likely to provide the best results.

Shape-based distance measures can be divided into two categories; the lock-step measures and the elastic measures. Figure 3.6 presents a comparison of two time series made with lock-step and elastic measures, respectively. Lock-step measures always compare the i th sample of time series X to the j th sample of time series Y , with $i = j$. Elastic measures methods take into account the surrounding points in time to allow for shifts in time, so that the i th sample of time series X can be compared to the j th sample of time series Y , with $i \neq j$. We used the Squared Euclidean distance as lock-step measure, and the Dynamic Time Warping distance as an elastic measure. These two measures were chosen because they represent the two different categories of shape-based distances as well as the most commonly used distance measures according to literature^{18,19,21,23}.

3.4.2 Squared Euclidean Distance

One of the most commonly used lock-step distance measures is the Euclidean distance. All lock-step measures require both time series to be of equal length ($i = j$). In our data, the EEG segments were all detected on the time of the highest amplitude and surroundings were selected based on predefined length. Therefore, finding similar waveforms corresponds to finding EEG segments which show a similar pattern over time. The Euclidean distance was therefore likely to be a suitable method to define similarity. In this project, we decided to use the Squared Euclidean (SE) distance, which is almost equal to the Euclidean distance, except that the calculation of this distance measure is faster, as it does not take the square root.

The Squared Euclidean distance D_{SE} between two EEG segments X and Y is defined as:

$$\begin{aligned} D_{SE}(X_i, Y_j) &= (x_1 - x_1)^2 + (x_2 - x_2)^2 + \dots + (x_i - x_j)^2 \\ &= \sum_{i,j=1}^t (X_i - Y_j)^2 \end{aligned}$$

Note that lock-step distance measures, and thus the SE distance, are sensitive to noise, scale and time-shifts, and thus must be used with care, especially when applied on time-series data.

3.4.3 Dynamic Time Warping distance

In contrast to the lock-step measures, the elastic shape-based methods take into account the surrounding points in time to allow for shifts in time. Although the EEG segments are all the same length and aligned in time based on the position of the highest amplitude, the waveforms can still show time-warping effects. If this is the case, they will be matched best when elongating or shrinking parts of the EEG segments over time. Therefore, Dynamic Time Warping (DTW) distance was applied as elastic distance measure.

DTW calculates the smallest distance between two signals in a non-linear way. It distorts the signals and creates a $(t \times t)$ local cost matrix (LCM), where each cell (i, j) corresponds to the distance between elements x_i and y_j . Note that t represents the length and thus the number of samples of the EEG segments. This distance is defined as the quadratic distance $D(x_i, y_j) = (x_i - y_j)^2$. Subsequently, a warping path W is created, with $W = w_1, w_2, \dots, w_K$ and K the length of the warping path. The warping path always starts at the beginning of the time series and finishes at the end, so that each sample of both time series is included in the warping path. Another constraint of the warping path is that it is restricted by the following moves:

- Vertical moves: $(i, j) \rightarrow (i + 1, j)$
- Horizontal moves: $(i, j) \rightarrow (i, j + 1)$
- Diagonal moves: $(i, j) \rightarrow (i + 1, j + 1)$

A window parameter can be added as additional local constraint. The window parameter sets the maximum value for $|i - j|$. Figure 3.7 illustrates the minimum warping path for two EEG segments through the LCM.

The DTW distance is obtained by finding the warping path with the minimum cumulative distance for each next possible move. The total distance for the warping path is found by taking the sum of the individual distances of the LCM trough which the warping path traverses.

$$D_{DTW}(X_i, Y_j) = \sum_{k=1}^K w_k \tag{3.1}$$

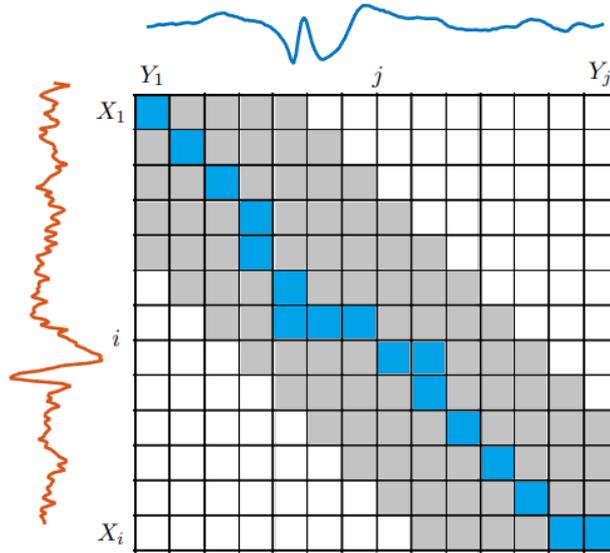


Figure 3.7: Minimum warping path through the LCM of two EEG segments X and Y . The grey area represents the boundaries of the warping window.

Note that this distance is equal to the SE distance when the minimum warping path traverses only the diagonal of the LCM.

3.5 Clustering

Many different applications of cluster analysis exist, and therefore many different clustering techniques have been developed. Generally, five different categories of cluster algorithms can be distinguished: distance-based methods, sub-divided into partitional and hierarchical methods, density-based, grid-based, model-based and multi-step methods¹⁹. Each of these categories can be divided into many more sub-categories and combinations. Since this an exploratory study, we wanted to start with an algorithm which was as simple as possible, but suitable for the time series data at hand. Distance-based methods are considered the most simple and easy to implement²⁰. Moreover, they can be used on time series data when an appropriate distance measure is applied. Therefore, we chose the K-means clustering algorithm, developed by Lloyd in 1982²⁴. The K-means is one of the oldest and most widely used distance-based algorithms²⁵. The main reason for choosing the K-means algorithm was its computational simplicity^{20,26}.

3.5.1 K-means

The K-means algorithm clusters the data into a predefined number of K clusters by minimising the distance between the cluster centroids and the objects within the clusters²⁷. The algorithm starts with K initial data points as centroids. These initial centroids are selected according to

Basic K-means algorithm

- 1: Select K points as initial centroids.
 - 2: **repeat**
 - 3: Form K clusters by assigning each point to its closest centroid
 - 4: Recompute the centroid of each cluster.
 - 5: **until** Centroids do not change
-

Table 3.2: Formal description of the basic K-means algorithm. Reprinted from '*Cluster Analysis: Basic Concepts and Algorithms*' by Kumar et al. (2005) In *Introduction to Data Mining*²⁰.

the K-means++ algorithm, a randomised seeding technique which applies a weighted probability to the random selection of the initial centroids²⁸. This technique decreases the computation time of the original K-means and improves the quality of the final clustering²⁸. Figure 3.8 A. shows an example of a data set with three initial clusters.

When all initial centroids are determined, the standard K-means algorithm can be applied. All data points (EEG segments) are then assigned to the closest centroid based on the distance between them. Each collection of data points assigned to a centroid forms a cluster, as shown in Figure 3.8 B. The centroids of all clusters are then updated based on the points belonging to each cluster, by using the mean of all points as new centroid (see Figure 3.8 C.). Based on the newly computed centroids, all points are re-assigned to the closest centroid, which might differ from the first assignment (see Figure 3.8 D.). The assigning of points and recalculation of the centroids is repeated until no points change clusters. Kumar et al. (2005, chapter 8) provide a clear pseudo-code of the basic K-means algorithm, which is presented in Table 3.2²⁰.

The assigning of points to the closest centroid is done based on the distance measure, where the algorithm seeks to minimise the distance of each point to its closest centroid. Since the initial centroids of the clusters are selected randomly, the outcome of the clustering can vary when a local optimum is found instead of the global optimum. Therefore, we run the K-means algorithm 50 times and select the result with the lowest sum of all distances between each point and its cluster centroid.

We set the maximum number of clusters per brain region to five since we did not expect that more than five different waveforms would be present in a brain region within a person. The number of optimal clusters, which was predefined to be between one and five clusters, was estimated with the gap statistic, as proposed by Tibshirani et al. (2001)²⁶. This technique is based on the difference between the within-cluster sum of squared errors for different values of K . For an increasing number of K , the sum of squared errors will decrease monotonically, but depending on the data set, this decrease will flatten at some point. Gap statistics estimate the number of clusters for which the sum of squared errors has the largest difference to its expected value. The principals of gap evaluation are explained in more detail in Appendix A.1.

The output of the clustering is saved as indices corresponding to the original EEG segments. Whenever a cluster contains less than three events, the cluster is deleted, and the events are included in a separate group of residuals.

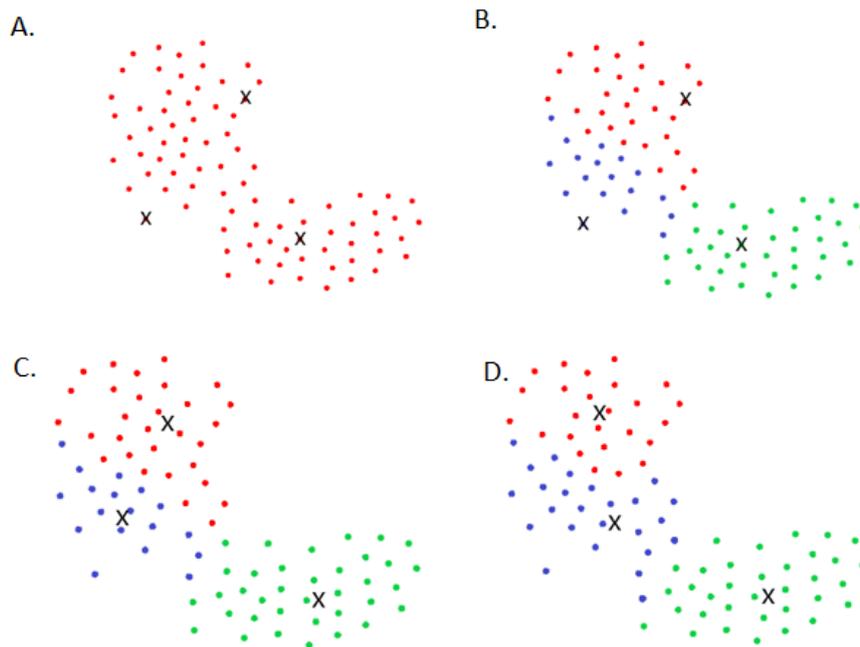


Figure 3.8: Iterations of the K-means algorithm. **a)** shows the initial cluster centroids. The data points are assigned to their closest centroid as shown in **b)** and **c)** displays the recalculation of the centroids. Based on the new centroid, the data points are again assigned to their closest centroid, as shown in **d)**.

4 | Design of the graphical user interface

4.1 Introduction

The results of the cluster algorithm are visualized in a Graphical User Interface (GUI). This chapter provides an overview of the GUI. The purpose of the GUI is to facilitate interaction between the cluster algorithm and the neurologist (user).

Kawamoto et al. (2005) showed that ‘automatic provision of decision support as part of the clinician workflow’ increases the success rate of clinical decision support systems with 75%²⁹. Hence, the workflow of the clinical department contains valuable information for the definition of the requirements. The potential users of the GUI were asked which features and requirements they found essential. The neurologist of the EMU at SEIN described that they desire a user interface to make the results of the clustering algorithm, and thereby automatic detection

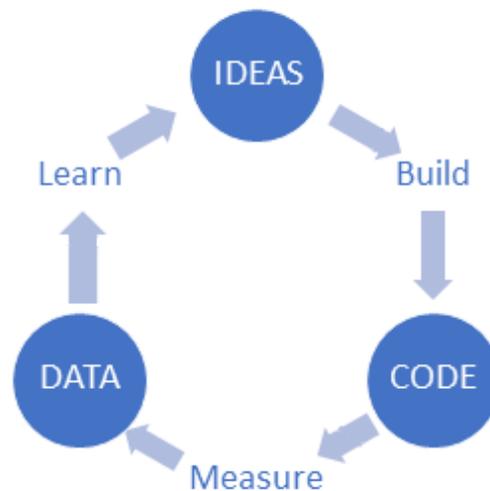


Figure 4.1: Build-Measure-Learn cycle. This loop represents the process of iterating a Minimal Viable Product (MVP). One starts with an idea, which is built into an MVP. This MVP is shown to the potential users and the results are measured. This feedback is then analyzed, and the developer will learn if they should persevere the initial idea or pivot and make drastic changes

Table 4.1: Requirements of the Graphical User Interface. A differentiation was made between the hard requirements, which define the features the GUI needs, and the soft requirements, which indicate the wishes of its potential users, but which are not considered to be necessary for clinical adoption.

Hard requirements (needs):
<ul style="list-style-type: none"> • Present the average waveform of each cluster • Indicate the variation of morphologies within a cluster • Show the localization of the detection • Point out the temporal occurrence of events in each cluster
Soft requirements (wishes):
<ul style="list-style-type: none"> • Ability to view events in a cluster individually • Review the surrounding EEG of a single detection • Ability to use different montages and filter settings

software in general, usable in the clinical practice. The user stories gave more insight into the desired functionalities and requirements. The basic GUI prototype was created, developed according to the Minimum Viable Product (MVP) concept.

The MVP concept is a part of the Lean Startup methodology, developed by Eric Ries³⁰. An MVP is a very basic prototype of the desired product, which can be evaluated to gather the maximum amount of feedback. The feedback from this first product iteration is then used to learn if the development of the product still goes into the right direction, or if changes must be made, which will lead to a second product iteration. This is done through the build-measure-learn process, as shown in Figure 4.1.

During the development of the GUI, several versions of MVPs were shown to the neurologists, starting with just some plots of the results, until a real GUI which allowed for user interaction. Each time, the feedback from the potential users was analysed and new functions were added to the MVP, or changes were made. This process resulted in the list of requirements shown in Table 4.1. It was differentiated in hard requirements, which represent the features the GUI must contain to reach its goal, and the soft requirements, which represent the wishes of the users.

4.2 Design

The GUI was designed with MATLAB App Designer (MATLAB 2019a), which is a MATLAB environment created for App building. The tool is designed in a way that multiple GUIs interact. On startup of the tool, the first GUI ‘Mainapp’ is opened. This is the main GUI of the tool which visualises all clusters per brain region. The Mainapp GUI only presents detections with a perception value of 0.9 or higher. When opening the tool, the Mainapp GUI is empty. At the top of the GUI, the tabs for all different brain regions are shown in the following order: Frontal left, Frontal right, Frontolateral left, Forntolateral right, Centroparietal left, Centroparietal right, parietooccipital left, parietooccipital Right, Midline, Generalized,

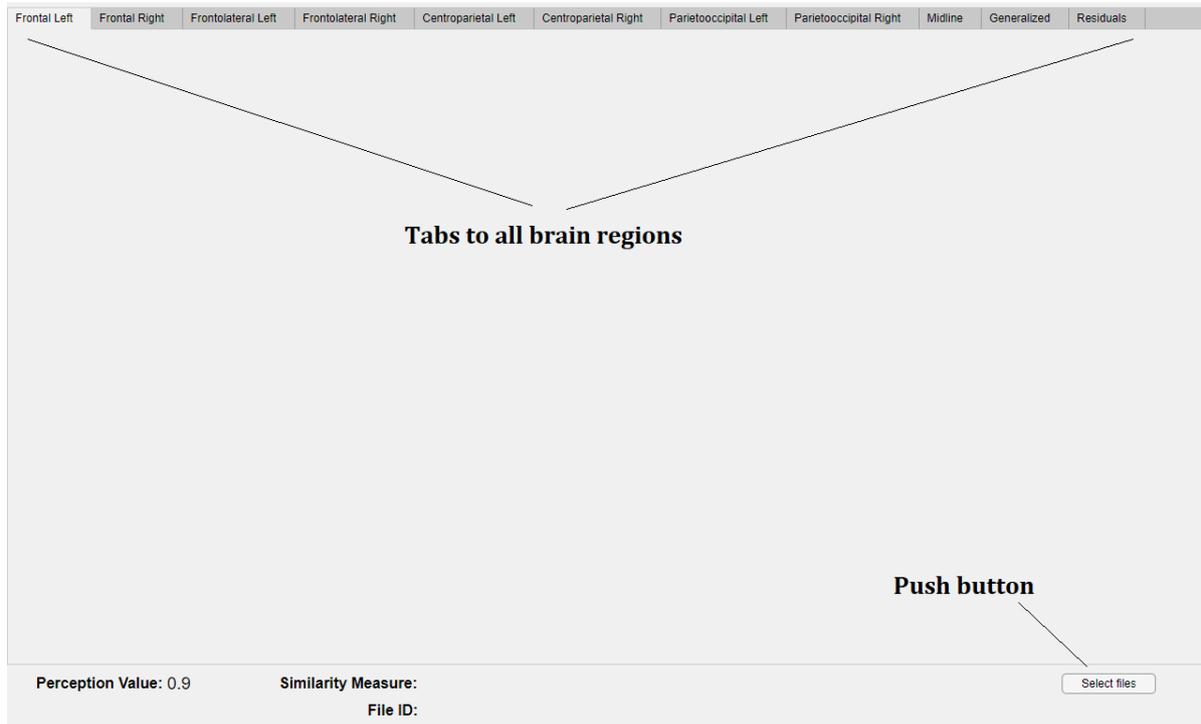


Figure 4.2: Screenshot of the startup screen of the Cluster tool. The tool always opens an empty version of the Mainapp GUI. The top row shows the tabs to all brain regions. In the left lower corner, the information about the visualised EEG and cluster settings will be displayed. In the lower right corner a push button is present which can be used to open the `Select_files` GUI to select and visualise a specific EEG recording.

