Identifying observer-based features for seizure detection in 4-electrode EEG in the paediatric ICU

Master thesis for Technical Medicine

Sanne Lange

Brain Neural Network is a photograph by Jesper Klausen/science Photo Library.

IDENTIFYING OBSERVER-BASED FEATURES FOR SEIZURE DETECTION IN 4-ELECTRODE EEG IN THE PAEDIATRIC ICU

Sanne, Lange

Student number : 4500407

24 april 2023

Thesis in partial fulfilment of the requirements for the joint degree of Master of Science in

Technical Medicine

Leiden University ; Delft University of Technology ; Erasmus University Rotterdam

Master thesis project (TM30004 ; 35 ECTS)

Dept. of Biomechanical Engineering, TUDELFT

May 2022 – May 2023

Supervisor(s):

Prof. dr. ir. Alfred C. Schouten

Dr. Robert van den Berg

Thesis committee members:

Prof. dr. ir. Alfred C. Schouten, TU Delft (chair)

Dr. Robert van den Berg, Erasmus MC

Dr. Maayke Hunfeld, Erasmus MC

Universiteit Leiden

TU Delft Delft University of Technology

Erasmus Universiteit Rotterdam

Acknowledgements

Dear readers,

I am delighted to present my master thesis on the topic of seizure identification in EEG. During the past year, I focused my research on the identification of seizures by creating EEG-based features related to epilepsy and identifying their ability to predict the presence of epilepsy in 4-electrode EEGs of children in the paediatric intensive care unit.

I wish to express my gratitude to Professor Alfred Schouten and Robert van den Berg for their guidance throughout this master thesis. Robert, your endless enthusiasm has been a constant source of motivation. My gratitude also extends to Maayke Hunfeld for her contribution as a committee member.

My appreciation also goes to Karla for providing me with the opportunity to observe intra-operative neuromonitoring surgeries and for sharing her extensive knowledge and expertise in this field.

I would like to thank my family and friends for being there for me and encouraging me throughout this journey. To Pam, Femke, and Fred for the long days spent together in the study rooms. A special mention goes to Daniek, Manon and Marit, thank you for the coffee breaks, helpful discussions, and lending an ear when I needed to vent.

Lastly, I would like to express my appreciation to Jasper for his countless hours spent helping me with reading, programming, and creating figures. Your dedication and support have been invaluable.

Sanne Lange
24-04-2023

Abstract

Epilepsy has been reported in 10-40% of children in the paediatric intensive care unit (PICU). Amplitude-integrated electroencephalography (aEEG), often used as neuromonitoring in the PICU, has some limitations and as a result, caretakers in the PICU may find it challenging to interpret aEEG. This may lead to uncertainty during diagnosis and increases the need for the assistance of the neurophysiology department, a costly EEG and a late diagnosis which might result in irreversible harm.
The aim of this exploratory study was to identify and evaluate observer-based features as stand-alone classifiers and combined in a random forest classifier, that could aid in the accurate classification of seizures in paediatric critical care patients with 4-electrode EEG.
Several features achieved an AUC above 0.60, with the lower border of the aEEG signal yielding the highest AUC of 0.79. The best-performing features maintained their classifying ability when tested on a larger, independent dataset. The random forest classifier did not provide better results.
Our results show that various observer-based features can aid in the identification of seizures in 4-electrode EEG. These additional features could increase the accuracy and decrease the delay and uncertainty in diagnosing epilepsy in aEEG by PICU staff, reducing the need for full setup EEG monitoring and improving patient outcomes in paediatric critical care.

Contents

Nomenclature

Abbreviations	Definitions
ACF	Autocorrelation function
aEEG	Amplitude-integrated electroencephalography
ASR	Artifact subspace reconstruction
AUC	Area under the curve
C3-C4	Central left and right electrode positions
cEEG	Continuous electroencephalography
EDF	European Data Format
EEG	Electroencephalography
FPR	False positive rate
ICU	Intensive care unit
METC	Medical Ethical Testing Committee (dutch: Medisch Ethische Toetsings Commissie)
NPV	Negative predictive value
P3-P4	Parietal left and right electrode positions
PICU	Paediatric intensive care unit
PLV	Phase locking value
PPV	Positive predictive value
Q1-Q3	Interquartile range
RA	Alpha1-total band power ratio
RAD	Alpha1-delta band power ratio
RB	Beta-total band power ratio
RBD	Beta-delta band power ratio
RD	Delta-total band power ratio
RF	Random forest
ROC	Receiver operation curve
RS	Alpha2-total band power ratio
RSD	Alpha2-delta band power ratio
RT	Theta-total band power ratio
RTD	Theta-delta band power ratio
SEF	Spectral edge frequency
TPR	True positive rate
TsEntropy	Tsallis entropy

1

Introduction

Epilepsy is a neurological disorder characterized by recurrent seizures, which are episodes of abnormal electrical activity in the brain. It is one of the most common neurological disorders, affecting around 50 million people worldwide [1]. The incidence of epilepsy in children ranges from 0.04% to 0.19% [2]. Children in the pediatric intensive care unit (PICU) are highly vulnerable to brain injuries, including seizures. Electrographic seizures have been reported in 10-40% of children in the PICU and emergency department who underwent neuromonitoring [3]. Often, the seizures were non-convulsive [4–6]. Clinical examination is a common method to diagnose seizures, however, non-convulsive epileptic seizures may go unnoticed. Additionally, the often sedated and ventilated children in intensive care may not display typical signs of epilepsy [7–10]. This could leave epilepsy cases untreated, leading to severe and irreversible harm. Therefore, a neuromonitoring modality is highly desirable for the detection of brain injury [11–14].

Electroencephalography (EEG) is a technique used to record the electrical activity of the brain's surface by placing electrodes on the scalp. It measures the electrical potentials of cortical neuronal dendrites near the surface of the brain, which are generated by the influx and efflux of ions during the synaptic transmission of pyramidal cells. This flux generates an extracellular field potential, forming the basis of potentials recorded on a scalp electrode [15]. As EEG is sensitive to changes in brain activity [16], it is the common modality for long-term neuromonitoring. Indications for continuous EEG (cEEG) in the PICU include the detection of seizures, the identification of the underlying cause of altered mental states, and the monitoring after acute or traumatic brain injury or after resuscitation [17].

Although continuous EEG is the gold standard in the ICU [18], some limitations call for an alternative. Using an EEG requires the expertise of both a clinical technician to perform the recording and a trained neurologist for interpretation. These resources may not always be available, the procedures are time-consuming, and restrict the possibility to monitor in real-time. Amplitude-integrated electroencephalography (aEEG) offers a simplified and continuous alternative. It is a transformed and time-compressed EEG signal using two to four electrodes. It was developed as a neuromonitoring method for adults after resuscitation [19] and is now popularly used in neonates [20, 21]. Use on the PICU has been increasing [22], but still lacks guidelines and scientific evaluation [23]. Further limitations, specific to epilepsy, include the fact that short seizures may go unnoticed in the time-compressed signal and that artefacts might mimic epileptic activity. These difficulties can lead to limited accuracy and uncertainty in identifying true seizures. As a result, caretakers in the PICU may find it challenging to interpret aEEG and often call in the help of neurophysiologists and cEEG when in doubt, highlighting the need to make it easier to identify abnormalities in brain activity in aEEG.

One possible solution to improve the diagnosis of epilepsy directly by PICU staff, is to provide additional information beyond the aEEG to visualise the behaviour of the brain activity in a patient. This supplementary information can be based on the characteristics of epileptic activity normally observed by neurophysiologists in EEG recordings. Epileptic seizures result from the malfunctioning of the brain's electrophysiological system, which leads to sudden excessive electrical discharges in

groups of neurons. This phenomenon is caused by hypersynchronous activity of neurons, resulting in higher voltages, rhythmicity, and changes in frequency, morphology, symmetry, and localisation in EEG recordings. Incorporating these observer-based characteristics into the neuromonitoring tool could help reduce uncertainty and improve the accuracy of seizure detection.

Numerous studies have investigated various techniques to identify seizures in EEG and aEEG signals [24–28]. However, many studies have focused on adults, did not show optimal performances or have created intricate models, often with the basis of a machine learning algorithm. It is important to address that, to accept and adhere to new technologies, physicians have to be able to understand what the result of an algorithm is based on [29]. Therefore, it is necessary to find an approach that can identify epileptic activity in children in the PICU, that can be easily understood and interpreted by the PICU staff.

The goal of the current study is to identify and evaluate observer-based features that have added value besides the standard aEEG signal. It is hypothesised that features similar to characteristics of an EEG used by clinical staff to interpret epileptic activity could help in classifying ictal periods. We intend to make the first steps in creating a real-time automatic epilepsy detection model that will aid in the identification and diagnosis of epilepsy in children in the PICU using 4-electrode EEG, improving the care of this vulnerable patient group.

2

Literature background

This chapter contains a brief overview of the previously conducted literature study on the relevant and most promising observer-based EEG features with respect to epilepsy identification implemented and evaluated in this study.

2.1. Seizures on EEG

A seizure is the result of an imbalance between excitatory and inhibitory interactions of the cortical neurons. During a seizure, there is a hypersynchronous discharge of neurons that can spread from a specific region of the cortex to neighbouring areas. The start of a seizure is marked by high-frequency bursts of action potentials and hypersynchronisation of a neuronal population, resulting in a spike discharge on the EEG. Seizure propagation occurs when there is sufficient activation to recruit surrounding neurons [30]. The manifestation of epileptic activity in EEGs can vary greatly among patients and among seizures. Epileptic activity always disrupts the background, may evolve in frequency distribution, changes morphologically over the course of the seizure, shows higher voltages and is often very rhythmic. These abnormalities will appear in the affected area and therefore there is a correlation between proximate regions, while a difference occurs between distant regions. Based on insights from clinical neurophysiologists, standardised terminology and previous literature, a set of observer-based features have been identified;

1. High amplitude (especially of high frequencies)
2. Evolution in frequency
3. Rhythmicity
4. Evolution in morphology
5. Symmetry between electrodes close by
6. Asymmetry between hemispheres
7. Evolution in location/spreading

2.2. Feature extraction methods

The features mentioned in the previous section are possible candidates to be used to distinguish seizures in EEG recordings. In the literature study, many possible extraction methods of these features have been discussed, of which the main findings are summarised in Table 2.1. This table states whether the extraction method fulfilled the requirements of being intuitive (understandable for clinical staff), showed classifying abilities in paediatric EEGs and was used in real-time and in EEGs with limited electrodes. The most promising methods for this study, based on these requirements, are highlighted in the table and are described in the following paragraphs.

Table 2.1: Characteristics of epileptic activity in EEGs and the reviewed corresponding features

Characteristic	**Feature**	**Intuitive**	**Tested for:**		
			Paediatrics	Real time	Limited electr.
High voltage	Amplitude	✓	✗	✗	✗
	Spectral analysis	✓	~ [31]	✗	✗
	aEEG	~	✓[22]	✓[22]	✓
Evolution in	Spectral analysis	✓	✗	✗	✗
frequency	SEF	✓	~ [32]	✗	✗
Rhythmicity	Maximum voltage	✓	✗	✗	✗
	Autocorrelation	✓	~ [33]	✓[34]	✗
	Entropy	~	✓[35, 36]	✗	✓[37]
Evolution in	Derivatives	✓	~ [31]	~	✗
morphology	Wavelet functions	✗	✓[38]	✓[39, 40]	✗
	Morphological filters	✗	✗	✗	✗
Spatial	Coherence	~	✓[41, 42]	✓[34]	✓[41–43]
(a)symmetry	Phase locking value	~	✓[41, 42]	✗	✓[41, 42, 44]
	Phase lag index	✗	~ [45]	✗	✗

aEEG = Amplitude-integrated electroencephalography, SEF = Spectral edge frequency, Blue are the suggested methods to be used in the project.

Amplitude

The simultaneously firing neurons can result in high amplitude signals, often in higher frequency ranges. One way to determine if a signal contains epileptic seizures based on this phenomenon is by looking at the mean amplitude. If the mean amplitude at a certain point deviates from a previously established baseline or a continuous baseline, it could indicate a seizure-like event. Mean amplitude has been found to be one of the most effective biomarkers of epileptic activity [31].

aEEG

Amplitude-integrated electroencephalogram (aEEG), is a compressed and filtered EEG trend that enables long-term monitoring of brain function across patients of different ages. The lower and upper borders of the aEEG represent the variations in respectively the minimum and maximum amplitudes in the EEG. A normal continuous pattern is defined by a lower border trace above 5μV and an upper border trace above 10μV [46]. A seizure can be recognised by a slow rise in lower margin amplitude, a small bandwidth (both due to higher voltages present in the signal [47]) and an abrupt drop of the lower border to baseline. Previous studies have used varieties of this method in a variety of detection models [48–51].

Spectral analysis

One way of differentiating seizure vs non-seizure segments based on the high voltage characteristic is by comparing the power of frequency bands per segment EEG, using the power spectral density. Because of the higher amplitude activity, the power in the different frequency bands will generally increase during a seizure. Furthermore, the frequency distribution can also be used to identify the evolution in frequency during a seizure in which there is a shift of signal power from the lower to higher frequency bands [52]. Both frequency band powers and ratios have shown good results in previous studies identifying epileptic activity [31, 53].

Autocorrelation

Autocorrelation is a signal-processing technique that measures the correlation between a signal and a delayed version of itself. It is commonly used to identify repeating patterns in a signal and could therefore find the rhythmicity related to seizure in an EEG signal. For a signal, the autocorrelation function provides a value between 1 for perfect correlation, 0 for no correlation, and -1 for negative correlation. In the context of epilepsy detection, a higher autocorrelation function indicates higher rhythmicity in the EEG signal, which may indicate the presence of epileptic activity. Autocorrelation has shown promising results when used in observer-based decision trees [34] and combined observer-based automatic detection [33].

Entropy

Many methods for feature extraction to capture rhythmic oscillations in EEG signals have been limited by their reliance on linear methods. However, EEG signals are nonlinear and complex, making them well-suited for analysis using nonlinear methods. One such method is entropy, which measures the randomness and irregularity of a signal. During a seizure, the rhythmic expression of the signal becomes more regular and less chaotic, resulting in a lower entropy value. Signal analysis has numerous forms of entropy, many of which have been used in the detection of seizures in EEG signals. Shannon entropy is a simple form of entropy commonly used in signal analysis and measures the degree of uncertainty or randomness in a signal and has shown some promising results as a classifier of seizures [54]. A variation on Shannon entropy is Tsallis entropy, with which one can vary the influence of outliers on the overall measure. High entropy in an EEG signal implies a high degree of disorder, which means that there are multiple individual signals contributing to the final signal. A low entropy value indicates a more simple build-up of the signal. During a seizure, many neurons fire at the same rate and frequency causing the final signal caught by the electrode to be fairly simple leading to a low value of entropy and a more predictable signal.

Coherence

Coherence is the linear correlation between two signals calculated in the frequency domain, representing the oscillatory coupling based on amplitude and phase synchronisation. At the beginning of a seizure, when the focus of activity expands to adjacent brain regions, there is an increase in coherence between signals recorded by neighbouring electrodes. In cases where the seizure is localized to one hemisphere, the coherence between signals recorded by electrodes on the opposite hemisphere decreases. As the seizure progresses, incorporating more and more brain regions, the overall coherence between signals increases. Several studies have proven that coherence, the overall, inter- or intra-hemispheric values, can be used for distinguishing ictal and non-ictal EEGs [34, 41–43, 55].

Phase locking value

The phase locking value (PLV) determines if the phase difference between two signals is relatively constant, providing the degree of phase locking [56]. It is similar to the coherence but normalised to the amplitude, so it more strictly identifies phase synchronisation [57]. The PLV is a method that has been used in seizure detection for children in critical care [41], with limited EEG channels [41, 42, 44], with different frequency decomposition methods and as a feature in combination with a machine learning algorithm [58].

3

Methods

We performed a retrospective study on the EEG data from children in the PICU to find features identifying epileptic seizures. Firstly, individual methods were computed to establish the performance of individual features in identifying a seizure. Afterwards, a random forest model used these features to create a combined automated model.

3.1. Data

Two sets of data were used in this study, a small optimisation set and a larger test set. For both datasets, the EEG data was selected from patients in the database of the Erasmus MC Sophia Children's Hospital from 2018 to 2020. The optimisation set consisted of relevant EEGs of ten children with different types of hard-to-identify epileptic activity encountered during clinical practice. The test set consisted of EEGs that were recorded within 24 hours after the initial aEEG recording had stopped running. The EEGs were recorded in BrainRT™with 11 to 19 electrodes, based on the international standard configuration of the 10-20 system. Experienced neurophysiologists or lab technicians classified the events during the recordings through visual inspection. These formed the ground truth to which the results of the model's classifications were compared. All recordings were performed as standard procedure in clinical evaluation of the patients and no information can lead back to the individual patient. It has been reviewed and approved by the Medical Ethical Testing Committee (METC) (nWMO-MEC-2021-0145).