Residuals. At the bottom of the GUI, the minimal perception value, the similarity measure used to create the clusters, and the file ID of the EEG are presented. These last two values are empty upon startup because no file is selected. The bottom line also includes a push-button to select files. Figure 4.2 shows a screenshot of the opening screen of the Cluster tool.

The Mainapp GUI has several callback buttons to other GUIs, such as the `'select_files'` GUI, the `'databrowser'` GUI and the `'0.4threshold_plot'` GUI. The architecture of the GUIs is shown in Figure 4.3. Each GUI has different functionalities. The `select_files` GUI is used to select and import data, to select a distance measure and to start the clustering algorithm. The Mainapp GUI is used to visualize the results of the cluster algorithm, of all events with a perception value of 0.9 or higher, whereas the `0.4_threshold` GUI does the same for all events with a perception value above 0.4. The Data Browser GUI can be used to visualize the individual events within a cluster, including four seconds of the surrounding EEG.

4.2.1 Data import, distance measure selection and clustering

The select files push button in the Mainapp GUI, opens a pop-up window with the `Select_files` GUI, as shown in Figure 4.4a. Through this GUI, the user can select the EEG file and

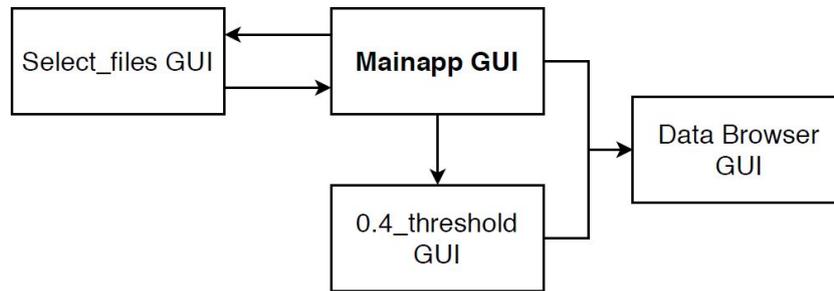
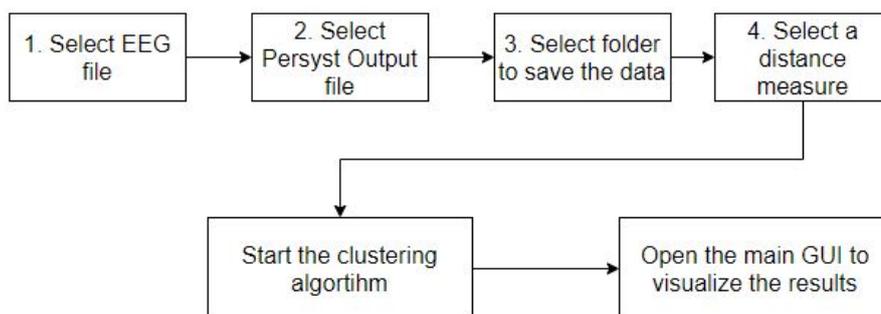


Figure 4.3: Architecture of the GUIs of the Cluster tool. The Mainapp GUI is always the first GUI to open. From here, other GUIs can be opened as pop-up window.



(a) Layout of the Select_files GUI



(b) User action flowchart

Figure 4.4: a) Overview of the Select_files GUI and b) the User actions required in the Select_files GUI to select input data and cluster settings and to visualise the cluster results of a specific EEG.

corresponding Persyst output file and select a folder where the cluster results will be saved. Through the ‘Browse’ -button, a pop-up window lets the user browse the file system to locate the specific EEG, .csv file and folder that the user wants to select. From the check-box in Figure 4.4a, the user can select the distance measure to be used in the cluster algorithm. After selecting all input, the user must press the ‘Run’ -button, which starts the cluster algorithm as described in the previous chapter. When the clustering is done, or the cluster results of that specific EEG were already saved in the selected folder, the OK push button will be enabled. This initiates the plotting of the cluster results in the Mainapp GUI, as described in the next section. These user actions are visualised in the flowchart in Figure 4.4b.

4.2.2 Visualization of the clusters

The results of the cluster algorithm are presented in the Mainapp GUI. Each cluster is plotted in the tab of the corresponding brain region. The number of clusters per brain region can vary between zero and five. Figure 4.5 shows an example of the frontolateral right region of a patient with five clusters. Each cluster is visualized by all electrodes belonging to that specific brain region (see Figure 3.4 in Chapter 3). That means that, although the event is detected on F8, the EEG signal of F10 and T8 is also displayed in the cluster plot, and these signals are also included in the calculation of the average waveform. The average waveform is presented by the fat coloured line, where each cluster has a different colour within a brain region. The grey waveforms which are seen in the background of the average waveform, are the individual

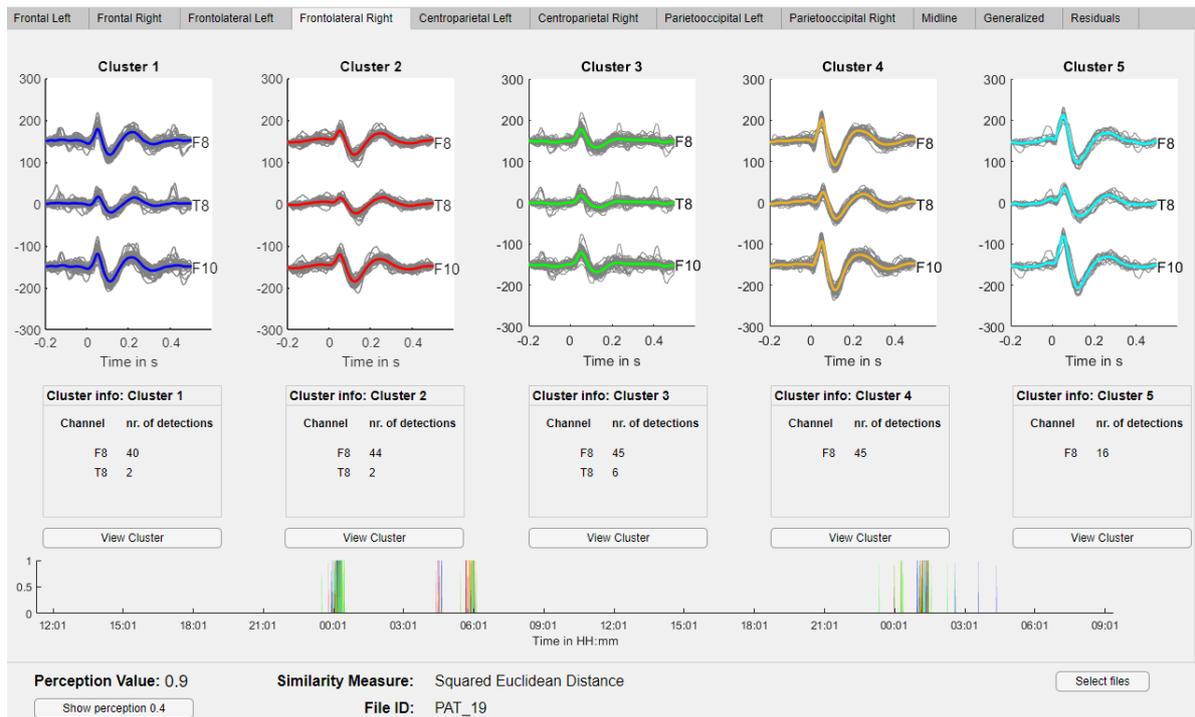


Figure 4.5: Layout of the Mainapp GUI. The tab that is shown is the Frontolateral right brain region. Five clusters have been found by the cluster algorithm in this region, with the Squared Euclidean Distance and a perception value of 0.9 or higher.

detections. This view can be used as an indication of the variation of the individual detections from the average waveform. The text box below each cluster plot contains information about the exact number of detections found on each electrode. Cluster 5 in Figure 4.5 for example contains 16 detections which are all detected on F8.

Each tab contains a timeline which visualises the temporal occurrence of the detected events per minute. The colour of the vertical line corresponds to the cluster number, and the height shows how often an event is detected per minute. Figure 4.5 only contains detections which occur once per minute. The “View cluster”-button enables the user to view all individual events within a cluster. By pressing this button, a pop-up window will open the Data Browser GUI. This GUI exists of the FT_databrowser function, implemented in the Fieldtrip Toolbox, which is a free MATLAB toolbox for EEG analysis.

Figure 4.6 shows an example of a frontolateral left brain region with only one cluster, which is inspected in detailed through the ft_databrowser function. This GUI allows scrolling through all the individual detections within a specific cluster. The EEG shown in the Data Browser GUI contains 2 seconds of EEG before and after the exact detection time and therefore provides more information about the context of the detection than the average plot. The “Show perception 0.4”-button can be used to open the ‘0.4threshold_plot’ GUI. This GUI is a

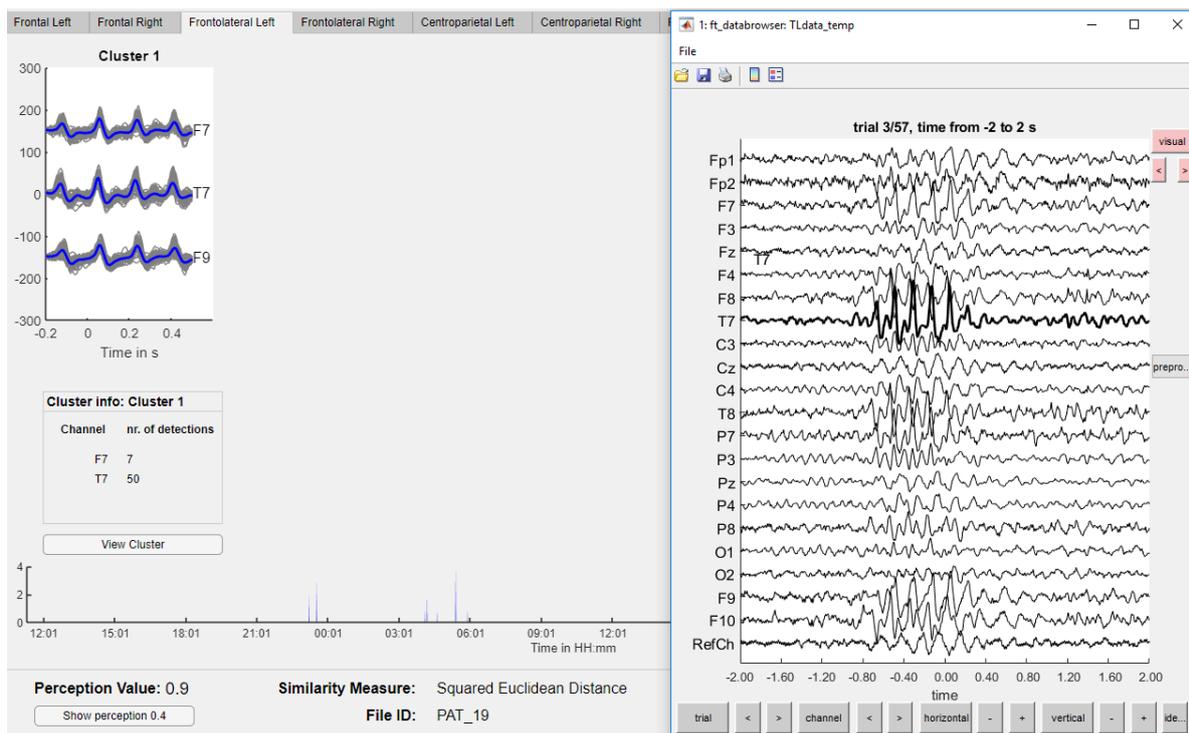


Figure 4.6: Mainapp GUI with pop-up window which shows the Data Browser GUI. The Data Browser GUI shows 2 seconds of EEG data before and after the exact detection time. All events included in the cluster can be viewed individually, and all EEG electrodes can be shown.

copy of the Mainapp GUI, except that the Perception value (left bottom corner) shows a value of 0.4. This option should only be used when the 0.9 threshold has only a few, or no detected events, and the user doubts if there are any epileptiform detections.

The tab 'Residuals' contains all events, from all brain regions, which were in a cluster with less than three events. These clusters were then removed from the brain region and included in the Residuals. Since these events are not clustered anymore according to their morphology and topology, it would not make sense to plot the average waveform. Therefore, this tab presents the events in a Table, where the number of detections per electrode is shown. A "view Cluster"-button is included in the Residuals tab, to view all individual events in this group. The 'Generalized' tab contains all events which were detected as such by Persyst P13. These events are also not clustered according to morphology and topology and are therefore presented in the same way as the Residuals.

5 | Evaluation of the Cluster Tool

5.1 Introduction

In the previous chapters, we described the design of the clustering algorithm and the design of the GUI. The clustering algorithm was implemented in the GUI and subsequently used for visualisation of the cluster results. The combination of algorithm and GUI is from now on, referred to as the Cluster Tool. Evaluation of the Cluster Tool is necessary to measure the performance of the clustering algorithm and to assess if the tool fits into the clinical workflow of SEIN. The current chapter introduces methods to evaluate the performance and usability of the Cluster Tool. First, a review of existing methods on cluster evaluation is provided. Then, the methods applied for evaluation of the Cluster Tool are presented, followed by the results. Finally, we discuss the results and evaluation methods applied.

5.2 Literature review

Cluster evaluation is a difficult part of cluster analysis^{19,20,31}. The main problem of cluster evaluation is captured well by Aggarwal et al. (2013), who state that "*clustering is a problem in which precise quantification is often not possible because of its unsupervised nature.*"³². Generally, a cluster is defined to be good when objects in a cluster are similar to each other and different from the objects in other groups. Therefore, cluster evaluation is directly linked

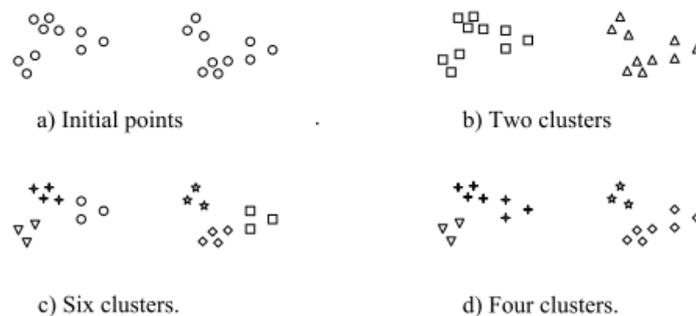


Figure 5.1: Different ways of clustering the same set of points. Adapted from ‘The Challenges of Clustering High Dimensional Data - Basic Concepts and Techniques of Cluster Analysis’ by Steinbach, M. (2003)[?]

to the definition of the distance measure, which defines the similarity between objects in a data set. An understanding of the data is required to define the quality of cluster results. Figure 5.1 presents multiple ways to divide a data set into clusters, which all seem to be possibly correct. This overview of the data enables us to get an impression of the data structure and helps us to determine which cluster results are potentially good. Note that it is complex to visualise the structure of the high dimensional time series in 2D. Nevertheless, the data always need to be placed in context, and the purpose of the clustering needs to be considered to decide which clustering is correct.

The main techniques for cluster evaluation of time series are visualisation and scalar measurements¹⁹. Scalar measures can be divided into two types: external and internal indices. The possibilities and limitations of these techniques are discussed below. This information is subsequently used to decide which evaluation methods are suitable to evaluate the Cluster Tool.

5.2.1 Scalar measurements

External validation

One possible way to validate clusters is to compare each object within the cluster with an externally provided class labels, which represent the ground truth²⁰. External validation measures represent the degree of agreement between the cluster results and the class labels¹⁹. This approach requires external information about the true clustering and is therefore called external validation. Note that these externally provided class labels are defined subjectively, by a human expert. Although these external labels contain subjective judgement, they do present the shortcomings and strengths of clustering algorithms. Therefore, external validation provides the best possible way to examine the performance of a cluster algorithm^{19,20,33}.

Internal validation

Internal validation provides another possibility to validate cluster algorithms quantitatively^{19,20}. These measures define the quality of a clustering algorithm based on the cohesion and separation of the clusters. Cohesion is defined as the degree of similarity between objects within a cluster, whereas separation is defined as the difference between objects of different clusters (see Figure 5.2). Internal measures can be used to evaluate cluster results when no cluster labels or any other kind of external information about the accuracy of the clustering is present. However, these validation measures can only be used to compare cluster algorithms, that apply the same metric¹⁹.

5.2.2 Case studies

Alternative methods need to be applied if no measure can be defined to assess the quality of a clustering³⁴. Visualisation of the clusters provides valuable information about cluster quality. Therefore, qualitative case studies are well suited to evaluate the performance of an algorithm, in case quantitative measures are absent³⁵. Examples of good and bad performances provide a general impression of the quality of the output of the clustering algorithm. An advantage of

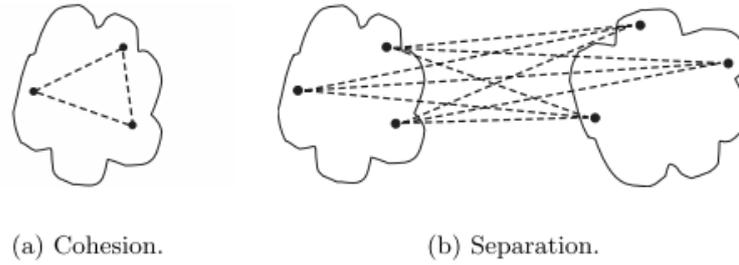


Figure 5.2: Illustration of the definition of a) Cohesion and b) Separation. Adapted from ‘Cluster Analysis: Basic Concepts and Algorithms’ in Introduction to Data Mining, by Kumar, V., Tan, P.-N., & Steinbach, M. (2005)²⁰

case studies is their intended focus on a particular issue or feature that provides the possibility to point out specific drawbacks or advantages of an algorithm. On the other hand, case studies only present individual cases. Consequently, the interpretation of a case study can only lead to assumptions on the behaviour of the algorithm on a larger scale. It must be considered that evaluation of the clustering through visualisation does not only evaluate the performance of the clustering algorithm but also evaluates the quality of the visualisation of the clusters.

5.2.3 Usability tests

Usability tests provide another method to get an impression of the performance of a digital product and simultaneously assess its usability. Usability testing is the process of watching an actual user while they use the product. The so-called moderator sits together with the test participant and helps them through the task, answers their questions and replies to the feedback. Moderated usability testing provides live user feedback which contains valuable information about the usability of the product, including the achieved accuracy and speed for reaching the products goal³⁶.

5.3 Methods

A visualisation-based approach, including case-studies, was chosen to evaluate the performance and usability of the Cluster Tool. Qualitative methods offer an effective way of measuring user experience and discovering the pitfalls and advantages of a product. However, qualitative results are more difficult to compare systematically. Quantitative evaluation through scalar measures was considered but was found not to be feasible. Due to the absence of class labels, we could not perform external validation. The usage of two internal validation measures was explored during this project: the sum of squared errors (SSE) to measure the cohesion and the between the sum of squared errors (BSS) to measure the separation of the cluster results. However, these were no good indicative measure for the quality of the clustering, as presented in Appendix A.3.

The performance of the clustering algorithm was also assessed on a more global level to

evaluate the accessibility of the cluster results. As stated by Kumar et al. (2005), there is more to cluster evaluation than obtaining an exact numerical measure of the validity of a clustering²⁰. More importantly, cluster evaluation should always consider the usefulness of cluster results. The purpose of our clustering algorithm was to summarise the output of an automatic spike detection algorithm and subsequently present the detected events in a way that enables efficient clinical interpretation. The clinical conclusion in the EEG report contains information about the morphology, localisation, and temporal occurrence of the IEDs, and concludes with a clinical diagnosis. For the Cluster Tool to be clinically useful, the user of the tool should be able to get to the same clinical conclusion as is currently described in the EEG report. Therefore, the performance of the Cluster Tool was also evaluated based on these criteria.

5.3.1 Test data

For this project, EEG recordings of 35 patients were retrospectively selected from the EMU at SEIN, Heemstede. Inclusion criteria were age above fifteen, and an EEG with approximately 24 or 48 hours of recording, which had to be of good quality. The EEGs were recorded with the Micromed system at a sampling frequency of 256 Hz, in a frequency band of 0.01 to 100Hz. The electrodes were placed according to the 10-20 system with additional sub-temporal electrodes. All EEGs were separately analysed by the Persyst P13 spike detector, resulting in a .csv file containing information about the time, channel and perception value of all detections found in the EEG. The detections with a perception value of 0.9 or higher were included for further clustering.

5.3.2 Usability testing

Test setup

Moderated usability tests were performed to gather information about the review time and usability of the Cluster Tool. A participant-observer sat together with the test participant and would help him or her through the task of analysing a long-term EEG recording with the Cluster Tool. Each EEG was reviewed twice, once with the SE distance, and once with the DTW distance. No participant would review the same EEG with a different distance measure. Due to logistic reasons, it was not possible to let each participant review the same number of EEGs. Therefore, some participants used the Cluster Tool more often than others.

The Cluster Tool was installed on a 64-bit virtual Windows computer. This virtual machine was reachable from every computer within the network of SEIN so that each participant could perform the tests from their preferred workplace. Five EEG experts, four neurologists and one physician assistant, were included as test participant. All five participants were experienced EEG readers and regularly evaluated long-term EEGs for diagnostic purposes at the EMU of SEIN. Before using the Cluster Tool, the participants were instructed to review a maximum of one hour of the wake EEG, including the diagnostic test, the first hour of sleep and the first half-hour after waking. These hours were selected because it represents the workflow SEIN wants to achieve in the future (see section 2.1). The test participants could review this part of the EEG in Micromed, the EEG viewer which they currently use in SEIN, and they could use their preferred montage. The rest of the EEG was subsequently reviewed by using the Cluster

Tool. The test participants were allowed to ask questions to the participant-observer while they used the Cluster Tool. After reviewing the complete EEG, the participants formulated their conclusion, including information about the morphology, localisation, and temporal occurrence of IEDs. They also formulated a clinical diagnosis based on their findings.

Data collection

Data about the review time and usability of the tool were obtained during the usability tests by the participant-observer who kept track of all feedback. The review time could only be assessed through estimation of the participants, who indicated whether the EEG evaluation was quicker with the Cluster Tool. In order to identify the advantages and pitfalls of the Cluster Tool, the participants were asked to think aloud and encouraged to give feedback on their experience. The participant-observer took notes of all positive, negative and neutral comments and monitored on what EEG certain comments applied. It was also registered how often a test participant used the feature to inspect clusters by their individual detections through the *'view cluster'*-button. Data on the number of clusters per brain region and cluster size were automatically saved in an Excel spreadsheet by the Cluster Tool.

Data analysis

Descriptive statistics were performed using R (version 3.5.3) in RStudio (Version 1.1.463). Data on individual experiences of the performance, review time and usability of the Cluster Tool were summarised and described per category. Case studies were performed on selected EEGs, based on remarkable or frequent findings from the usability tests.

5.3.3 Performance evaluation

Test setup

The performance and usefulness of the Cluster Tool were evaluated by comparing the conclusions that resulted from the use of the Cluster Tool with the conclusions described in the clinical report of SEIN. This resulted in three conclusions per EEG: one from the clinical report, one based on the Cluster Tool with SE distance, and one based on the Cluster Tool with the DTW distance. The clinical report of SEIN was created as part of the routine clinical practice, as described in section 2.1. The data that originated from the usability tests could not be used for the performance evaluation, because the clustering algorithm contained a mistake in the calculation of the similarity measures at the time the usability tests were performed. Although the cluster results did not seem to change significantly based on visual inspection, we decided not to use the clinical conclusions but only the user experiences resulting from the usability tests. Therefore, the clinical conclusions based on the Cluster Tool had to be defined again.

Due to time restrictions, the extensive usability study could not be repeated. The clinical conclusions were obtained by letting two test participants use the Cluster Tool and create a mutually agreed on the conclusion. An experienced and novice EEG reader, the latter was also involved in the design of the Cluster Tool, served as test participants for this task. The clinical conclusion was solely based on the output provided by the Cluster Tool. No parts of

Table 5.1: Example of a clinical conclusion based on the cluster tool. The information is divided into morphology, localization, temporal occurrence and the final clinical conclusion

	Morphology	Localization	Temporal occurrence	Diagnosis
<i>Conclusion</i>	Polyspikes-and- slow-waves,	Frontal	Increase	Epilepsy
<i>description</i>	spike-and-slow- waves	bilateral	during the night	

the EEG were reviewed in Micromed. Because we were interested in the detection of IEDs, we considered only the clinical conclusion regarding epileptiform waveforms. The test participants reviewed the clusters per brain region. They were instructed to use the detailed view of the individual detections as least as possible. However, when in doubt, they were allowed to use this feature. The timeline of the GUI was used to get an impression of the temporal occurrence of the events in each cluster.

Data collection

The clinical conclusion based on the Cluster Tool was described per EEG, based on the morphology, localisation and, if remarkable information was present also the temporal occurrence of the IEDs. The two EEG readers decided on a mutually agreed-upon clinical diagnosis, which could be ‘epilepsy’, ‘normal EEG’, ‘abnormal non-epileptic’, or ‘uncertain’. An example of the obtained information is shown in Table 5.1. Each EEG was reviewed two times, once with the SE and once with the DTW as the similarity measure. The clinical conclusion as described in the EEG report of SEIN was analysed and divided into categories containing the information about the morphology, localisation, temporal occurrence and clinical diagnosis.

Data analysis

The clinical conclusions created with the Cluster Tool were compared to the events described in the clinical report. Whenever an event was described similarly in the final report, it was counted as detected by both, the Cluster Tool and the EEG report. This comparison was made for all information categories, being morphology, localisation, temporal occurrence and diagnosis. The conclusion described in the clinical report of SEIN was considered the gold standard.

5.4 Results

5.4.1 Included data

Due to time restrictions, 23 of the initial 35 EEGs were included. The 23 included EEGs had a total of 591 hours of EEG recording. The age of the included patients varied from 20 to 65 years, with an average of 37. These EEGs were used for both the usability test and the performance test. Persyt P13 detected between 3 and 2531 events per EEG. The distribution

of the number of detections per EEG was positively skewed, with a mean of 480 and a median of 128 detections per EEG. The exact number of detection per EEG is presented in Table A.1 in Appendix A.5.

5.4.2 Performance evaluation

The clinical conclusion based on the EEG report and the conclusion created with the Cluster Tool with both, the SE and DTW distance, are presented in Appendix A.5. Comparison of these conclusions revealed that the clinical diagnoses made with the SE and DTW distance were the same for all EEGs. Figure 5.3 shows whether the diagnosis created with the Cluster Tool corresponded to the diagnosis as described in the EEG report. It can be observed that the clinical diagnosis ‘epilepsy’ (red colour) corresponded in all EEGs, except one. This exception was EEG ID 9, for which the diagnosis made with the Cluster Tool was ‘uncertain’. It stands out that the results in Appendix A show similar morphology, localisation and even temporal occurrence as described in the EEG report, but differs regarding the clinical diagnosis. A detailed case study on the output of the Cluster Tool for this EEG is presented in section 5.4.5.

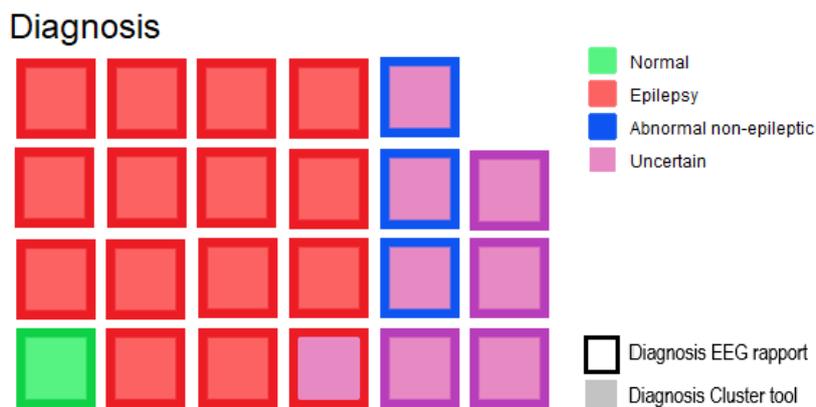


Figure 5.3: Results of the comparison between the clinical diagnosis from the EEG report (gold standard) and the clinical diagnosis based on the usage of the Cluster Tool. The diagnosis based on the Cluster Tool represents both, the diagnosis made with SE and the diagnosis made with DTW distance. Each square represents an EEG. The border of the square shows the diagnosis based on the EEG report, while the fill of the square represents the diagnosis based on the Cluster Tool. One EEG was normal and also correctly diagnosed with the Cluster Tool (green square). Fifteen EEGs were epileptic, of which fourteen were diagnosed correctly with the tool (red squares). All other EEGs were marked with an uncertain diagnosis based on the Cluster Tool (pink filled squares). One of these should have been diagnosed as epileptic, and three as abnormal non-epileptic.

Table 5.2: Results of the comparison of the morphologies described in the clinical conclusion of the EEG rapport and the clinical conclusion based on the usage of the cluster tool, separated into the use of the SE and DTW distance. The table shows the number of times the morphology was described per method

	EEG rapport	SE	DTW
Sharp waves	10	9	9
Sharp-and-slow-waves	9	8	8
Poly spikes	1	1	1
Poly-spike-and-slow-waves	3	2	3
Spike-wave-complexes	4	4	4
Sharp transients	8	7	7

One EEG was defined as ‘normal EEG’ by the EEG report. This EEG was also described as normal by the EEG readers when using the Cluster Tool. All other clinical diagnoses were defined as uncertain by the users of the Cluster Tool. It stands out that no diagnosis of ‘abnormal non-epileptic activity’ was defined when using the Cluster Tool. These cases were always described as an uncertain diagnosis.

In general, most IEDs that were described in the EEG report were also described while using the Cluster Tool. Table 5.2 compares the number of different IEDs described in the conclusion based on the Cluster Tool, with the ones described in the EEG report. The only difference between the SE and DTW distance was one event of poly-spike-and-slow-waves that was not found by using SE distance. In a few cases, a certain type of IED was not described while using the Cluster Tool. However, the usability study revealed that this was mostly due to interpretation. If sharp-and-slow-waves were found in an EEG, some EEG readers did not bother to describe that sharp waves are also present at that localisation. In case sharp transients were observed in an EEG, it depended on the rest of the findings in the EEG whether the EEG reader would define the sharp transients as sharp waves or sharp transients. Note that sharp waves are a type of IED whereas sharp transients are not specific epileptiform. However, these few differences in the description of IEDs did not influence the clinical diagnosis. The results in Table A.1 also present the localisation of the epileptiform detections. The localisation as described by the test participants while using the Cluster Tool corresponded with the ones described in the EEG report for all EEGs.

The test participants noticed that the timeline gave a good impression of the frequency and distribution of the epileptiform events. They even considered the information about the temporal occurrence to be more accessible compared to the current workflow. However, it was also noticed that some clusters contained detections which were not considered to be epileptiform. For these clusters, the timeline gave a biased impression of the frequency of epileptiform detections.