3.2. Preprocessing

Both sets underwent several preprocessing steps before feature calculation, as visualised in Figure B.2. For the creation of the aEEG features, the same data underwent different preprocessing steps. First, the raw data was exported in EDF+ (European Data Format) from BrainRT™. Further analysis was done using MATLAB™[59] with EEGLab™[60] and Fieldtrip™[61] plugins. All channels, other than the standard channels for aEEG (C3, C4, P3, P4), were removed from the EDF datasets. The signals were referenced to the average of the entire set, while for the creation of the aEEG features, as later explained, the montage was changed to C3-P3 and C4-P4, resulting in two channels instead of four as this is standard practice. The EEG signals were then filtered with a Hamming window sinc FIR filter, with the edges set to 1.5Hz and 25Hz. The data was downsampled to 128Hz and cleaned for artefacts using the $clean\_artifacts$ function within the EEGLAB toolbox that applies real-time artefact subspace reconstruction (ASR) to detect and reject high-amplitude activity including eye-blinks, muscle activity and sensor motion [62, 63]. Each second of the data was labelled as 'seizure' or 'regular' based on the visual inspection of the neurophysiologists.

3.3. Feature extraction

Various epilepsy feature extraction methods were previously reviewed and are summarised in Table 2.1. These features were based on characteristics that are used to identify ictal activity in clinical practice by neurophysiologists. The high-lighted extraction methods have been developed in this study. All

Table 3.1: Epileptic EEG features and corresponding methods of extraction

Characteristic	Extraction method	Features
High voltage	Amplitude	Maximum higher envelope, Maximum variance
	aEEG	Maximum lower envelope, Minimum variance
Evolution in frequency	Delta/Theta/Alpha/ Alpha2/Beta/Total power band	Maximum
	Theta-delta/Alpha-delta/ Alpha2-delta/Beta-delta power ratio	Maximum
	Delta-total/Theta-total/Alpha-total/ Alpha2-total/Beta-total power ratio	Maximum
Rhythmicity	Autocorrelation peak-ratio	Peak-ratio closest to 1, Minimum variance
	Tsallis entropy	Minimum, Mean variance
Spatial (a)symmetry	Coherence	Maximum, (Minimum)
	Phase locking value	Maximum, (Minimum)

aEEG = Amplitude-integrated electroencephalography, PPV = Positive predictive value, () indicate features only included in the random forest classifier

extraction methods followed a similar setup. For the signals of the four electrodes, the main variable of interest was calculated for every 3-second window with a 2-second overlap. Further analysis, such as computing the envelope or the variance of the variable, was also conducted. In the following chapters, each extraction method will be specified in more detail. Table 3.1 list the features corresponding to each extraction method.

Amplitude

For every electrode, the mean higher envelope and the moving variance of the EEG data with a 3-second window were calculated with a 2-second overlap. To combine the four electrodes into one feature, we took the maximum absolute value of the higher envelope and the maximum variance of these four electrodes. Both parameters were used as two separate features in the individual method analysis and the random forest classifier.

aEEG

To determine aEEG-based features, firstly the EEG signal was manipulated to create an aEEG signal using an existing open-source code by Vesoulis et al. [64]. The signal underwent filtering and rectification, followed by envelope extraction using a Butterworth filter. Next, we subjected the envelope to scaling so that values up to 10mV were displayed linearly, while the remaining values followed a logarithmic scale. This resulted in two 2-channel aEEG signals. The maximum value for the lower envelope and the minimum variance of the lower envelope of the two aEEG signals were calculated every three seconds, with a 2-second overlap.

Frequency analysis

For each 3-second epoch, we calculated the powers of five frequency bands (delta (0.5-4Hz), theta (4-8Hz), alpha (8-12Hz), alpha2 (12-16Hz), and beta (16-25Hz)) for each electrode for every second. The power spectral density for each of these frequency bands was determined using Welch's method with a Hamming window segmenting the data. Other parameters that were computed using these frequency

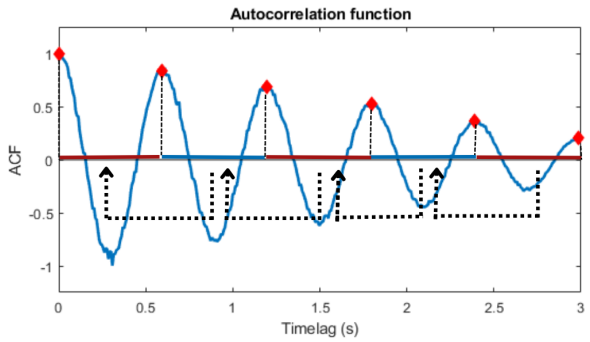


Figure 3.1: Example of the autocorrelation function of an EEG epoch. The peaks are detected and the intervals between the peaks are compared to each other. The more closely the comparison is to 1, the more rhythmic the signal is.

band powers include the ratio between the five band powers and the total power, as well as the ratio between four band powers and the delta power. For the final features, the maximum value of the four electrodes was taken and set as parameters.

Autocorrelation

Our approach was developed based on similar methods presented in previous papers [33, 34, 65]. The autocorrelation function was calculated for each epoch of three seconds, with a 2-second overlap in the EEG signal for the four electrodes. We then identified the peaks of this function and calculated the mean distance between all the peaks within the 3-second window. A peak-ratio close to one indicates a periodic signal which might hint at the occurrence of a seizure. An example is shown in Figure 3.1. The more the ratio differs from one, the less rhythmic the signal is. The average and variance of the peak-ratio were both considered as individual features for detecting epileptic activity in the EEG signal and as features for the combined random forest classifier.

Entropy

The entropy measure used in this study was derived from Shannon entropy, namely Tsallis entropy. We determined the Tsallis entropy measure using the wavelet entropy measure. This first splits a signal into wavelet coefficients using wavelet decomposition, after which the entropy measures are determined using the normalised wavelet coefficients as the probability distributions. The more rhythmic the signal, the less chaotic it tends to be. This indicates that epileptic activity should decrease the level of entropy.

Coherence

Just like the other methods, the coherence calculation was done for every 3-second window with a 2-second overlap. The coherence over all frequency bands was determined for all the different combinations of electrodes, so six channels in total. The maximum and minimum values of these six combinations per second were calculated. For the preliminary method analysis, the maximum coherence, as suggested by some of the previously mentioned studies [34, 41, 55] was used as its parameter. In the combined model analysis, the maximum and minimum coherence were both included in the search for the best combination of methods in the random forest classifier.

Phase locking value

The PLV for all frequency bands was calculated in the following way for the four electrodes separately. First, a Hilbert transform created the analytical signal of the EEG data, consisting of a signal between 0 and 1 with a real and imaginary part. Then, the angle was determined. For each pair of electrodes, six channels in total, the PLV was calculated for every second with a 3-second window. Like the coherence parameters, the maximum and minimum PLV between the electrode combinations were determined. For the individual method analysis, the maximum PLV was considered as the feature, while both parameters were included for the combined method analysis.

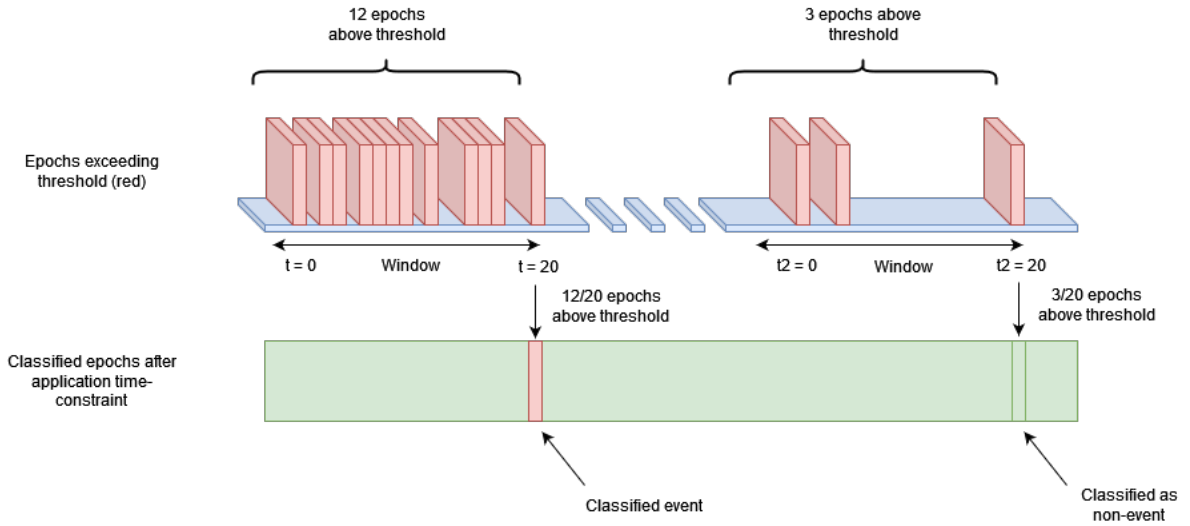


Figure 3.2: Overview of minimum window-sum condition, explaining its influence on the classification of events. In the left example, the feature exceeds the threshold in 12 time points within the 20-second window. This leads to the current epoch being classified as an event. The right example shows that the feature only exceeded the threshold at 3 points in the time window, thus leading to the current epoch being classified as a non-event.

3.4. Individual method analysis

We determined the optimal window length for all previously mentioned features by running all the feature extraction analyses for 1, 3, 5 and 10-second windows and comparing their distributions. The goal is to classify every second of EEG signal as either seizure or non-seizure for every feature and compare it to the classification of the visual inspection of the neurophysiologist. We needed three variables to do this; a threshold to which we compare the feature to determine if it classifies as seizure or not, a mean to base this threshold on and a window-sum condition to make the classification more robust and to prevent overestimating. This window-sum condition is defined as the minimal amount of points in a certain previous time window that has to be positive to classify the current epoch as a seizure. Figure 3.2 provides an example of a window-sum condition of 12 points exceeding the threshold within a window of 20 seconds. The threshold for each feature is based on the mean of the total EEG and a generalised variable c:

$$Threshold = c * \mu_{signal}.$$

Each feature was normalised per patient to the mean, to be able to compare the relative difference between seizure and non-seizure epochs. To systematically determine the optimal combination of a minimal amount of points during which the feature exceeds a threshold and time window (the minimum window-sum condition), receiver operating curves (ROC) curves were created for every feature, for the possible time windows of 5, 12, 20, 30, 40 and 60 seconds, with a minimum number of points exceeding the threshold of respectively $1/5^{th}$, $2/5^{th}$, $3/5^{th}$ and $4/5^{th}$. These time windows were empirically determined. Employing such a minimum window-sum condition will create a delay in the detection of the seizure by at least the minimum number of required positively identified points, and might miss shorter seizures more easily. However, it ensures a higher robustness of the classifier by not classifying every outlier as an event. We generated ROCs for all methods, which depict the trade-off between sensitivity and specificity with varying discriminatory thresholds. Per feature, the optimal window-sum condition corresponding to the best AUC was determined.

3.5. Statistical evaluation of individual features

The statistical analysis and performance evaluation of every method involved the calculation of median and interquartile ranges for each feature, followed by the application of the Mann-Whitney U test to determine the significance of the differences between the 'seizure' and 'non-seizure' groups (Appendix D). These differences or similarities were also visualised with violin plots. As stated before, the best AUC for the corresponding window-sum condition was searched. For these ROC curves, further performance

evaluation was done. The optimal threshold was determined by finding the point on the ROC curve with the highest combined score of sensitivity and specificity. The sensitivity, specificity, positive predictive value, negative predictive value, and accuracy corresponding to this optimal threshold were determined for both the test set and the optimisation set (Appendix D). Correlations between all features and the presence of events were calculated and visualised using the features calculated for the test set.

3.6. Random forest classifier

After evaluating the individual methods in seizure identification, all collected normalised parameters were used as input for the search for an optimal combination of these features in a random forest classifier. The classifier was built in Python 3.7 using the Scikit-learn toolbox [66]. To find the optimal combination of hyperparameters and train the model, we used the features of the ten EEGs from the optimisation dataset. First, the data was resampled to have 2000 data points per patient, with 1000 event epochs and 1000 non-event epochs, so that each patient had an equal part in training the data, which avoided overfitting on the longest EEG recordings or the patients with a high number of seizures. For smaller sets, duplicates were allowed. The resampled normalised data of the optimisation set was split into test and training data with a stratified 5-fold split, ensuring that all samples of individual patients were included in one group. The hyperparameters were optimised in a 10-iteration, 5-fold randomised search. The classifier was given various options on different hyperparameters to find the best combination fitted to the data (see Appendix H for more information on the hyperparameters). The classifier was able to use different criteria in splitting; entropy or gini, was given the option of bootstrapping and had a variety of options for the number. Furthermore, the depth of trees, how many branches one main tree could obtain and the minimal amount of samples that should result in another split were iterated in search of the best combination. The trained model was then validated on the larger set of features calculated for the 127 EEGs. To determine the performance of the classifiers, the AUC, sensitivity, specificity and accuracy were calculated for the five folds. The different features included in this classifier were checked for their importance in the model.

4

Results

The upcoming results section will begin by determining the optimal epoch length for calculating the features, followed by a comparison of feature value distributions between seizure and regular periods. Most importantly, the performance of the individual features as classifiers will be analysed and the correlations between features will be discussed. Additionally, the results of the random forest classifier will be addressed.

4.1. Patient data

The optimisation set consisted of ten EEG recordings of ten children and the test set of 127 EEG recordings of 115 children, with, due to the manner of data collection, an overlap of four patients with seven EEG recordings. Both subsets contained EEGs from patients of varying chronological ages. Many of the included patients were younger than one year, while a few patients included in the test set were up to 17 years of age. The variation in age was larger in the test set compared to the optimisation set. The plots in Appendix A present the variation in patient characteristics between the two sets of EEG recordings. Most of the older patients did not have electrographic seizures. A trend is visible showing an inverse correlation between the age of patients and the frequency of seizures. The optimisation set contained recordings of a total of 477.8 hours, of which almost 5% consisted of epileptic seizures. Great variability was found in the occurrence of seizures between patients and the duration of the epileptic activity (Table 4.1). 29% of the 127 recordings in the test set contained seizures. Notably, the duration of seizures ranged widely, more so in the test set than in the optimisation set, with some lasting up to 40 minutes, while others were as brief as a few seconds.

Table 4.1: Characteristics of EEG data in optimisation and test set.

Characteristic	**Optimisation set**	**Test set**
Number of patients	10	115
Males (%)	5 (50%)	67 (58%)
Age in days (median (Q1-Q3)), max	57 (26-117), 811	98 (36-495), 6452
Number of recordings	10	127
Duration of recordings (total (range))	477.82h (2.0h-142.9h)	3302.4h (0.16h-206.9h)
Recordings with seizures (%)	10 (100%)	37 (29.37%)
Amount of seizures (range)	600 (2-267)	1685 (0-267)
Percentage of seizure activity (total% (range))	4.87% (0.63- 40.8%)	2.19% (0-32.9%)
Duration of seizures (mean (range))	12.4m (2.57m-24.0m)	2.58m (0.05m-40.9m)

h = hours, m = minutes

4.2. Optimal epoch length

For all methods, the features were calculated using 1, 3, 5, and 10-second epochs. The distributions of all the features for the different epoch lengths are shown in Appendix C, Figures C.1a through C.1i. For the features maximum amplitude, higher envelope of aEEG, autocorrelation peak-ratio and Tsallis entropy, the distributions generated by different epoch lengths were found to be relatively consistent. Although longer epoch lengths did slightly reduce the distribution size and eliminated some outliers, the basic form of the distribution remained similar. In contrast, for other features such as ratios to delta and total power, coherence, and PLV, changing the epoch length resulted in substantial changes in the distribution shape. The use of a one-second epoch was deemed insufficient to produce reliable values, while epochs longer than three seconds did not significantly alter the distributions. Therefore, a 3-second window was chosen as the optimal length for all features. Additionally, to ensure that no data was missed due to the window length, a 2-second overlap was used between each window resulting in a feature value for every second.