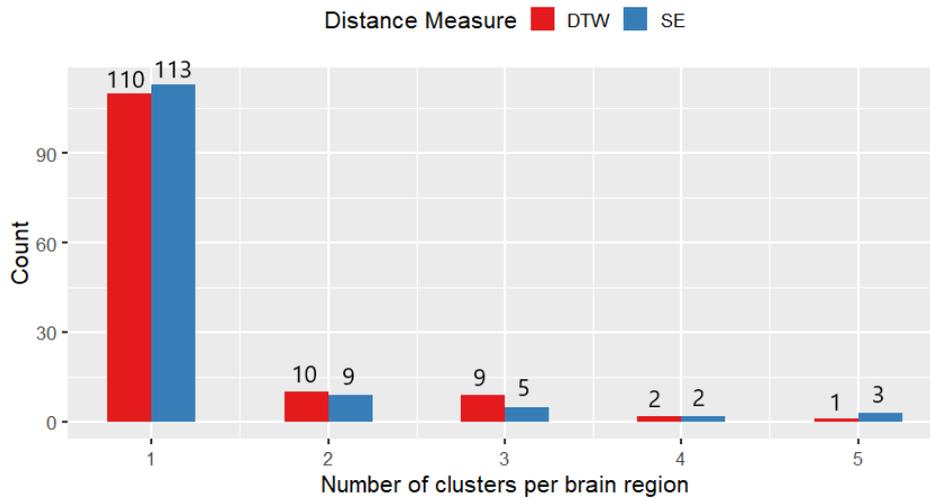


Figure 5.4: Number of times a certain number of clusters is created in a brain region. Most of the times, only one cluster is formed per brain region.

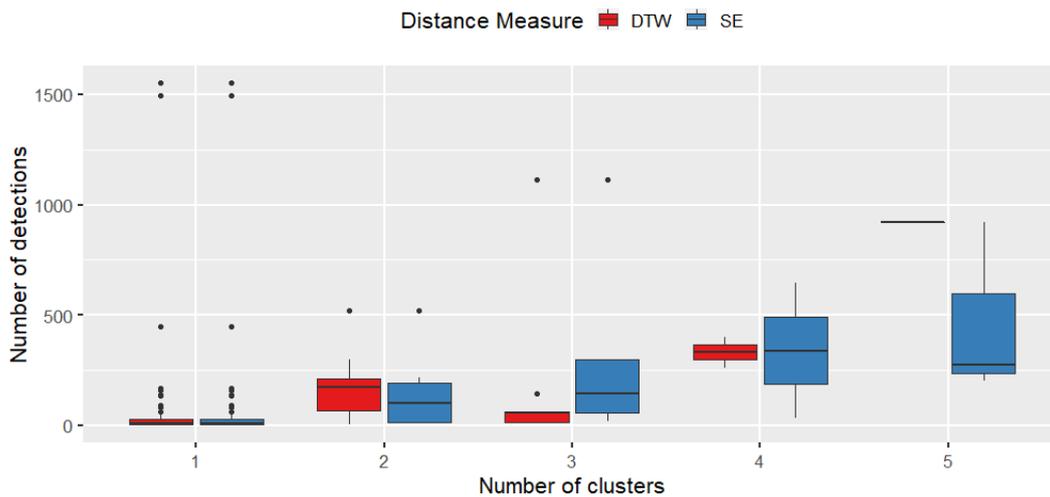


Figure 5.5: Number of detections in a brain region in relation to the number of clusters in that brain region. Most clusterings that contained one cluster contained a relative small number of detections.

5.4.3 Usability evaluation

Figure 5.4 shows that the Cluster Tool mostly partitioned a data set into one cluster, and rarely in two or more clusters. This was the case for both similarity measures. Figure 5.5 shows the relation between the number of detections and the number of clusters per brain region. Brain regions with a higher number of detections, seem to result in more clusters. However, several outliers can be observed, where a high number of detections resulted in just one cluster.

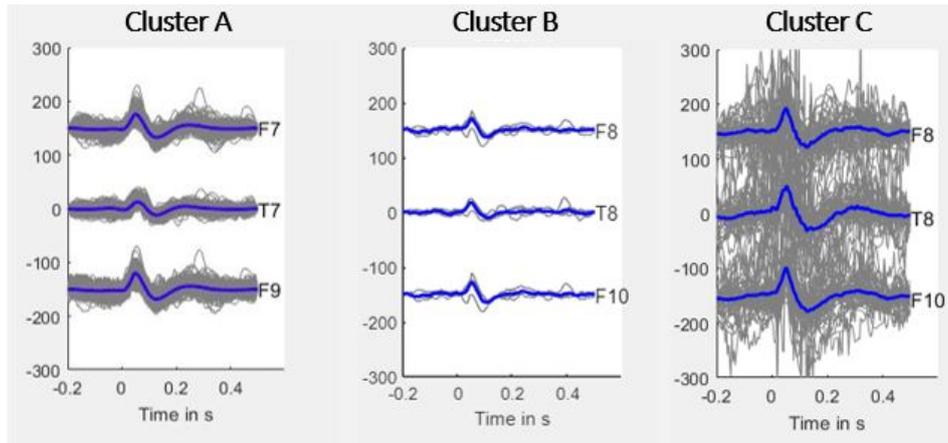


Figure 5.6: Example of three different clusters, where all events were put together into one cluster. **A.** shows a cluster which contains 1493 events. **B.** shows a cluster that contains only five detected events and **C.** shows a cluster which contains 63 events, disrupted by many artefacts

The test participants distrusted the performance of the clustering algorithm, and thereby the overview of the clusters, because the it did not show clearly when different morphologies were present. Especially when the data set was partitioned into one cluster, the amount of different morphologies within a cluster was high. Figure 5.6 shows three different examples of clustering results where just one cluster was created. In Figure 5.6 A, a cluster which contains sharp-waves and sharp-and-slow-waves is presented. Test participants could not distinguish these two morphologies based on the cluster overview. This supported the distrust of the test participants in the performance of the clustering algorithm. Figure 5.7 shows an example of different morphologies which were found within a cluster.

Clusters that contained a small number of detections were often often easy to interpret for the test participant. Figure 5.6 B shows an example of a cluster with only five detections. On the other hand, clusters that contained many artefacts, as shown in Figure 5.6 C, were experienced as difficult to interpret. For clusters that contained a high number of detections or many artefacts, the ‘*view cluster*’-button was used frequently by the test participants to inspect the detections individually. Since the test participants spend a lot of time on manually checking the detections within the clusters, the review time of an EEG was experienced as relatively long.

5.4.4 Distance measures

The DTW distance seems to result in a slightly better clustering in a few cases. It was not possible to determine which distance measure performed best based on clinical evaluation. Therefore, the difference in distance measures was evaluated through usability tests and case studies. Figure 5.8 shows different cluster results based on the SE and DTW distance. The clustering based on the DTW distance, clusters all spike-and-slow-wave EEG segments together

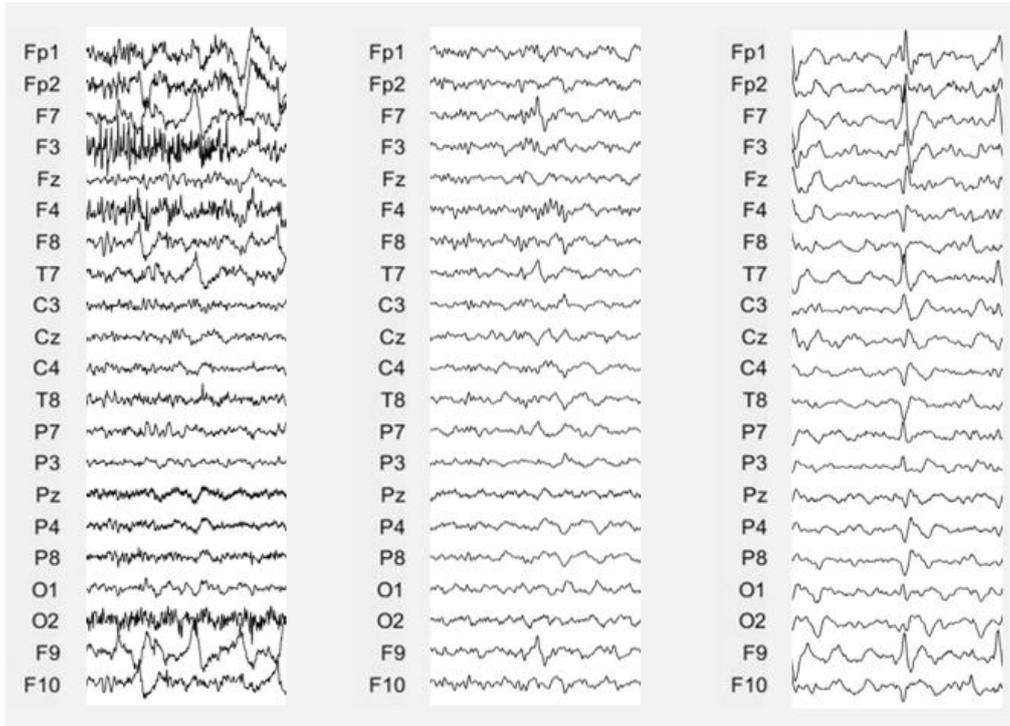


Figure 5.7: Example of three detections from the same cluster, originating from the fronto-lateral left brain region. The morphology from the different detections is not similar.

in one cluster, which resulted in a better separation than the clustering with SE distance. It occurred more frequently that the DTW clustering resulted in a higher separation between clusters compared with the SE clustering.

5.4.5 Case study

The evaluation of the EEG with ID 9 revealed several interesting facts about the performance and usability of the Cluster Tool. A total of 101 events were detected by Persyst P13 in this EEG. The only difference between the DTW and SE distance measures was observed in the frontolateral right brain region, which is depicted in Figure 5.9 and 5.10. It shows that DTW clustering resulted in more clusters, which waveforms can be distinguished more clearly than in the single cluster created with the SE distance. However, in both cases, the users could not draw a conclusion based on solely the cluster overview and preferred to inspect the detections in detail.

EEG ID 9 is the only EEG for which the clinical diagnosis was marked as 'uncertain' by the users of the Cluster Tool, while it should have been 'epilepsy' according to the EEG report. It is remarkable though, that the morphology, localization and temporal occurrence were described similarly for the EEG report and the Cluster Tool, as can be seen in Table A.1 in Appendix A.5. The lack of information about the surrounding EEG and the patient's history as well as the absence of filter and montage settings in the Cluster Tool were named as the cause for insufficient information for the EEG readers to define a certain clinical diagnosis. This

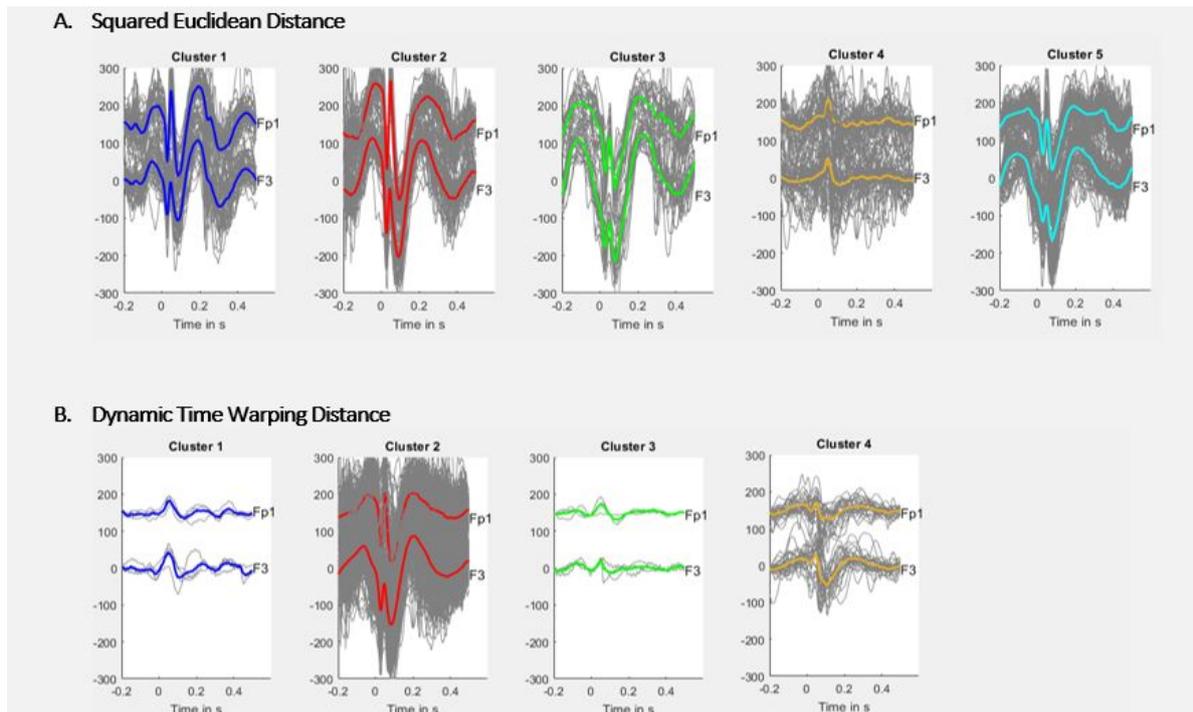


Figure 5.8: Screenshot of the GUI presenting the results of SE and DTW clustering. **A.** shows the results of clustering with SE distance, where the wave forms of the different clusters look similar to each other. **B.** shows the results of clustering with DTW distance. The DTW captures all spike-and-slow-waves into one cluster.

information was needed because the waveforms reviewed in this EEG were unclear. The EEG readers discussed that the waveforms mainly looked like small sharp-and-slow-waves, which could be interpreted as physiological small sharp spikes, also known as Benign sporadic sleep spikes. This shows the difficulty of the interpretation of these waveforms and the importance of additional information of the EEG in such cases.

5.5 Discussion

Our prototype shows a striking visual comparison between the clinical diagnosis based on the Cluster Tool and the EEG report. The diagnoses 'normal' and 'epilepsy' were made correctly for all except one case. Three EEGs with the diagnoses 'abnormal non-epileptic activity' were also diagnosed incorrectly. The Cluster Tool is designed with the purpose to present IEDs in a comprehensive overview. It does not provide any information about slow rhythmic activity or other abnormal non-epileptic phenomena. Therefore, it is acceptable that the EEGs with diagnoses other than 'normal' and 'epilepsy' are not correctly diagnosed.

The main focus for the design of our prototype was that the diagnosis 'epilepsy' could be drawn correctly based on the usage of the Cluster Tool. The case study presented in section 5.4.5 showed that the one EEG that was diagnosed incorrectly contained difficult to interpret

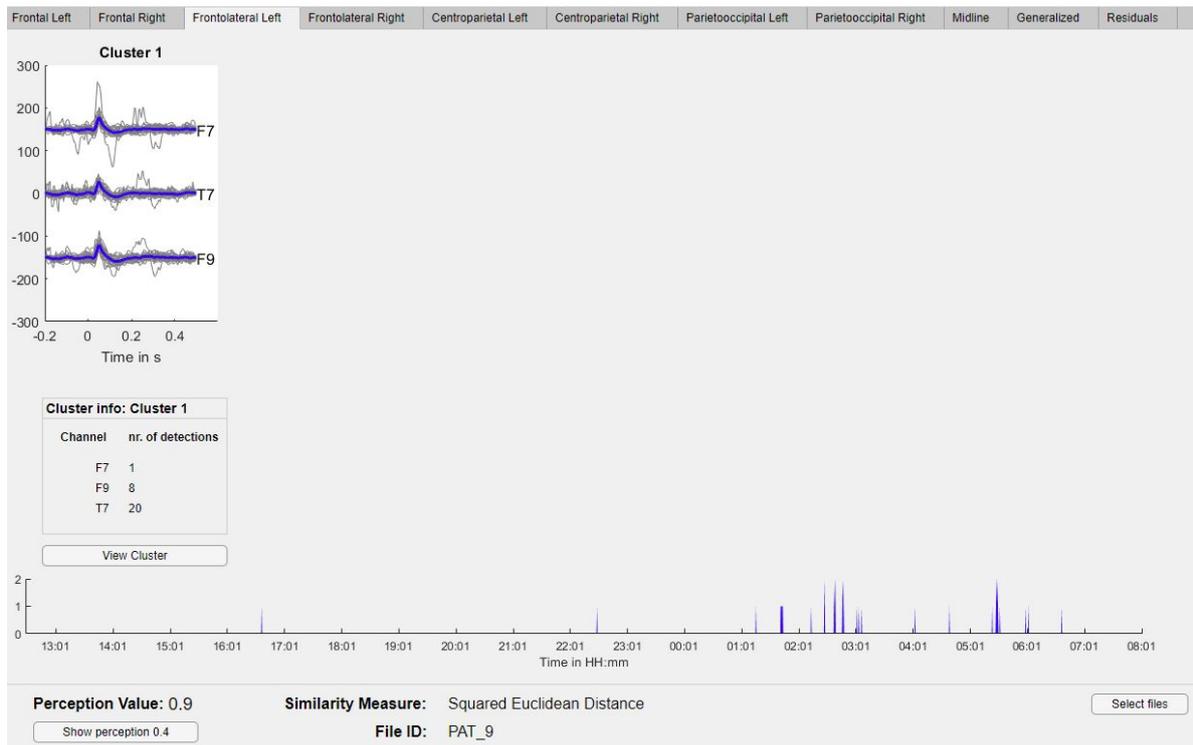


Figure 5.9: General overview in the GUI for the results of the clustering with squared euclidean distance in the Frontolateral brain region of EEG ID 9

waveforms. It is known that IEDs lack a clear definition, and due to high interrater variability, the gold standard is doubtful^{6,9}. This is also shown by the fact that the EEG report, which is considered our gold standard, defined four EEGs with an uncertain diagnosis. Nevertheless, the Cluster Tool must provide all information that is needed so that its users can apply their knowledge about the EEG to their full extent. This requires access to the surrounding EEG and the patient’s history as well as to different montages, filter settings and a more detailed version of sensitivity settings. Lagerlund (2002) also stresses the importance of access to different montages and filters for accurate EEG evaluation³⁷. It is recommended that these features will eventually be included in such a visualisation tool.