4.3. Feature-value distribution

A first glimpse of the difference between the seizure (event) and non-seizure (non-event) groups in the distribution of the values per feature is portrayed in the violin plots created for each feature for the ten datasets. The most-promising features are shown in Figures 4.1 and 4.2 while the other results can be found in Appendix E. The 1% of the upper and lower outliers were removed in the plots for better visual effects. The plots show that the distributions of the values for seizures and non-seizures overlapped in many features, and no clear separation can be made. Plots for certain features, however, including the amplitude features, the frequency band power features, the theta-total ratio and especially the lower border of the aEEG signal, indicate that the event group had higher values compared to non-events. Conversely, minimum entropy and variance in autocorrelation peak-ratio show the opposite of slightly lower values for the seizure group, while the coherence and PLV feature displayed broader variation in value and the alpha-total and alpha2-total ratio show smaller distributions on both sides for the samples during seizures.

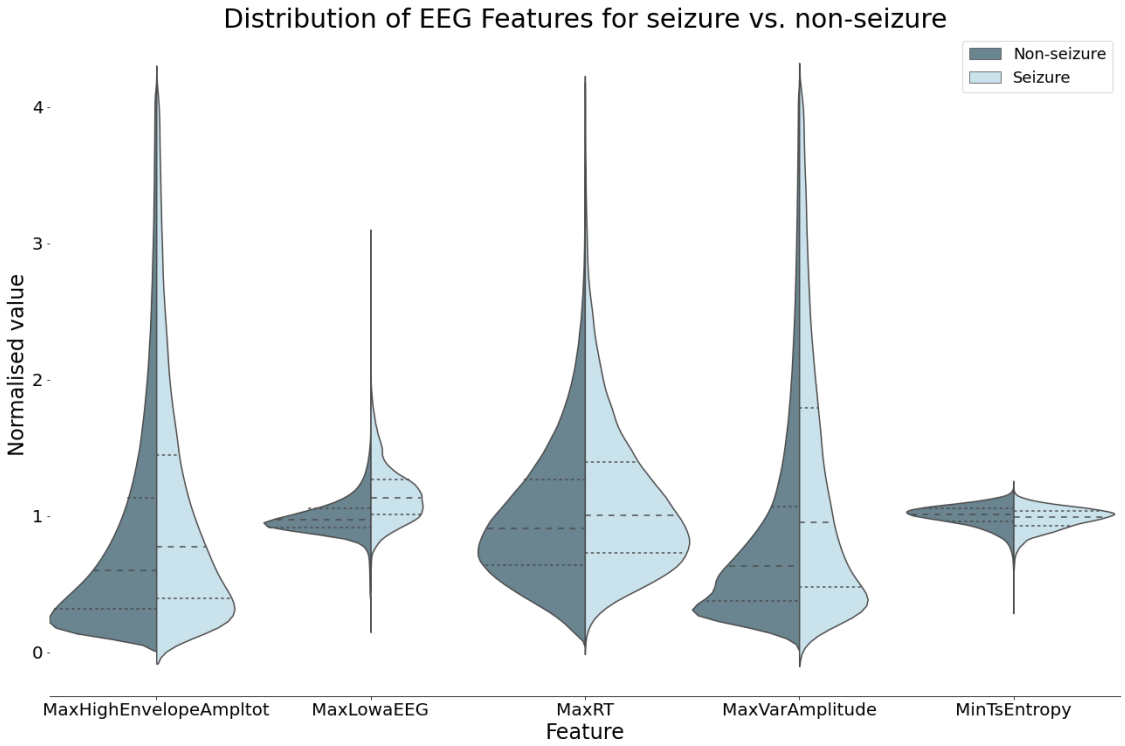


Figure 4.1: Violin plots of the maximum amplitude, maximum lower border of aEEG, maximum theta-to-total ratio, variance in amplitude and minimum entropy features, with upper and lower 1% outliers removed. The distribution of the values of these features is shown for both the samples of the non-seizure (dark blue) and seizure (light blue) groups. The striped lines present the medians and the dotted lines indicate the interquartile ranges.

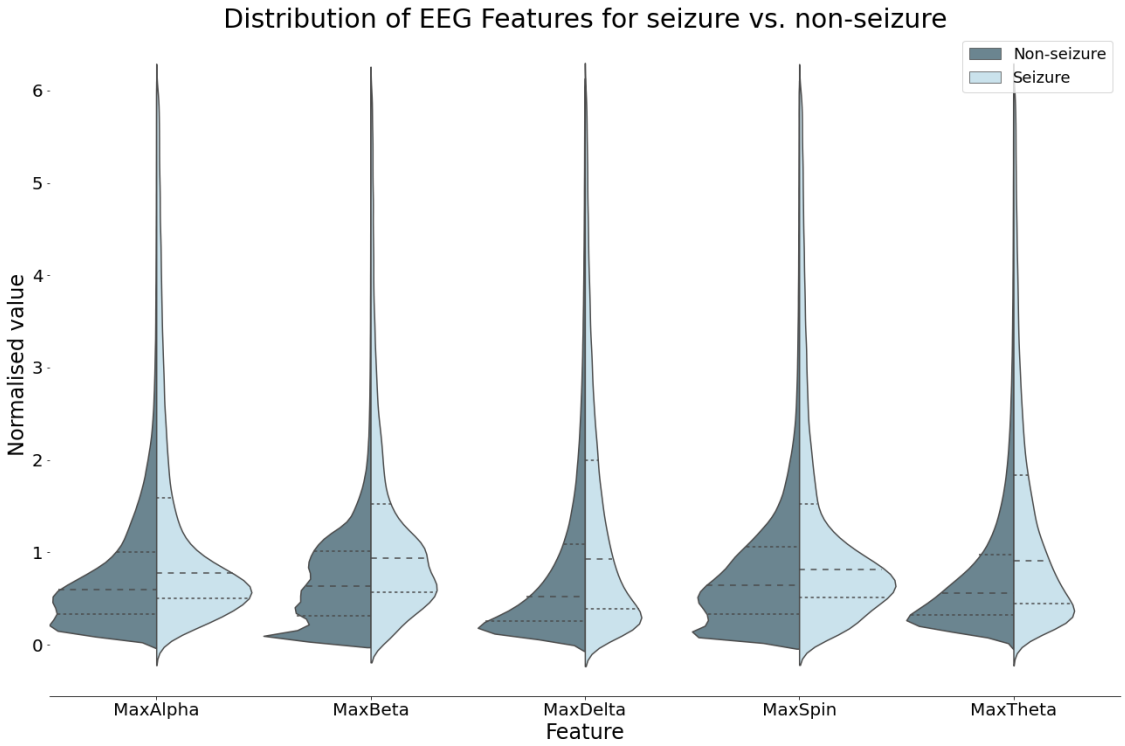


Figure 4.2: Violin plots of the maximum frequency power band features, with upper and lower 1% outliers removed. The distribution of the values of these features is shown for both the samples of the non-seizure (dark blue) and seizure (light blue) groups. The striped lines present the medians and the dotted lines indicate the interquartile ranges.

This is consistent with the results of the Mann-Whitney U-test comparing the medians and interquartile ranges between the seizures and non-seizures in Table F.1. Some variations were very small, but still, all features showed significant differences in the medians and interquartile ranges (Q1-Q3) between seizures and non-seizures for all features. While substantial differences between both groups were found for the amplitude, aEEG and power band features, most others demonstrated only minor differences.

4.4. Feature performance

ROC curves were made representing different window-sum conditions; time windows in which a specific number of positively detected epochs are required to classify the current epoch as a seizure. An example for the best-performing feature, the lower border of the aEEG is shown in Figure 4.3, the same ROCs for the other features can be found in Appendix G. Based on the AUC, the best minimum window-sum condition was chosen for each feature which resulted in a variety of combinations of positive points and windows for the different features. A comparison of the resulting AUCs with and without a window-sum condition demonstrated that using such a condition enhanced the classification performance of each feature.