In approximately 83% of the cases, the clustering resulted in only one cluster. The presence of just one cluster indicates that all EEG segments in the data set are similar. However, test participants frequently noticed that the waveforms within a cluster were highly variable and that the average waveform presented in the overview of the Cluster Tool did not capture this variability. This resulted the fact that the test participants distrusted the performance of the Cluster Tool and caused them to spend time on manually checking the detections within clusters. Therefore, the time spend on the EEG evaluation was not positively influenced by the use of the Cluster Tool. Furthermore, it was not possible to differentiate true IEDs from false positive detections based on the cluster overview. The bad performance of the clustering algorithm impacted the clinical implementation possibilities strongly to the negative. It is expected that an improvement of the clustering algorithm will quickly resolve this issue.

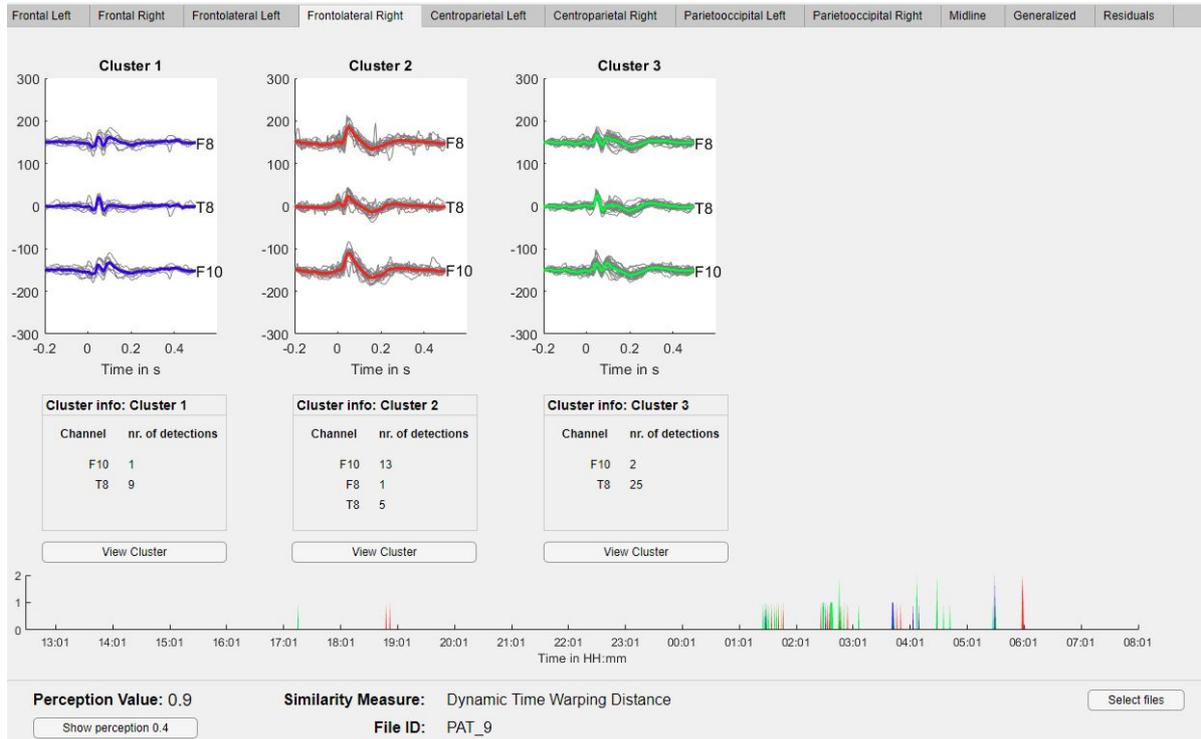


Figure 5.10: General overview in the GUI for the results of the clustering with dynamic time warping distance in the Frontolateral brain region of EEG ID 9

It was expected to observe a learning curve based on the number of clusters that was inspected individually. An increased usage of the Cluster Tool was presumed to result in a decreased usage of the 'view cluster'-button. Figure A.4 in Appendix A.4 shows that this was not the case. We assume that this is also caused by the distrust of the EEG readers towards the performance of the clustering algorithm.

Although the performance of the clustering algorithm was experienced as poor, the usage of the GUI resulted in more satisfying experiences. The localisation and temporal occurrence of the detections were captured well by the tool. The time line that presents the number of detections per minute was found to give a clear and objective overview, which was considered better than the current estimation of the temporal occurrence of IEDs. The localisation of the IEDs described when using the Cluster Tool corresponded completely with the localisation described in the EEG report. Nevertheless, the test participants noted that they would like to be able to view different brain regions at the same time and merge similar clusters from neighbouring brain regions. This indicates that the current choice of brain regions is not optimal. Ideally, the localisation of the IEDs is presented by their potential source. This does not need to be in one brain region as we defined them. The division into brain regions as defined in the Cluster Tool was therefore experienced as sub-optimal.

The usage of the DTW distance seemed to result in clusters with a higher separation. The goal of the Cluster Tool was to group all events with similar morphology in the same cluster. It is known from previous research, that IEDs within patients tend to be morphologically more similar than between patients¹⁵. Since the DTW distance results in a slightly better separation of the clusters compared with the SE distance, it seems like it captures similarity of EEG waveforms more effectively. Thomas et al. (2016) also compared Euclidean and DTW distance when clustering IEDs and found that DTW provided a more effective approach than the non-elastic Euclidean distance²⁷. Therefore, DTW distance is found to be a more promising distance measure.

All in all, this evaluation identified that the visualisation of the IEDs by the GUI was experienced as satisfying. However, the Cluster Tool is still a prototype and the performance of the clustering algorithm was too inaccurate to impact the clinical workflow positively. We believe that the Cluster Tool is the first step towards a faster, reproducible and more objective method for EEG evaluation.

6 | General Discussion

The evaluation of the clinical performance and usability of the Cluster Tool indicates the potential of using a comprehensive overview, which summarises the results of an automatic spike detection algorithm through clustering techniques. It shows the importance of post processing the results of automatic spike detection algorithms, to visualise them in a way that enables correct interpretation. Clustering is an essential part of the post processing, and entails an important step towards a proper visualisation and thereby clinical implementation of automatic detection algorithms.

The Cluster Tool designed in this project is a prototype, and this was the first time such a tool was created and evaluated in SEIN. Therefore, many choices have been made to keep things simple for this initial prototype. The Cluster Tool still suffers from many limitations, impeding its implementation in the clinical practice. However, by building this prototype, many insights have been gained on what features are important for the visualisation of IEDs, how to improve the clustering, and on alternative methods that might provide better results. This chapter reviews the methods used in the Cluster Tool in the context of the gained insights and discusses which alternative methods are likely to lead to better clustering performance, necessary for clinical adoption.

6.1 The cluster algorithm

The visual assessment of the Cluster Tool indicated that the current clustering algorithm was too inaccurate. In order to use the Cluster Tool in the clinical practice the clustering algorithm must be improved. The clustering algorithm developed in this project consists of three main steps: the data preparation, the definition of the distance measure and the choice of cluster algorithm. The three steps are reviewed separately in the following sections.

6.1.1 Data preparation

Noisy EEG segments have been observed in the Cluster Tool, which are likely to have a negative impact on the clustering. The preparation of the data has a large impact on the cluster results. Noisy time series can result in clusters where groups are created based on similarity in noise rather than on similarity of the feature of interest¹⁹. Arbelaitz et al. (2013) showed that an inclusion of 10% noise on time series data, resulted in reduction of 33% of the quality of a clustering³¹. The filters applied on the input EEG data of the Cluster Tool were not sufficient to filter out all noise. The EEG was filtered with a 2Hz highpass filter, and Hanning window was applied on the four seconds segments. We recommended to add a 40Hz lowpass filter, and apply the Hanning window on a smaller EEG segment to get rid of all high-frequency noise

and artefacts.

The EEG that surrounds a spike or transient incorporates essential information for accurate detection of IEDs. The Cluster Tool reduces the EEG data to the single channel where Persyst P13 detected the spike, thereby reducing the information content of the input data. If we would use all EEG channels as input data, the cluster performance is likely to increase. However, this step would increase the dimension of the input data drastically.

The current choice of dividing the detections into brain region was found to be sub-optimal. This step might be redundant if all EEG channels are included. Information about the localisation of the IED will then be present in the data itself. This would allow for clusters to combine IEDs from electrodes which are currently in different brain regions, like for example F3 and F7. The information about the channel where the event was detected by Persyst could then serve as an extra feature used for the clustering. It is recommended to study to the feasibility of clustering EEG segments which include all channels.

The choice for an automatic detection algorithm has a major impact on the accuracy of Cluster Tool. The quality of the Cluster Tool can only be as accurate as the quality of the input it receives. We used Persyst P13 as detection algorithm, because its sensitivity of 43.9% was proven to be human-like, and it seemed to be the best detection algorithm which was available on the market. Clustering and visualisation of the detections were proposed to deal with the high FPR of 1.65/min. We found that clustering detections based on morphology and localisation is a very promising way of presenting the results of a detection algorithm. Recently, a novel algorithm for the detection of IEDs was presented, which states to reach a sensitivity of 47.4% with a false positive rate of 0.6/min³⁸. This algorithm was thereby able to reach a sensitivity which is slightly higher than the one of Persyst P13, but with a remarkable lower FPR. The implementation of a more specific detection algorithm should be considered.

6.1.2 Distance measure

The Squared Euclidean distance and the Dynamic Time Warping distance were chosen as distance measures to quantify the dissimilarity between EEG segments. Recall that SE is a non-elastic measure, whereas DTW is an elastic measure. The K-means calculates the cluster centroids based on the mean of all objects within that cluster and seeks to minimise the distance of all objects to its mean. No time warping is applied in the calculation of the cluster centroids, as this is always a non-elastic calculation. Therefore, the cluster centroid, as calculated by the K-means, does not provide a good representation of the cluster which is found based on the DTW distance. This is visualised in Figure 6.1 where the total Sum of Squared Error (SSE) for a clustering is plotted per iteration of the K-means. The SSE is equal to the sum of the distances of every object to its centroid. The K-means seeks to minimise the SSE and it is expected that the SSE decreases with every iteration. However, the usage of DTW distance results in an SSE which fluctuates and even increases in successive iterations of the K-means. This is caused by the way the mean is calculated for the objects in a cluster created based on DTW distance. Therefore, the regular K-means algorithm should not be used in combination with the DTW distance. Still, the DTW seemed to capture the morphology of the EEG waveforms better than the SE. In many applications for time series clustering, DTW performs better than Euclidean distances³⁹. An alternative for the regular K-means is called

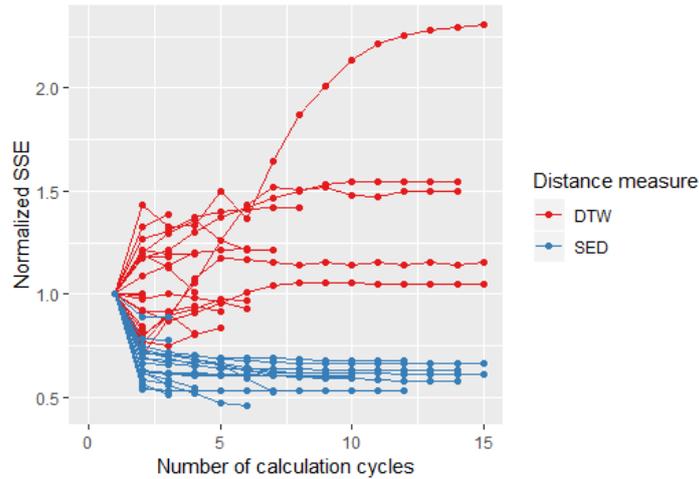


Figure 6.1: Normalized SSE over the iterations of the K-means algorithm

DTW Barycenter Averaging (DBA) K-means. DBA enables the calculation of an average time series that is consistent with DTW as similarity measure, by taking into account the time-warping⁴⁰. Figure 6.2 presents the difference in cluster average when calculated according to the mean and to the DBA. It shows that DBA captures the warping of a time series better. It is expected that the usage of DBA K-means will improve the cluster performance.

Besides from DTW, there are numerous other measures which can be used to capture the (dis)similarity between EEG segments^{20,39,42}. In our current algorithm, we only use the EEG segment originating from one channel. As stated before, it is recommended to experiment with the clustering of EEG segments containing information of multiple electrodes. Defining an appropriate similarity function, which captures the shape of a time series becomes even more challenging for multivariate time series³⁹. It is therefore proposed to compare different similarity measures and study which measure captures similarity of multivariate EEG segments best. Additionally, the performance of the cluster algorithm can be improved by combining different similarity measures. The perception value, as defined by Persyst, already contains valuable information about the likeliness of a detection to be truly epileptiform. Correlation of the perception value or the channel of detection can be used as additional similarity measure.

6.1.3 Clustering algorithm

The cluster algorithm used in the Cluster Tool was the K-means, one of the most widely used clustering methods. K-means is easy to implement, and the algorithm is known to perform well for finding non-overlapping spherically shaped clusters in small to medium-sized data sets¹⁸. However, the number of clusters needs to be pre-defined in the K-means algorithm. In this research, gap statistics are used to estimate the optimal number of clusters. Due to the lack of a labelled data set, it was not possible to evaluate if the number of clusters was estimated correctly. The need to pre-assign the number of clusters, while it is not known if the data contains any natural clusters, is a major drawback of the K-means algorithm¹⁹. Besides, it is uncertain if the underlying assumption of sphericity holds for complex multidimensional data.

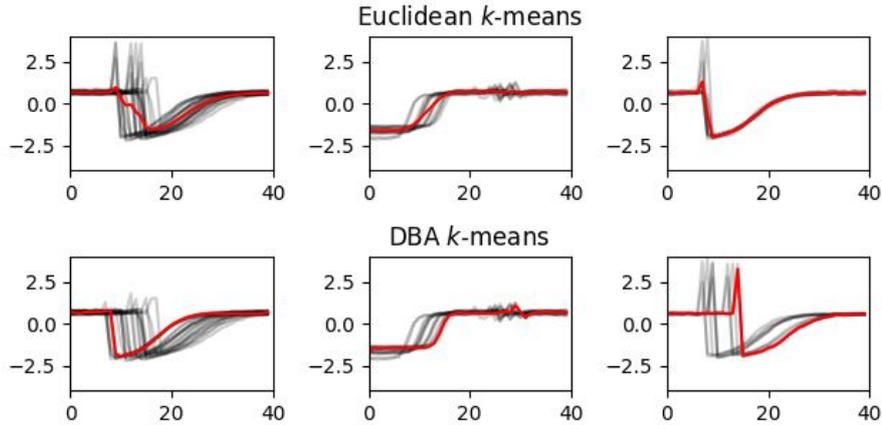


Figure 6.2: Difference between a euclidean and DBA k-means. The x-axis represent time, whereas the y-axis shows the amplitude. Adapted from *tslearn: A machine learning toolkit dedicated to time-series data* by Tavenard et al. (2017) Retrieved October 14, 2019, via https://tslearn.readthedocs.io/en/latest/auto_examples/plot_kmeans.html⁴¹

Various clustering techniques have been developed for time series clustering¹⁸. Figure 6.3 gives an overview of various well-known clustering algorithms and their impact on different data structures. Previous studies on clustering interictal epileptiform detections have used, among others, agglomerative hierarchical clustering, affinity propagation (AP), K-means and sequential clustering with subsequent template matching^{12,27,43,44}. A comparison between a K-means and AP algorithm for IED clustering showed that the AP algorithm outperformed the K-means, based on DTW distance²⁷. Another study on clustering noisy time series compared the K-means to DBSCAN and concluded better results for the DBSCAN⁴⁵. This encourages further research on the performance of other algorithms on the clustering of IEDs.

6.1.4 Dimension reduction

In this research, no dimension reduction techniques were applied. The data size of the EEG was already reduced by selecting a single channel and only a small segment of the EEG. However, dimension reduction offers the possibility to visualise high dimensional data in a low dimensional space, of two or three dimensions. This way, the structure of a high dimension data set can be studied, which is important to get an impression of the structure of the data. This information is valuable for the selection of a proper similarity measure and cluster algorithm, as is shown in Figure 6.3. T-distributed Stochastic Neighbor Embedding (t-SNE) is a technique for dimension reduction, which is capable of capturing the structure of the high dimensional data very well⁴⁶. By using t-SNE we could display EEG data two-dimensional. A data set with externally provided class labels would enable us to study which distance measure captures the natural structure of the data best. Subsequently, this labeled data set could be used to find a proper clustering algorithm. It is highly recommended to use t-SNE for the exploration of the structure of the high dimensional EEG data and to create a labeled data set. This method can assist in finding the optimal combination of similarity measure and cluster algorithm.

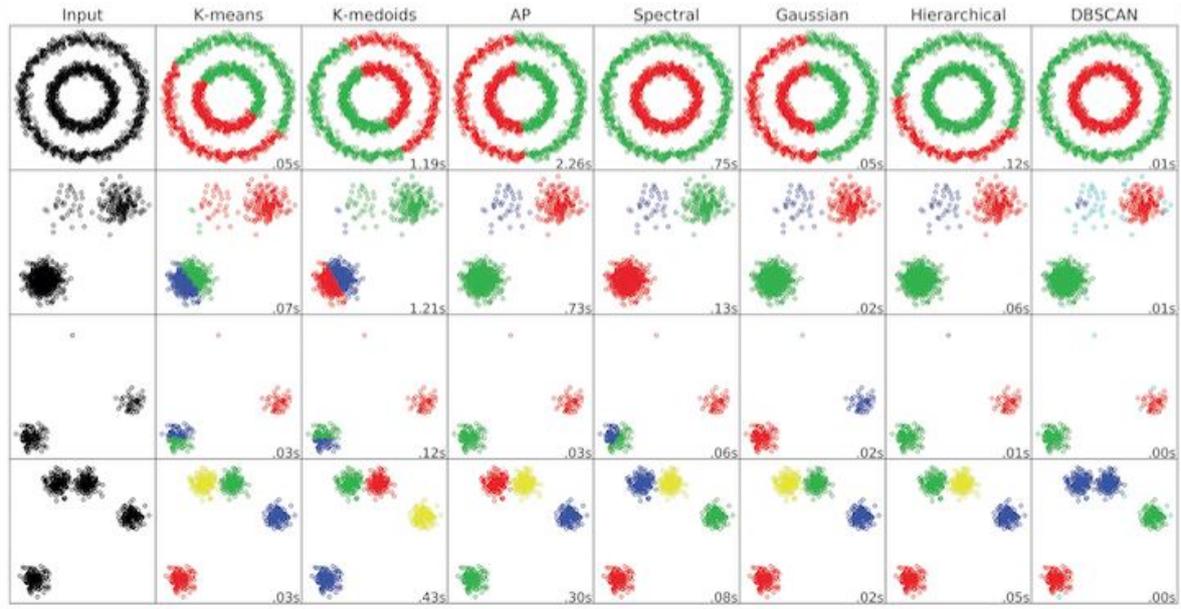


Figure 6.3: This table compares clustering results and run-times for six different algorithms (columns) applied across four input datasets (rows) with different structures (non-linear noisy circles, boxes with different densities, data with outliers, and close boxes). Colours represent cluster labels. Spectral clustering and DBSCAN beat other methods when the dataset has a circular structure and boxes with different densities (top two rows). For the second type of dataset, affinity propagation works better than others in most cases. The third dataset includes an outlier not far from the clusters, and Gaussian mixture model clustering does best at finding the outlier. The last row shows a dataset with four clusters, two of them are close to each other. All the algorithms do well, but the ways they partition the two close clusters are different. Figure retrieved from '*Exploring Patterns in Big Data Using ClusterEnG: A Clustering Engine for Genomics*' by Manjunath, M., Zhang, Y. (2017). Retrieved October 14, 2019, from <http://biomedicalcomputationreview.org/content/exploring-patterns-big-data-using-clustereng-clustering-engine-genomics>

6.2 The Graphical User Interface

We believe that the poor presentation of the results of automatic detection algorithms is one of the key factors that limits clinical adoption. Horsky et al. (2012) studied design principals for usable clinical decision support systems and found that the use of appropriate visual representation of clinical data belongs to the most important principals⁴⁷. The GUI forms an essential part of the Cluster Tool, adding a novel method for visualisation to the existing detection algorithm.

6.2.1 Comparison with previous literature

A few cases are known where clustering was used to improve the visualisation of automated detection for more efficient EEG evaluation. Wilson et al. (1999) introduced a hierarchical clustering algorithm to divide automatically detected transients into groups according to their localisation and morphology¹². The user interface allowed the EEG reader to open the hierarchical tree and display all levels of the tree until the nodes only contained single detections. Although this user interface allowed the EEG readers to quickly review the automated detections, it still required opening parts of the hierarchical tree. This action limits a direct interpretation of the results and decreases the reproducibility of the results.

Scherg et al. (2012) present a clustering algorithm designed to increase review time of interictal waveforms in long-term EEG by clustering⁴⁸. This algorithm is implemented in BESA Epilepsy 2.0, a commercially available software for the automatic detection of IEDs. An internal study at SEIN tested this software and found that the presentation of the IEDs through clustering and so-called hyper-clustering was very convenient. However, the detection algorithm did not perform well. The epileptic events found by BESA Epilepsy 2.0 corresponded only in 19% of the cases with events found through manual review and thus the software was rejected by the clinical experts¹³. This shows the importance of both, a sufficiently sensitive detection algorithm and a good user interface for the visualisation of the detections.

The Cluster Tool is based on a sufficiently sensitive detection algorithm. Moreover, the design of the GUI was experienced as effective, and simple. Simplicity of a user interface is one of the most important features for an effective presentation of clinical results^{47,49}. The GUI of the Cluster Tool was developed in a lean way, and valuable feedback has been gained during this process, about what features should be incorporated in the Cluster Tool to be clinically useful. Therefore, the GUI was designed in a way that corresponded with the clinical workflow of SEIN. Kawamoto et al (2005) showed that ‘automatic provision of decision support as part of the clinician workflow’ increases the success rate of implementing the system with 75%.

6.2.2 Implications for future work

In the current version of the Cluster Tool, the average waveform is taken as representative for the cluster. However, figure 6.2 already shows that in some cases, this is not the best way to visualise the main waveform of a cluster. Other representations, like the medoid, shape average prototypes using DTW, such as used by DBA, or a local search prototype as introduced by Hautamaki et al. (2008) might give a better representation, and hence a better visualisation of the cluster^{18,50}. The choice for the representation of the cluster should depend on the distance measure and clustering algorithm applied. Therefore, the possibilities

of different representations should be studied, while considering the distance measure and clustering algorithm.

The implementation of the timeline to indicate the temporal occurrence of the events in a specific cluster was found to be potentially very valuable. It enabled the user to give a quantitative measure of the occurrence of events in a cluster. Since not all events in a cluster were defined as truly epileptiform, the quantitative value of the timeline could not be used to give an accurate quantitative measure but only an indication of the temporal occurrence. However, as the performance of the clustering and the detection algorithm increases, the quantitative value of the timeline will increase as well. We recommend to keep using a timeline for the visualisation of the temporal occurrence, and improve the cluster algorithm to increase its quantitative value.

The division of the initial detections into brain regions was found to be sub-optimal to capture the information about the localisation of the IEDs. Furthermore, it is desired to include information about the surrounding electrodes in the cluster procedure. Currently, the GUI presents the clusters of each brain region per tab, and shows only the channels included in that specific brain region. Letting go of this division will result in a major change of the layout of the GUI. The most optimal way to structure the layout of the GUI will depend on the distance measure, the cluster algorithm and the representation method. Therefore, no further concrete recommendation can be given on the layout of the GUI. However, it is highly recommended to keep working according to the lean method and let potential users use prototypes to get feedback on how the GUI fits into the clinical workflow.

7 | Conclusion

This project presents the Cluster Tool, a digital prototype of a visualisation tool, consisting of a clustering algorithm and a Graphical User Interface (GUI), for efficient clinical interpretation of automatically detected interictal epileptiform discharges (IEDs). The GUI provided a visualisation of the IEDs, clustered according to their morphology and localisation. The GUI has been developed through an iterative design process which was based on the clinical workflow of EEG readers to eventually facilitate clinical implementation. The clinical performance evaluation and usability tests demonstrated the potential of the visualisation tool for improving EEG evaluation, but the clustering algorithm was too inaccurate to impact the clinical workflow positively. The usage of the tool resulted in remarkably similar clinical diagnoses in comparison to the EEG report. However, the clusters derived by the algorithm did not consistently meet the expectations of the neurologists, which decreased their trust in the performance of the tool and caused them to spend time on manually checking detections within clusters. We expect that an improvement of the clustering algorithm can provide a visualisation of the IEDs that complies with clinical expectations. After the implementation of a sufficiently accurate clustering algorithm, the designed prototype will enable a faster, reproducible and more objective method for EEG evaluation, ready for clinical implementation.

Bibliography

1. W. media centre World Health Organization, “Epilepsy [Fact Sheet],” 2019.
2. R. S. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. H. Cross, C. E. Elger, J. Engel Jr, L. Forsgren, J. A. French, M. Glynn, D. C. Hesdorffer, B. Lee, G. W. Mathern, S. L. Moshé, E. Perucca, I. E. Scheffer, T. Tomson, M. Watanabe, and S. Wiebe, “Ilae official report: A practical clinical definition of epilepsy,” *Epilepsia*, vol. 55, no. 4, pp. 475–482, 2014.
3. M. van Putten, “The EEG in epilepsy,” in *Essentials of Neurophysiology; Basic Concepts and Clinical Applications for Scientists and Engineers* (Springer, ed.), ch. Electroenc, Mairdumont GmbH & Co. Kg, 2009.
4. C. P. Panayiotopoulos, “Epileptic seizures and their classification,” in *A Clinical Guide to Epileptic Syndromes and their Treatment*, vol. 14, pp. 21–63, London: Springer London, 2010.
5. M. L. Scheuer, A. Bagic, and S. B. Wilson, “Spike detection: Inter-reader agreement and a statistical Turing test on a large data set,” *Clinical Neurophysiology*, vol. 128, no. 1, pp. 243–250, 2017.
6. S. B. Wilson and R. Emerson, “Spike detection: a review and comparison of algorithms,” *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, vol. 113, pp. 1873–1881, 2002.
7. C. N. Joshi, K. E. Chapman, J. J. Bear, S. B. Wilson, D. J. Walleigh, and M. L. Scheuer, “Semiautomated Spike Detection Software Persyst 13 Is Noninferior to Human Readers When Calculating the Spike-Wave Index in Electrical Status Epilepticus in Sleep,” *Journal of Clinical Neurophysiology*, vol. 35, pp. 370–374, sep 2018.
8. J. J. Halford, “Computerized epileptiform transient detection in the scalp electroencephalogram: Obstacles to progress and the example of computerized eeg interpretation,” *Clinical Neurophysiology*, vol. 120, no. 11, pp. 1909 – 1915, 2009.
9. W. Webber and R. P. Lesser, “Automated spike detection in EEG,” *Clinical Neurophysiology*, vol. 128, pp. 241–242, 2017.
10. S. Jawad, J. Oxley, W. Yuen, and A. Richens, “The effect of lamotrigine, a novel anticonvulsant, on interictal spikes in patients with epilepsy,” *British Journal of Clinical Pharmacology*, vol. 22, pp. 191–193, aug 1986.
11. J. J. Halford, M. B. Westover, S. M. LaRoche, M. P. Macken, E. Kutluay, J. C. Edwards, L. Bonilha, G. P. Kalamangalam, K. Ding, J. L. Hopp, A. Arain, R. A. Dawson, G. U. Martz, B. J. Wolf, C. G. Waters, and B. C. Dean, “Interictal Epileptiform Discharge Detection in EEG in Different Practice Settings,” *Journal of Clinical Neurophysiology*, vol. 35, no. 5, p. 1, 2018.
12. S. B. Wilson, C. A. Turner, R. G. Emerson, and M. L. Scheuer, “Spike detection ii: automatic, perception-based detection and clustering,” *Clinical Neurophysiology*, vol. 110, no. 3, pp. 404 – 411, 1999.
13. F. Spijkerboer, “A practical comparison of automatic detection software for interictal spikes in long-term EEG recordings at SEIN.” 2018.

14. N. Kane, J. Acharya, S. Beniczky, L. Caboclo, S. Finnigan, P. W. Kaplan, H. Shibasaki, R. Pressler, and M. J. A. M. V. Putten, "A revised glossary of terms most commonly used by clinical electroencephalographers and updated proposal for the report format of the EEG findings . Revision 2017," *Clinical Neurophysiology Practice*, vol. 2, pp. 170–185, 2017.
15. J. Jin, J. Dauwels, S. Cash, and M. B. Westover, "SpikeGUI: software for rapid interictal discharge annotation via template matching and online machine learning," *Conference proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, vol. 2014, pp. 4435–4438, 2014.
16. S. Winesett, S. R. Benbadis, and S. Florida, "Which Electroencephalogram Patterns Are Commonly Misread as Epileptiform ?," *US Neurology*, vol. 4, no. 2, pp. 62–66, 2008.
17. J. Jing, J. Dauwels, T. Rakthanmanon, E. Keogh, S. Cash, and M. Westover, "Rapid annotation of interictal epileptiform discharges via template matching under Dynamic Time Warping," *Journal of Neuroscience Methods*, vol. 274, pp. 179–190, dec 2017.
18. T. Warren Liao, "Clustering of time series data - A survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.
19. S. Aghabozorgi, A. Seyed Shirshorshidi, and T. Ying Wah, "Time-series clustering - A decade review," *Information Systems*, vol. 53, pp. 16–38, 2015.
20. V. Kumar, P.-N. Tan, and M. Steinbach, "Cluster Analysis: Basic Concepts and Algorithms," in *Introduction to Data Mining*, ch. Chapter 8, 2005.
21. A. Sarda-Espinosa, "Comparing time-series clustering algorithms in r using the dtwclust package," 2017.
22. D. J. Berndt and J. Clifford, "Using dynamic time warping to find patterns in time series," in *KDD Workshop*, 1994.
23. P. Esling and C. Agon, "Time-series data mining," *ACM Comput. Surv.*, vol. 45, pp. 12:1–12:34, Dec. 2012.
24. S. P. Lloyd, "Least Squares Quantization in PCM," *IEEE Transactions on Information Theory*, vol. 28, pp. 129–137, 1982.
25. D. Barbe, A. Debant, and X. Shang, "Time series clustering." 2016.
26. R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, vol. 63, no. 2, pp. 411–423, 2001.
27. J. Thomas, J. Jin, J. Dauwels, S. S. Cash, and M. B. Westover, "Clustering of interictal spikes by dynamic time warping and affinity propagation," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, vol. 2016-May, pp. 749–753, 2016.
28. D. Arthur and S. Vassilvitskii, "k-means ++ : The Advantages of Careful Seeding," in *SODA '07: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms.*, pp. 1027–1035, 2007.
29. K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success," *BMJ*, vol. 330, no. 7494, p. 765, 2005.
30. E. Ries, *The lean startup : how today's entrepreneurs use continuous innovation to create radically successful businesses*. New York: Crown Business, 2011.
31. O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona, "An extensive comparative study of cluster validity indices," *Pattern Recognition*, vol. 46, no. 1, pp. 243–256, 2013.
32. C. C. Aggarwal, "An Introduction to Cluster Analysis," in *Data Clustering: Algorithms and Applications*, ch. 1, pp. 1–23, CRC Press, 2013.