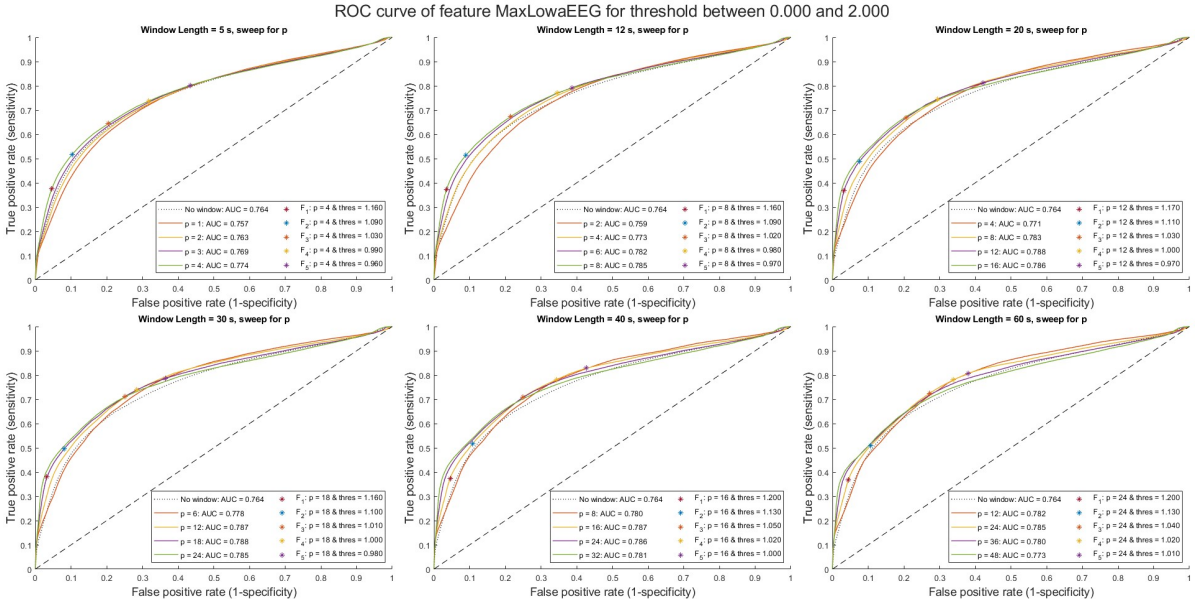


Figure 4.3: ROC curves of the maximum lower border of the aEEG signal for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

The lower border of the aEEG signal displayed the highest AUC of 0.79 (Table 4.2), as shown in Figure 4.4c. Other features with AUCs above 0.60 were the amplitude features, the minimum entropy, all frequency power bands and the theta-total ratio. The ROCs of the best-performing features are visualised in Figure 4.4. For these relevant features, optimal thresholds were determined based on the maximum combined sensitivity and specificity of the optimisation set (presented as dark blue dots in the ROC curves). These can be found in Table 4.2. The sensitivity, specificity, PPV, NPV and accuracy corresponding to the classification based on these optimal thresholds for the 10-EEG optimisation set and the 127-EEG test set are stated as well.

For the optimisation dataset, the lower border of the aEEG and the alpha2 band power showed the highest combined sensitivity and specificity scores. Notably, the aEEG feature maintained its high performance when tested on a larger independent test set, whereas the alpha2 band power and the theta-total power ratio exhibited a decrease in sensitivity. The other amplitude-related features, the entropy measures, coherence, PLV and the delta, theta and alpha1 band maintained their performance while testing for the larger set, while the other features combined with the determined threshold did not.

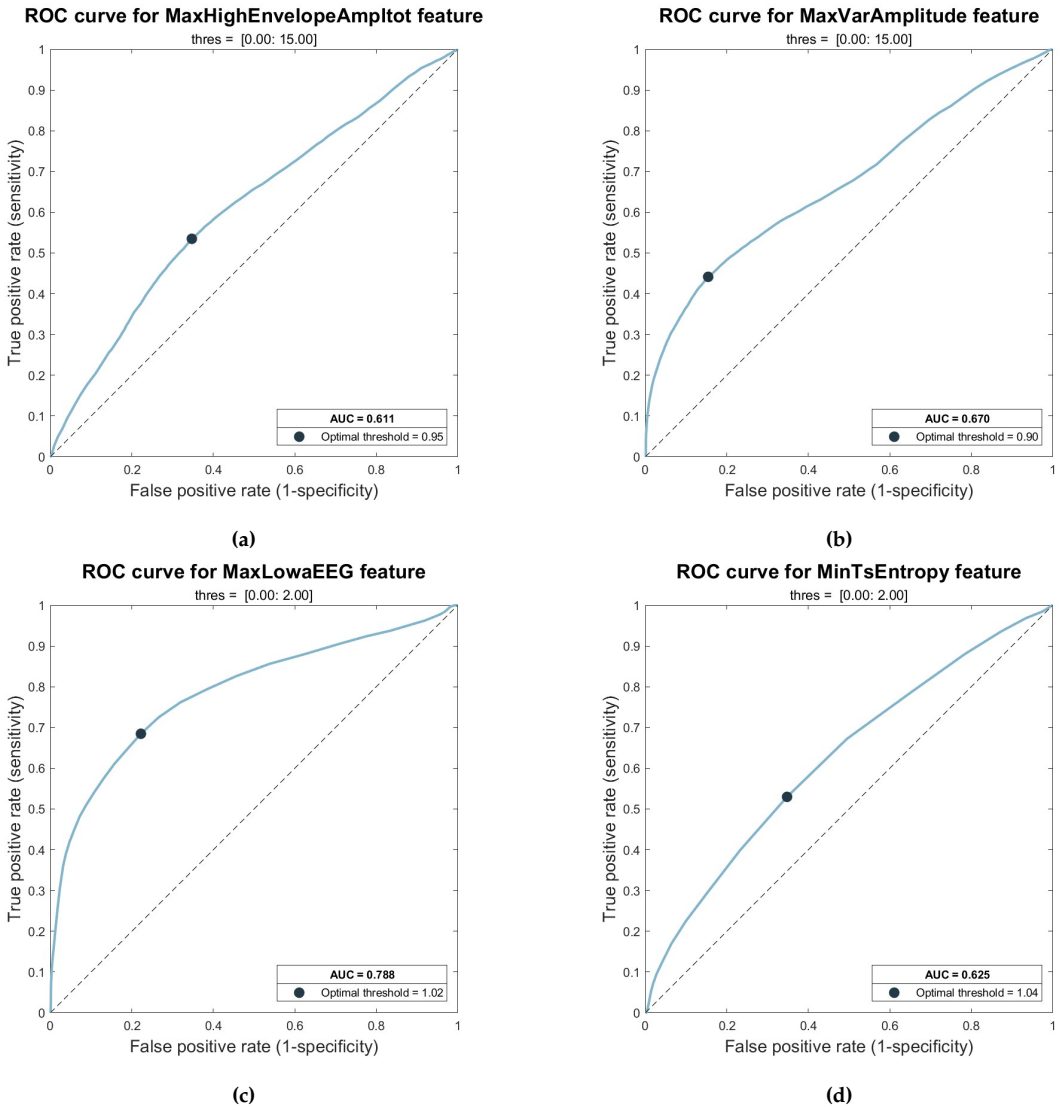


Figure 4.4: ROC's of the most relevant features (AUC>0.60). The optimal threshold (maximum of combined sensitivity and specificity) is specified and depicted in dark blue.