33. Q. Zhao, *Cluster Validity in Clustering Methods*. PhD thesis, University of Eastern Finland, 2012.
34. H. Haverkort, "Writing a report on experiments with algorithms," tech. rep., Eindhoven University of Technology, 2013.
35. M. N. Khairul baharein, "Case Study : A Strategic Research Methodology," *American journal of applied Sciences*, vol. 5, no. November, 2008.
36. C. M. Barnum, *Usability Testing Essentials: Ready, Set...Test!* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 1st ed., 2010.
37. T. Lagerlund, "Manipulating the Magic of Digital EEG: Montage Reformatting and Filtering," *Neurodiagnostic Journal*, vol. 40, pp. 121–136, jun 2000.
38. M. C. Tjepkema-Cloostermans, R. C. de Carvalho, and M. J. van Putten, "Deep learning for detection of focal epileptiform discharges from scalp EEG recordings," *Clinical Neurophysiology*, 2018.
39. D. Kotsakos, G. Trajcevski, D. Gunopulos, and C. C. Aggarwal, "Time-Series Data Clustering," in *Data Clustering: Algorithms and Applications*, ch. Chapter 15, pp. 357–380, CRC Press, 2013.
40. F. Petitjean and P. Gançarski, "Summarizing a set of time series by averaging: From Steiner sequence to compact multiple alignment," *Theoretical Computer Science*, vol. 414, no. 1, pp. 76–91, 2012.
41. R. Tavenard, J. Faouzi, and G. Vandewiele, "tslearn: A machine learning toolkit dedicated to time-series data," 2017. <https://github.com/rtavenar/tslearn>.
42. S. Lhermitte, J. Verbesselt, W. Verstraeten, and P. Coppin, "A comparison of time series similarity measures for classification and change detection of ecosystem dynamics," *Remote Sensing of Environment*, vol. 115, pp. 3129–3152, dec 2011.
43. A. Ossadtchi, R. M. Leahy, J. C. Mosher, N. Lopez, and W. Sutherling, "Automated interictal spike detection and source localization in MEG using ICA and spatial-temporal clustering," *Proceedings - International Symposium on Biomedical Imaging*, vol. 2002-Janua, no. February, pp. 785–788, 2002.
44. P. Van Hese, B. Vanrumste, H. Hallez, G. J. Carroll, K. Vonck, R. D. Jones, P. J. BONES, Y. D'Asseler, and I. Lemahieu, "Detection of focal epileptiform events in the EEG by spatio-temporal dipole clustering," *Clinical Neurophysiology*, vol. 119, no. 8, pp. 1756–1770, 2008.
45. L. Kirichenko, T. Radivilova, and A. Tkachenko, "Comparative analysis of noisy time series clusterin," in *Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Systems (COLINS-2019). Volume I: Main Conference, Kharkiv, Ukraine, April 18-19, 2019.*, pp. 184–196, 2019.
46. L. van der Maaten and G. Hinton, "Visualizing Data using t-SNE," *Journal of Machine Learning Research*, vol. 9(Nov), pp. 2579–2605, 2008.
47. J. Horsky, G. D. Schiff, D. Johnston, L. Mercincavage, D. Bell, and B. Middleton, "Interface design principles for usable decision support: A targeted review of best practices for clinical prescribing interventions," *Journal of Biomedical Informatics*, vol. 45, no. 6, pp. 1202–1216, 2012.
48. M. Scherg, N. Ille, D. Weckesser, A. Ebert, A. Ostendorf, T. Boppel, S. Schubert, P. G. Larsson, O. Henning, and T. Bast, "Fast evaluation of interictal spikes in long-term EEG by hyper-clustering," *Epilepsia*, vol. 53, no. 7, pp. 1196–1204, 2012.
49. K. Miller, D. Mosby, M. Capan, R. Kowalski, R. Ratwani, Y. Noaiseh, R. Kraft, S. Schwartz, W. S. Weintraub, and R. Arnold, "Interface, information, interaction: A narrative review of design and functional requirements for clinical decision support," *Journal of the American Medical Informatics Association*, vol. 25, no. 5, pp. 585–592, 2018.
50. V. Hautamaki, P. Nykanen, and P. Franti, "Time-series clustering by approximate prototypes," *Proceedings - International Conference on Pattern Recognition*, pp. 1–4, 2008.
51. D. J. Ketchen and C. L. Shook, "The application of cluster analysis in strategic management research: An analysis and critique," *Strategic Management Journal*, vol. 17, no. 6, pp. 441–458, 1996.

A | Appendix

A.1 Gap statistics

To estimate the optimal number of clusters, we applied gap statistics as proposed by Tibshirani et al. (2001)²⁶. Gap statistics compares the within cluster cohesion W_k with the expected cohesion. First, Tibshirani et al. define the within cluster sum of squares W_k for a range of values of k as:

$$W_k = \sum_{n=1}^k \frac{1}{2|C_n|} \sum_{x,y \in C_n} D_{SE}(x,y) \quad (\text{A.1})$$

where $|C_n|$ is the number of objects within cluster n . $D_{SE}(x,y)$ is the sum of pairwise SE distances for all points in cluster n . Note that this is a different way of calculating the cohesion than used in Appendix A.3, where the within cluster sum of squares is not divided by the number of objects within the clusters.

Figure A.1b. shows the value of W_k for different number of clusters k . The optimal number for k is the number where the value of W_k does not decrease significantly when adding a cluster. It can be observed that this is the case for two clusters. This method is known as the elbow method. However, the optimal number of clusters cannot always be identified clearly when using the elbow method⁵¹. Therefore, Tibshirani et al. suggest to standardize the graph of $\log(W_k)$ by comparing it with its expectation under an appropriate null distribution²⁶. The optimal number of clusters is then estimated by the value of k for which the function $\log(W_k)$ falls the farthest under the reference function. Therefore they define:

$$Gap_n(k) = \hat{E}_n^* \{ \log(W_k) \} - \log(W_k) \quad (\text{A.2})$$

with \hat{E}_n^* the expectation from the reference distribution under a sample size n .

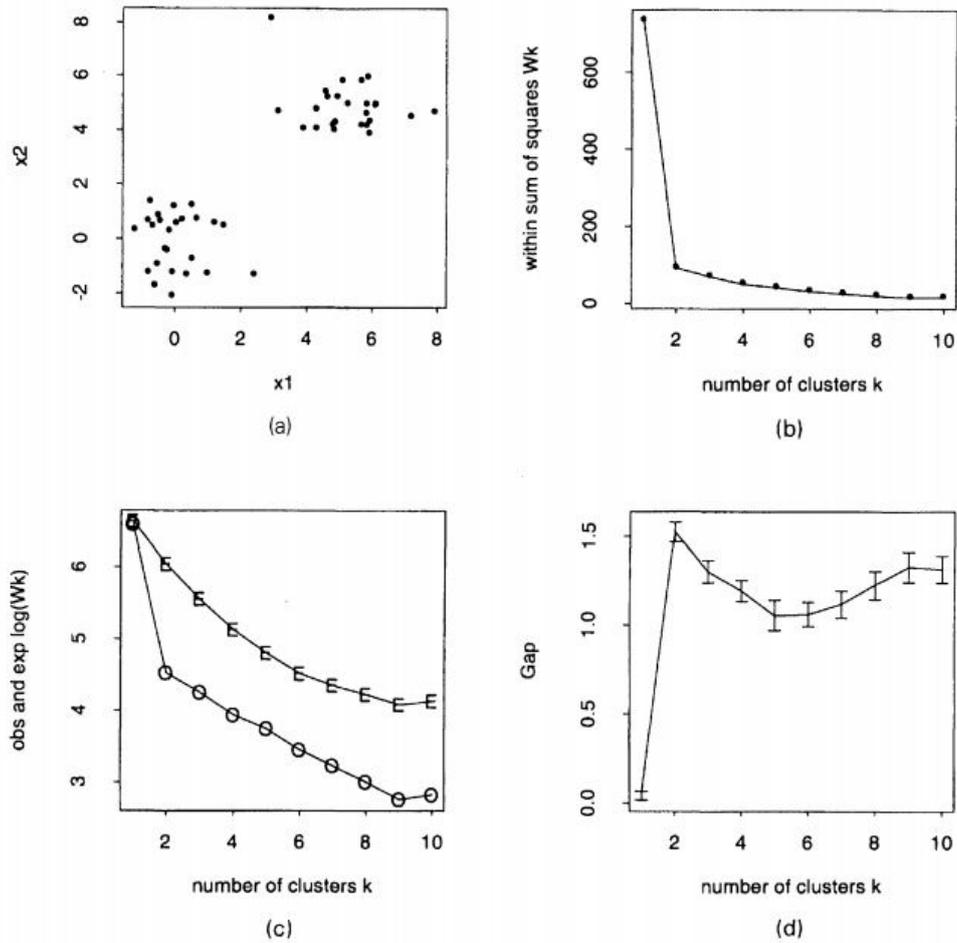


Figure A.1: Results for a two-cluster example: **a)** data; **b)** within sum of squares function W_k ; **c)** functions $\log(W_k)$ (O) and $\hat{E}_n^*\{\log(W_k)\}$ (E); **d)** gap curve. Figure retrieved from 'Estimating the number of clusters in a data set via the gap statistic' by R. Tibshirani, G. Walther, and T. Hastie (2001)²⁶

A.2 Flowchart data preparation

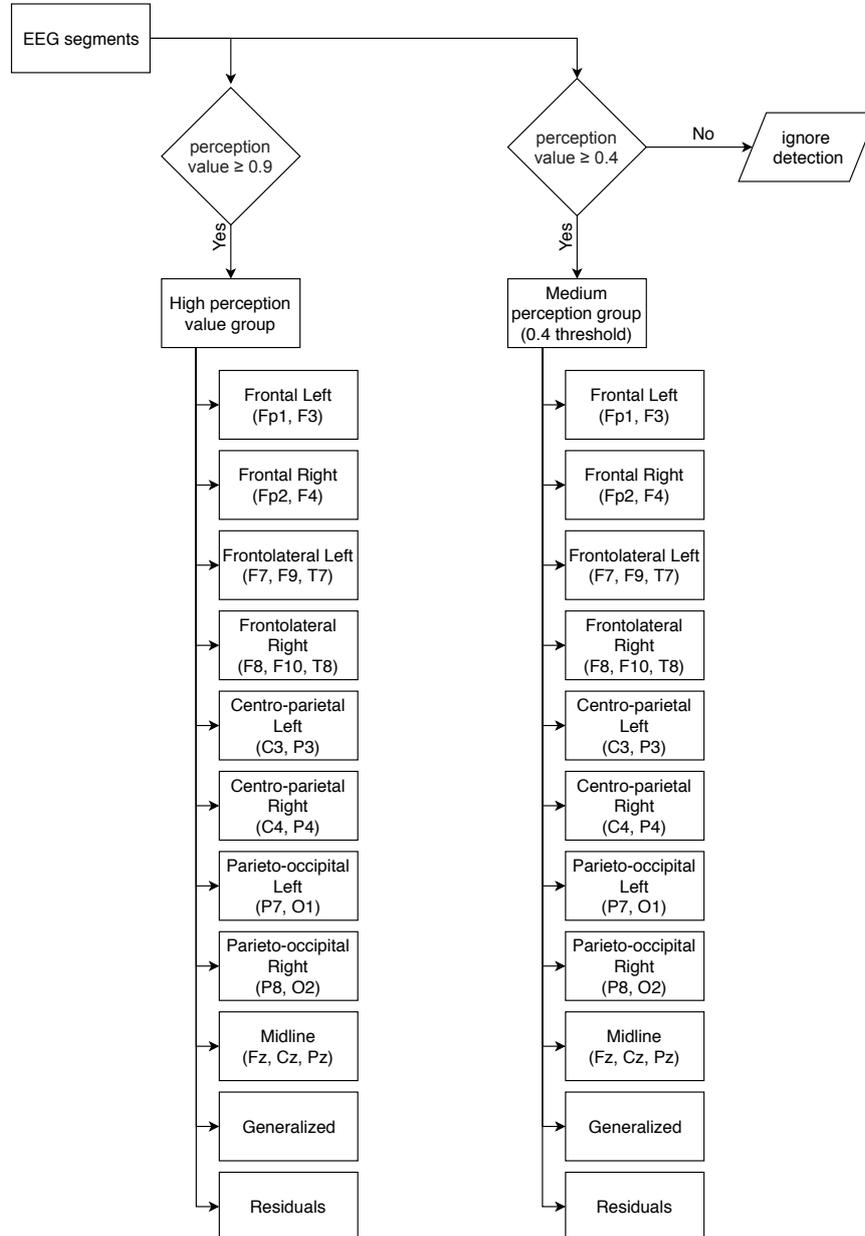


Figure A.2: Flowchart of all selection steps applied on the EEG segments. First, a separation based on the perception value is made, which divides the group into two groups. Secondly, the groups are separated per brain region

A.3 Internal validation

The usage of two internal validation measures was explored during this project: the sum of squared errors (SSE) to measure the cohesion and the between the sum of squared errors (BSS) to measure the separation of the cluster results.

A.3.1 Calculation of the SSE and BSS

The cohesion of a cluster is defined as the sum of the distances between the cluster centroid and the data points assigned to that cluster. Since we use the squared Euclidean distance, the sum of distances equals the sum of squared errors (SSE). The overall cluster cohesion is the sum of the SSE, as shown in equation A.3.

$$\text{Total SSE} = \sum_{i=1}^k \sum_{x \in c_i} D(x_m, c_i) \quad (\text{A.3})$$

with k the number of clusters, m the number of objects in a cluster with c_i the cluster centroid of the i^{th} cluster and D the SE distance as defined in equation ??.

The cluster separation is defined as the distance of a cluster centroid to the overall centroid, weighted according to the number of objects in the cluster. Since we use the squared euclidean distance, the cluster separation is equivalent to the between cluster sum of squares (BSS). The overall cluster separation is the sum of the BSS over all clusters, as shown in equation A.4.

$$\text{Total BSS} = \sum_{i=1}^k m_i D(C, c_i) \quad (\text{A.4})$$

with c_i the cluster centroid within the i^{th} cluster, m_i the number of EEG trials in that cluster, C the overall mean of all centroids and D the SE distance as defined in equation ??

A.3.2 Interpretation of the results

The use of K-means clustering with DTW distance has a large negative impact on the BSS. Figure A.3a shows the BSS of clusterings with SE and DTW. The BSS based on DTW clustering is very low. The DTW distance uses a non-linear approach to calculate the distance between two time series. However, the K-means algorithm takes the mean of a cluster as its centroid and does not apply time warping. By calculating the mean of objects which were defined as similar based on time warping effects, we create a mean signal which loses a lot of information about the true waveforms of the original objects. Therefore, the mean is not a good representative of a cluster created with DTW and should not be used as cluster centroid. Since the BSS is based on the mean as centroids of the clusters, it does not capture the true separation of a clustering created with K-means and DTW.

The SSE is also affected by the use of DTW in a K-means clustering algorithm. This can be seen in Figure 6.1. However, the SSE of based on the DTW distance is comparable to the

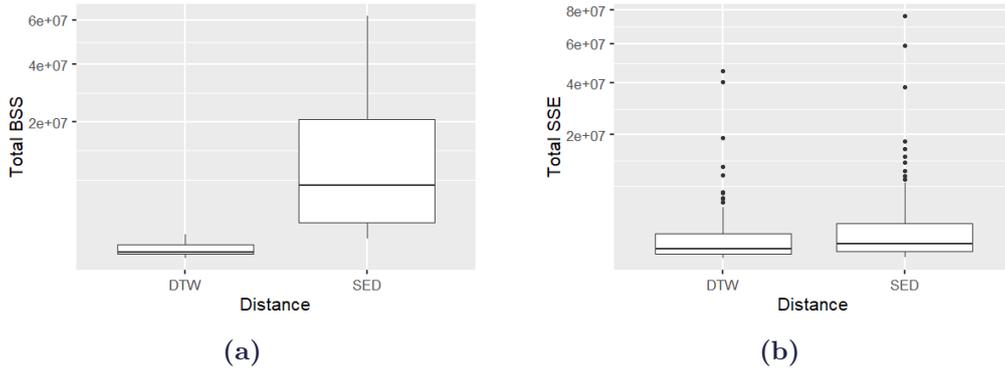


Figure A.3: Boxplots of the between sum of squared errors (BSS) and sum of squared errors (SSE) of the clustering based on the SE and DTW distance. **a)** BSS of clusterings with SE and DTW distance. The BSS represents the separation of a clustering. Note that the y-axis applies a squared scale. The DTW distance shows a remarkable lower BSS than the SE distance **b)** SSE of clusterings with SE and DTW distance. The SSE represents the cohesion of a clustering. Note that the y-axis applies a squared scale. The SSE of the DTW and SE distance are comparable, although the DTW distance seems to result in a slightly lower SSE.

SSE of the SE distance (see Figure A.3b). This is because the DTW distance is not restricted to the linear warping path, but can find smaller distances by comparing sample i to sample $j + 1$. By definition, the DTW distance will always find smaller distances.

A.4 Learn curve of the cluster tool

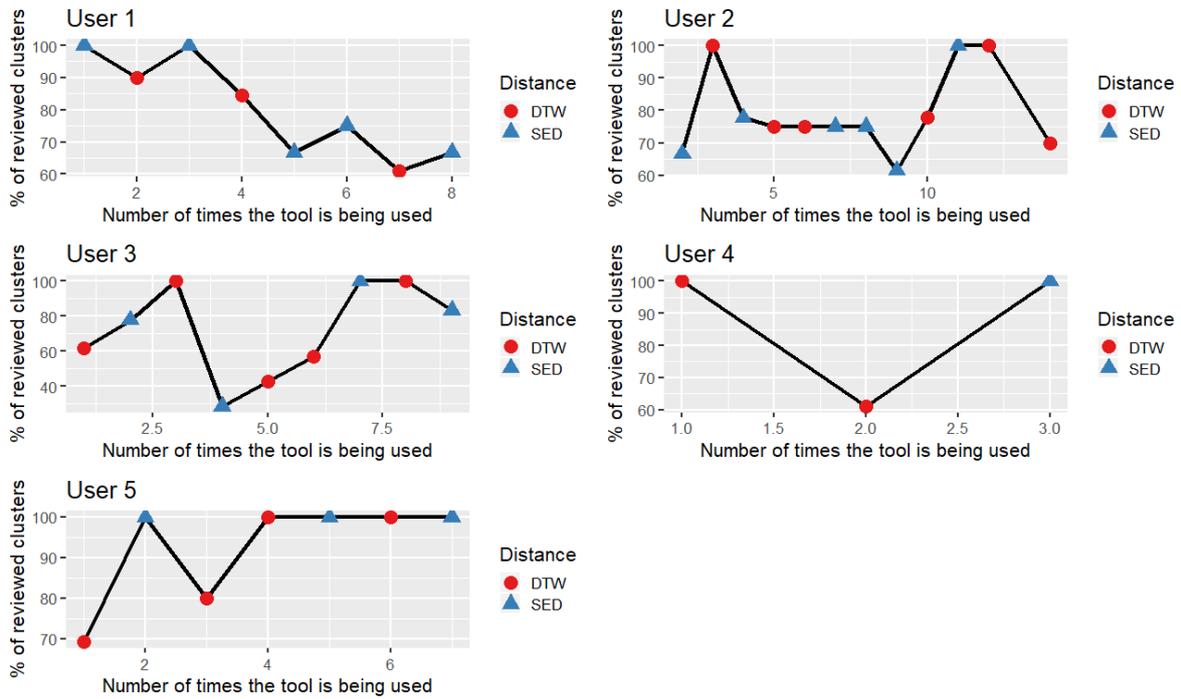


Figure A.4: Learn curve of the Cluster Tool. This figure presents the percentage of clusters that are inspected in detail through use of the *'view cluster'*-button, in regards to the number of times a specific test participant uses the tool. The coloured symbols represent whether the test participant used the cluster tool with the Squared Euclidean (SE) distance (blue pyramid), or with the Dynamic Time Warping (DTW) distance (red circle). It was expected to observe a trend (learning curve), where an increase of the number of times a test participant used the tool would result in a decrease of the percentage of clusters which were reviewed on a detailed level. However, no trend could be observed.

Table A.1: Comparison of the conclusion as described in the EEG rapport with the conclusion made by using the cluster tool. The conclusion is split into information regarding waveform, location, temporal occurrence and diagnosis.

EEG ID	Method	nr. of detections	nr. of clusters	Waveforms	Localisation	Temporal info	Notes	Diagnosis
5	EEG rapport			Sharp transients, sharp waves	right occipital > temporo-parietal		Both, isolated and in series	Epilepsy
	SE	9	1	Sharp transients, sharp waves	right parieto-occipital		Only one cluster, with only nine detections	Epilepsy
	DTW	9	1	Sharp transients, sharp waves	right parieto-occipital		Only one cluster, with only nine detections	Epilepsy
6	EEG rapport			Polyspike-and-slow-waves, spike-wave complexes	Frontal left and right	increase during sleep, sometimes series up to 6s		Epilepsy
	SE	149	8	Spike-wave complexes	Frontal left and right	increase during the night,	the series were in the same cluster as the isolated spikes and detected only by scrolling through the individual detections	Epilepsy
	DTW	149	10	Polyspike-and-slow-waves, spike-wave complexes	Frontal left and right	increase during the night,	separate cluster with mainly eye artefacts, series were in the same cluster as isolated detections	Epilepsy
4	EEG rapport			Sharp transients, series of sharp transients which could be wicket spikes	Frontotemporal Left >right	more clear during sleep		PNES, Abnormalities of uncertain significance
	SE	137	9	Sharp transients, no specific epileptiform discharges	Frontotemporal left, sometime also right	increase during sleep		Unclear meaning of sharp transients

Table A.1: Comparison of the conclusion as described in the EEG rapport with the conclusion made by using the cluster tool. The conclusion is split into information regarding waveform, location, temporal occurrence and diagnosis.

EEG ID	Method	nr. of detections	nr. of clusters	Waveforms	Localisation	Temporal info	Notes	Diagnosis
	DTW	137	10	Sharp transients, no specific epileptiform discharges	Frontotemporal left, sometime also right	increase during sleep		Uncertain significance of sharp transients
3	EEG rapport			Sharp waves/ sharp transients	Frontotemporal - parieto-occipital		Light functional neurological disorder frontotemporal - parieto-occipital (Left>Right) seen as erratic activity in the alpha band, sometimes mixed with a sharp wave.	Signs of a light functional neurological disorder
	SE	186	9	Sharp waves and sharp transients, often in series	Frontotemporal, parieto-occipital Right >Left			multifocal sharp transients uncertain diagnosis more information needed
	DTW	186	9	Sharp waves and sharp transients, often in series	Frontotemporal, parieto-occipital Right >Left			multifocal sharp transients uncertain diagnosis more information needed
8	EEG rapport			Sharp-and-slow-wave complexes	Frontotemporal left >>right	increase during sleep		Epilepsy
	SE	1644	13	Sharp-waves, Sharp-and-slow-wave complexes	Frontal, Frontotemporal left >>right	increase in the night		Epilepsy
	DTW	1644	13	Sharp-waves, Sharp-and-slow-wave complexes	Frontal, Frontotemporal left >>right	increase in the night		Epilepsy
9	EEG rapport			Sharp waves, sharp-and-slow waves	Frontotemporal right >left	Sometimes, mainly during relaxation/ light sleep	some sharp-and-slow waves, could be WHAMS	Epilepsy

Table A.1: Comparison of the conclusion as described in the EEG rapport with the conclusion made by using the cluster tool. The conclusion is split into information regarding waveform, location, temporal occurrence and diagnosis.

EEG ID	Method	nr. of detections	nr. of clusters	Waveforms	Localisation	Temporal info	Notes	Diagnosis
	SE	101	4	(small) sharp-and-slow waves, sharp waves	Frontotemporal right >left	increase in the night and early morning	Difficult to make a conclusion without further context	uncertain diagnosis
	DTW	101	6	(small) sharp-and-slow waves, sharp waves	Frontotemporal right >left	increase in the night and early morning	Difficult to make a conclusion without further context	uncertain diagnosis
15	EEG rapport			sharp waves, sharp-and-slow-waves, sharp transients	frontotemporal left>>right			Epilepsy and unspecific focal functional disorder
	SE	2093	6	sharp waves, sharp-and-slow-waves	frontotemporal left>>right			Epilepsy
	DTW	2093	6	sharp waves, sharp-and-slow-waves	frontotemporal left>>right			Epilepsy
16	EEG rapport			polyspikes-and-slow-waves, spike-and-slow waves	Frontal	increase during sleep	isolated and in series	Epilepsy
	SE	127	9	polyspikes-and-slow-waves, spike-and-slow waves	Frontal bilateral	increase during the night	polyspikes-and-slow-waves can be seen in cluster overview, series can be seen in the individual detections	Epilepsy
	DTW	127	9	polyspikes-and-slow-waves, spike-and-slow waves	Frontal bilateral	increase during the night	polyspikes-and-slow-waves can be seen in cluster overview, series can be seen in the individual detections	Epilepsy

Table A.1: Comparison of the conclusion as described in the EEG rapport with the conclusion made by using the cluster tool. The conclusion is split into information regarding waveform, location, temporal occurrence and diagnosis.

EEG ID	Method	nr. of detections	nr. of clusters	Waveforms	Localisation	Temporal info	Notes	Diagnosis
17	EEG rapport			sharp-and-slow waves	Frontal right			Epilepsy
	SE	389	8	sharp-waves	Frontal right >left	Cluster tool detects only during the night where IEDs are found clustered, many spikes within 20 minutes	the slow wave component is difficult to distinguish in the MATLAB viewer	Epilepsy
	DTW	389	7	sharp-waves	Frontal right >left		the slow wave component is difficult to distinguish in the MATLAB viewer	Epilepsy
7	EEG rapport			spike-and-slow-waves, some (not evident) polyspikes	Frontal bilateral interictal Ri>Le, but ictal Le>Ri	Only in wake, more often right after awakening		Epilepsy
	SE	613	17	spike-and-slow waves polyspikes	Frontal Le>ri and P8	mainly during the night and early morning	Ictal series are also detected (left >right)	Epilepsy
	DTW	613	17	spike-and-slow waves polyspikes	Frontal Le>ri and P8	mainly during the night and early morning	Ictal series are also detected (left >right)	Epilepsy
19	EEG rapport			Sharp-and-slow-wave, Wicket spikes	Frontotemporal right. wicket Le>Ri		series of spikes which are seen in the clusters, are interpreted as wicket spikes	PNES and epilepsy
	SE	291	11	sharp-and-slow waves, series of spikes, could be small sharp spike or wicket spikes	Frontotemporal right sss more left			Epilepsy

Table A.1: Comparison of the conclusion as described in the EEG rapport with the conclusion made by using the cluster tool. The conclusion is split into information regarding waveform, location, temporal occurrence and diagnosis.

EEG ID	Method	nr. of detections	nr. of clusters	Waveforms	Localisation	Temporal info	Notes	Diagnosis
	DTW	291	8	sharp-and-slow waves, series of spikes, could be small sharp spike or wicket spikes	Frontotemporal right sss more left			Epilepsy
20	EEG rapport							Normal EEG
	SE	3	1					Normal EEG
	DTW	3	1					Normal EEG
18	EEG rapport			Sharp-and-slow-waves, sharp waves	Frontal/ frontotemporal Left>Right			Epilepsy
	SE			Sharp-and-slow-waves, sharp waves	Frontal/ frontotemporal Left>Right			Epilepsy
	DTW			Sharp-and-slow-waves, sharp waves	Frontal/ frontotemporal Left>Right			Epilepsy
21	EEG rapport			sharp transients/ sharp waves	Frontotemporal right / left			light unspecific focal functional disorder
	SE	262	9	sharp transients/ sharp waves	Frontotemporal right / left			Unclear meaning of sharp transients
	DTW	262	9	sharp transients/ sharp waves	Frontotemporal right / left			Unclear meaning of sharp transients
24	EEG rapport			sharp waves, sharp-and-slow-waves	maximum P7 and surrounding	occurs a couple of time during sleep		Epilepsy
	SE	84	7	sharp waves, sharp-and-slow-waves	Mainly P7, sometimes F7 F9 and F4	only during the night		Epilepsy

Table A.1: Comparison of the conclusion as described in the EEG rapport with the conclusion made by using the cluster tool. The conclusion is split into information regarding waveform, location, temporal occurrence and diagnosis.

EEG ID	Method	nr. of detections	nr. of clusters	Waveforms	Localisation	Temporal info	Notes	Diagnosis
	DTW	84	7	sharp waves, sharp-and-slow-waves	Mainly P7, sometimes F7 F9 and F4	only during the night		Epilepsy
25	EEG rapport			low frequent activity (delta)	temporal left	during sleep		non-epileptic, unclear meaning
	SE	49	3	sharp transients (dubious)	frontotemporal mostly right	in the night		no specific epileptiform discharges
	DTW	49	3	sharp transients (dubious)	frontotemporal mostly righth	in the night		no specific epileptiform discharges
26	EEG rapport			sharp transients	temporal	during sleep		non-epileptic, no clear abnormalities
	SE	24	4	sharp transients	frontotemporal	in the night		unclear meaning
	DTW	24	4	sharp transients	frontotemporal	in the night		unclear meaning
22	EEG rapport			sharp waves sharp-and-slow-waves	frontotemporal left	mainly during relaxation and doezel, in long series long series are recognized very good in the timeline		epilepsy
	SE	539	6	sharp-and-slow waves	frontotemporal left			epilepsy
	DTW	539	6	sharp-and-slow waves	frontotemporal left			epilepsy
23	EEG rapport			Sharp waves, (poly-) spikes-and-slow-waves	temporal left and right, frontal bilateral	in series up to 2s, increase in the morning		epilepsy
	SE	512	9	Sharp waves, (poly-)spike-and-slow-waves	temporal left and right, frontal left and right	series can be seen when inspecting individual detections		epilepsy
	DTW	512	9	Sharp waves, (poly-)spike-and-slow-waves	temporal left and right, frontal left and right	series can be seen when inspecting individual detections		epilepsy

Table A.1: Comparison of the conclusion as described in the EEG rapport with the conclusion made by using the cluster tool. The conclusion is split into information regarding waveform, location, temporal occurrence and diagnosis.

EEG ID	Method	nr. of detections	nr. of clusters	Waveforms	Localisation	Temporal info	Notes	Diagnosis
27	EEG rapport			sharp transients	temporal left >right	mainly during light sleep		no evident epileptiform detections, but suspicious waveforms. Aspecifieke focale functie stoornis
	SE	109	6	sharp transients	parieto-temporal			suspicious graphic elements, not enough information for a clinical conclusion
	DTW	109	6	sharp transients	parieto-temporal			suspicious graphic elements, not enough information for a clinical conclusion
28	EEG rapport			sharp transients	parieto-occipital right		physiological rhythms parieto-occipital are remarkably high in amplitude	no evident epileptiform detections
	SE	67	4	sharp transients	parieto-occipital right			no evident epileptiform detections
	DTW	67	4	sharp transients	parieto-occipital right			no evident epileptiform detections
29	EEG rapport			sharp waves, sharp-and-slow-waves	frontotemporal right	occurs often, only in sleep		epilepsy
	SE	1104	8	sharp waves, sharp-and-slow-waves	frontotemporal right	only during the night		epilepsy
	DTW	1104	8	sharp waves, sharp-and-slow-waves	frontotemporal right	only during the night		epilepsy
30	EEG rapport			sharp waves, (poly-)spike-and-slow-waves	frontal left and bilateral synchronous	sharp waves sporadical during wake, PGC low frequent occuring in sleep		epilepsy

Table A.1: Comparison of the conclusion as described in the EEG rapport with the conclusion made by using the cluster tool. The conclusion is split into information regarding waveform, location, temporal occurrence and diagnosis.

EEG ID	Method	nr. of detections	nr. of clusters	Waveforms	Localisation	Temporal info	Notes	Diagnosis
	SE	22	4	(poly-)spike-and-slow-waves	frontal F3, Fz, sometimes F4	only 22 detections		epilepsy
	DTW	22	4	(poly-)spike-and-slow-waves	frontal F3, Fz, sometimes F4	only 22 detections		epilepsy