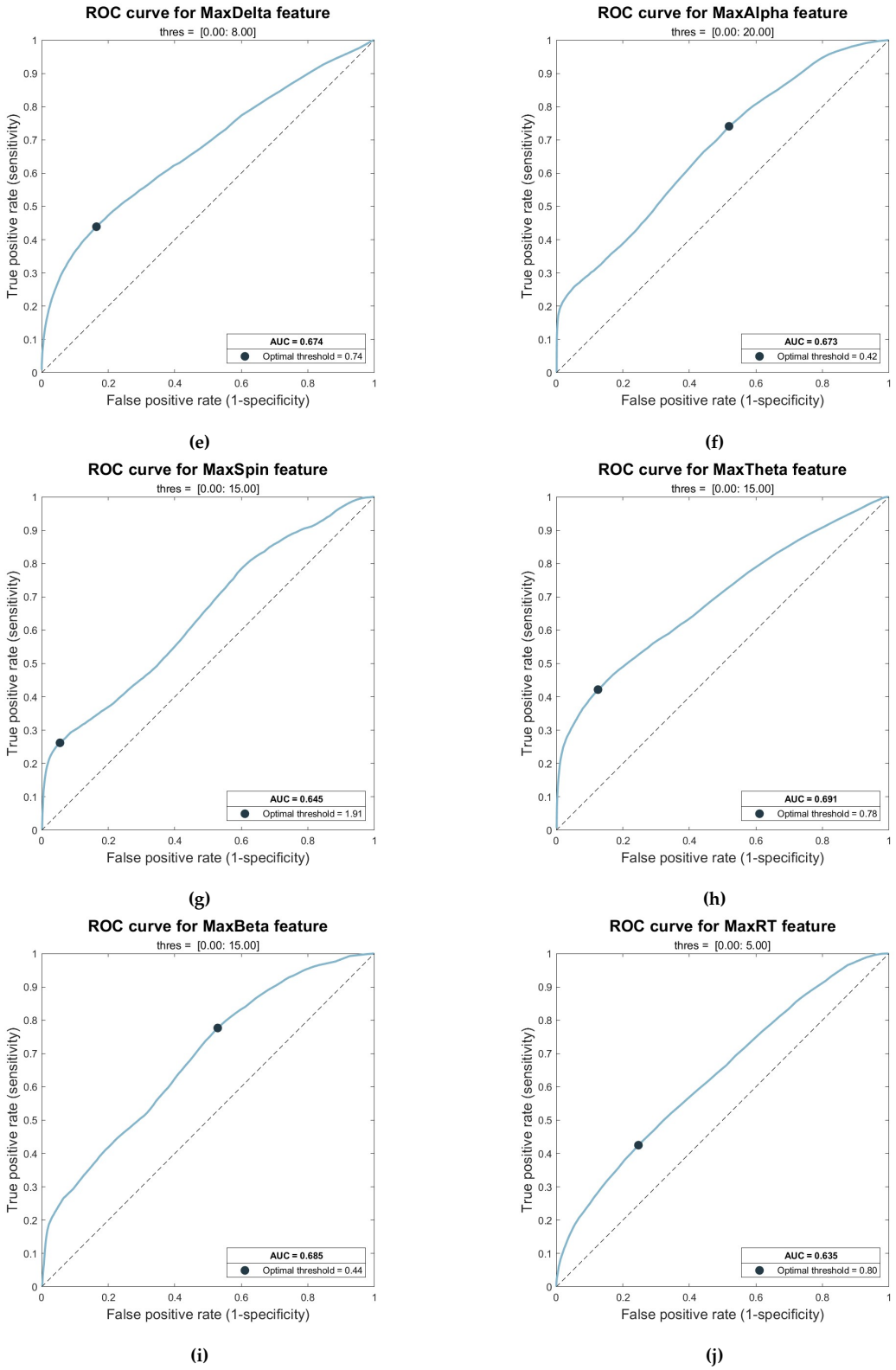


Figure 4.4: Continuation: ROCs of the most relevant features (AUC>0.60). The optimal threshold (maximum of combined sensitivity and specificity) is specified and depicted in yellow.

Table 4.2: Statistical values of the feature extraction methods as classifier for both the optimisation and test set

Extraction method	**AUC**	**Threshold**	**Sensitivity**	**Specificity**	**NPV**	**PPV**	**Accuracy**
			10 (127)	10 (127)	10 (127)	10 (127)	10 (127)
Maximum amplitude	**0.61**	>0.95	0.53 (0.54)	0.65 (0.62)	0.07 (0.03)	0.96 (0.98)	0.65 (0.62)
Variance in amplitude	**0.67**	>0.90	0.44 (0.34)	0.84 (0.85)	0.13 (0.05)	0.97 (0.98)	0.83 (0.84)
Lower border aEEG	**0.79**	>1.03	0.67 (0.62)	0.79 (0.68)	0.14 (0.04)	0.98 (0.99)	0.78 (0.68)
Variance in aEEG	0.54	>0.10	0.94 (0.87)	0.11 (0.16)	0.05 (0.02)	0.97 (0.98)	0.16 (0.18)
Maximum power							
Delta band	**0.67**	>0.72	0.44 (0.38)	0.83 (0.83)	0.12 (0.04)	0.97 (0.98)	0.81 (0.82)
Theta band	**0.69**	>0.72	0.45 (0.36)	0.85 (0.88)	0.13 (0.05)	0.97 (0.98)	0.83 (0.87)
Alpha1 band	**0.67**	>0.42	0.74 (0.53)	0.48 (0.62)	0.13 (0.03)	0.97 (0.98)	0.49 (0.61)
Alpha2 band	**0.65**	>1.99	0.56 (0.10)	0.95 (0.98)	0.22 (0.10)	0.96 (0.98)	0.92 (0.96)
Beta band	**0.69**	>0.44	0.77 (0.79)	0.47 (0.28)	0.07 (0.02)	0.98 (0.98)	0.49 (0.29)
Power ratio to delta							
Theta-delta	0.55	<0.38	0.74 (0.00)	0.3 (0.99)	0.05 (0.00)	0.96 (0.98)	0.35 (0.97)
Alpha1-delta	0.53	>0.19	0.76 (0.40)	0.35 (0.68)	0.06 (0.03)	0.97 (0.98)	0.37 (0.67)
Alpha2-delta	0.52	>1.39	0.73 (0.03)	0.35 (0.98)	0.05 (0.03)	0.96 (0.98)	0.37 (0.96)
Beta-delta	0.54	<1.01	0.64 (0.30)	0.44 (0.78)	0.06 (0.03)	0.96 (0.98)	0.45 (0.77)
Ratio to total power							
Delta-total	0.52	>0.89	0.73 (0.30)	0.32 (0.71)	0.05 (0.02)	0.96 (0.98)	0.34 (0.70)
Theta-total	**0.64**	>0.80	0.44 (0.01)	0.74 (0.97)	0.08 (0.01)	0.96 (0.98)	0.73 (0.96)
Alpha1-total	0.52	<1.27	0.67 (0.31)	0.38 (0.74)	0.05 (0.03)	0.96 (0.98)	0.39 (0.73)
Alpha2-total	0.55	<1.14	0.58 (0.25)	0.52 (0.80)	0.06 (0.03)	0.96 (0.98)	0.52 (0.79)
Beta-total	0.53	>1.01	0.50 (0.08)	0.58 (0.90)	0.06 (0.20)	0.96 (0.98)	0.58 (0.89)
Autocorr. peak-ratio	0.58	<1.01	0.54 (0.12)	0.57 (0.90)	0.06 (0.30)	0.96 (0.98)	0.57 (0.88)
Variance in peak-ratio	0.57	<1.14	0.63 (0.12)	0.49 (0.90)	0.06 (0.03)	0.96 (0.98)	0.49 (0.88)
Tsallis entropy	**0.63**	<1.05	0.52 (0.40)	0.66 (0.66)	0.07 (0.20)	0.96 (0.98)	0.65 (0.65)
Variance in entropy	0.55	<1.01	0.37 (0.35)	0.71 (0.71)	0.06 (0.03)	0.96 (0.98)	0.69 (0.70)
Coherence	0.55	>1.24	0.23 (0.22)	0.87 (0.85)	0.08 (0.03)	0.96 (0.98)	0.84 (0.83)
PLV	0.53	>1.27	0.25 (0.19)	0.86 (0.90)	0.09 (0.03)	0.96 (0.98)	0.83 (0.83)

aEEG = Amplitude-integrated electroencephalography, AUC = Area under the curve, PPV = Positive predictive value, NPV = Negative predictive value

4.5. Correlation between features

The correlation heat map in Figure 4.5 shows the correlation of all the normalised features. The true events are included as well to check the correlation between a feature and the occurrence of a seizure. Especially the maximum power in various frequency bands, the amplitude features and features related to the aEEG signal had a higher positive correlation with the event parameter compared to other features, while entropy showed a slight negative correlation with the true events. This aligns with the performance of these features in the method analysis.
Several features had high correlations between their values, either positive or negative. These include the features related to the frequency analysis, the coherence and PLV, and the amplitude measures. Comparing the correlation of the best-performing features, it appears that the lower border of the aEEG signal was positively correlated with the variance in amplitude, maximum power in frequency bands, and the ratio of theta to total power. However, little correlation existed between the aEEG feature and the minimum entropy or maximum amplitude features.

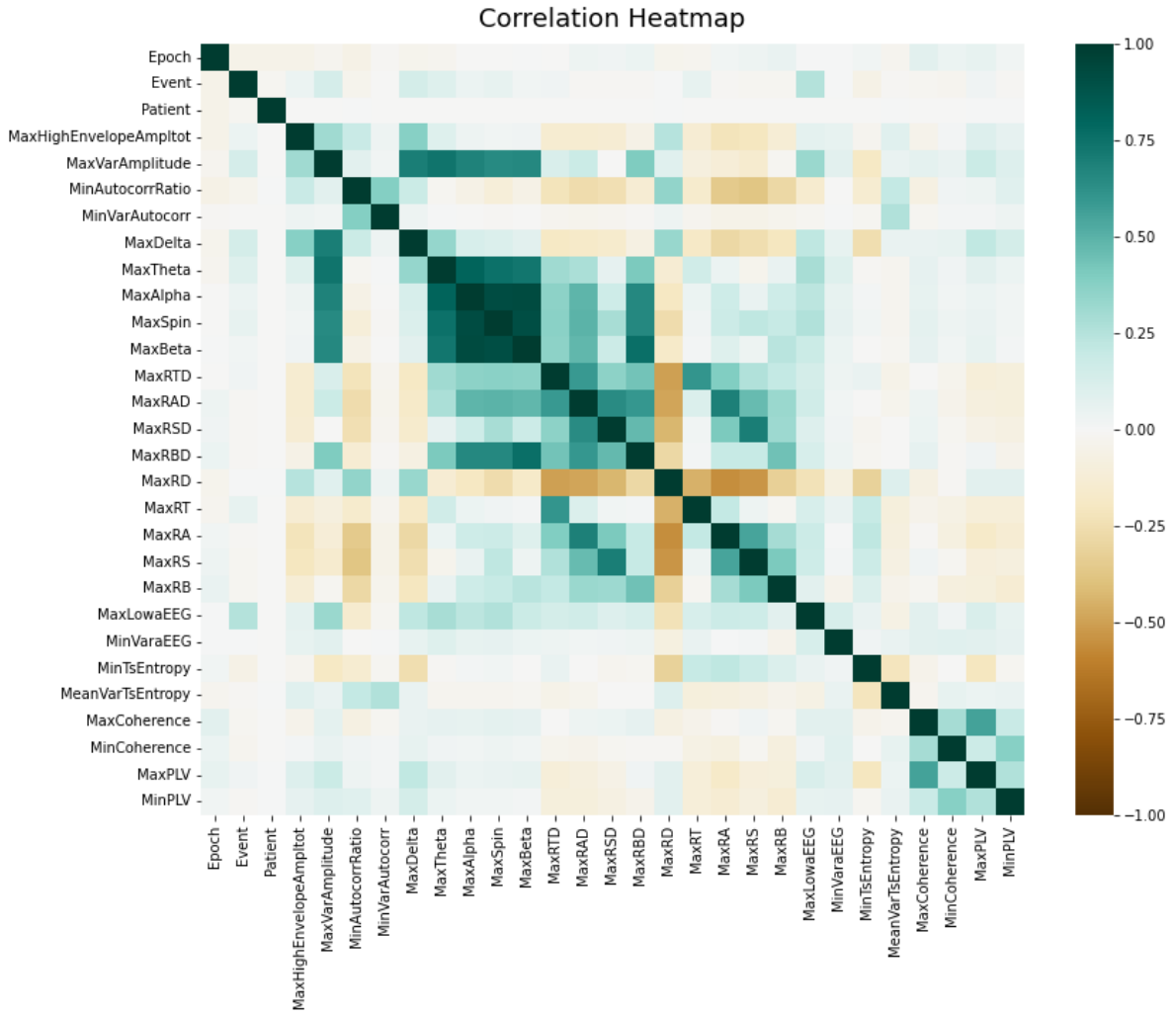
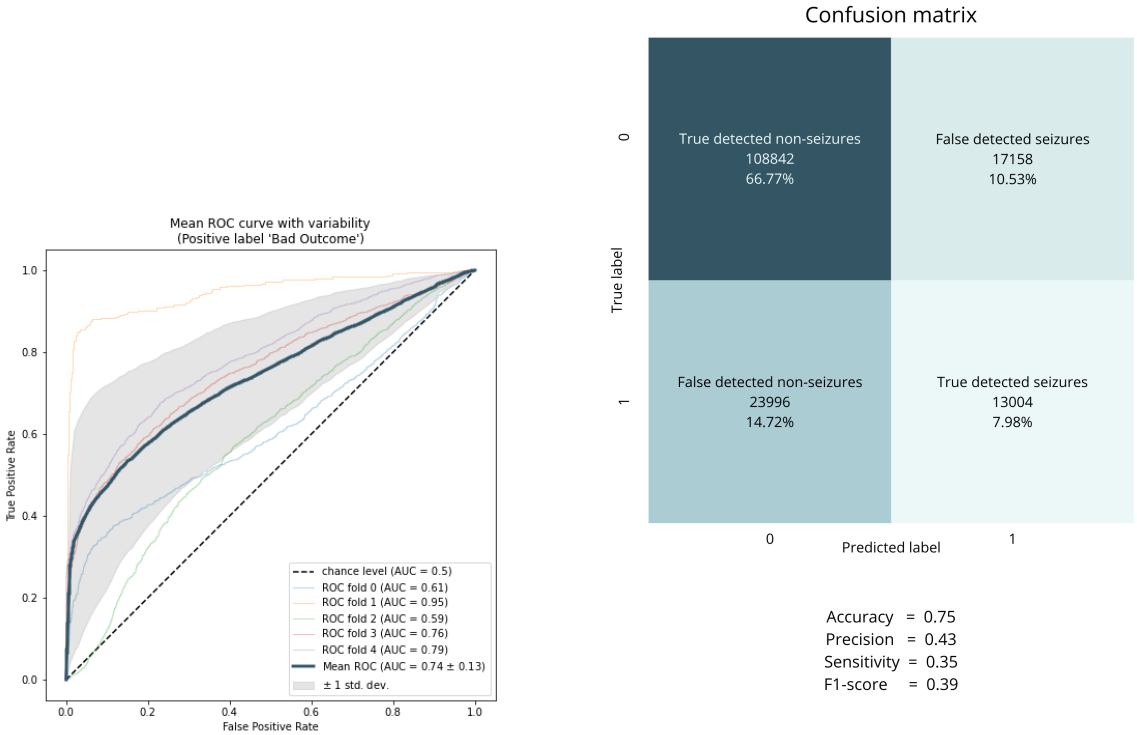


Figure 4.5: Heatmap representing the correlation between the features calculated for the 3-second window. All features were normalised per patient. Shading indicates the level of correlation from a positive correlation in blue, no correlation in white and a negative correlation in brown.

4.6. Random forest classifier

The ROC curve of the classifier trained with a 5-fold hyperparameter tuning for optimisation dataset in Figure 4.6a showed that the classifier overall had a mean AUC of 0.74 ± 0.13. This indicates that the combined RF classifier showed a lower performance for identifying the events in the optimisation dataset compared to the individual lower border aEEG feature with the window-sum condition. The variation in the different folds is substantial, indicating that the performance depends highly on which patients are trained and which are tested. Figures H.3 and H.2 show the most important features determined by the permutation methods based on a decrease in accuracy and impurity. Both show that the lower border of aEEG-feature highly influenced the model, followed by some of the absolute frequency band powers (alpha, theta, alpha2 and beta) and the variance in amplitude. The same features were found to be the best-performing features in the individual method analysis. The trained model was used to classify the data of the 127 patients. This resulted in the confusion matrix presented in 4.6b. The recall or sensitivity of 0.35 indicates that about two-thirds of the true seizures were missed, while the precision of 0.42 indicates that more than half of the classified events were wrongfully classified as such.



(a) **(b)**

Figure 4.6: a) ROC of the two random forest classifiers, trained using the optimisation 10-subject dataset and b) the confusion matrix created when predicting the 127-patient dataset with the trained model.

5 Discussion

In the PICU, neurological monitoring via aEEG represents a crucial diagnostic tool for detecting seizures. However, the identification of seizures using aEEG remains a challenge for healthcare professionals. To address this issue, we conducted an exploratory study aimed at identifying observer-based EEG features that could aid in the accurate classification of seizures. Subsequently, we evaluated the performance of each individual feature as a stand-alone classifier and examined whether combining these features in a random forest classifier could improve the predictive capacity of the model.

Our results demonstrated that several features, including maximum amplitude, amplitude variance, entropy, frequency band powers, and the theta-to-total ratio, achieved an AUC above 0.60, well above chance, with the lower border of the aEEG signal yielding the highest AUC of 0.79. The best performing features maintained their classification ability when tested on a larger, independent dataset. These features were also found to exhibit the strongest correlation with true seizure events, particularly in the delta band, aEEG border, and amplitude variance. The performance of the random forest classifier did not exceed that of the individual aEEG feature with window-sum condition but did find the same features to be the most predictive in classifying seizures.

5.1. Interpretation of results

The results of our exploratory study provide valuable insights into the potential use of additional features for accurate seizure classification in the context of paediatric critical care. It can be inferred that the lower border of the aEEG feature is the most effective in distinguishing between seizure and non-seizure periods. Additionally, other features with employed window-sum conditions have shown potential in aiding the classification of epileptic activity, based on their individual performance as classifiers and correlation with the presence of events. These features can be combined with the aEEG feature to improve the classification between epileptic activity and non-epileptic activity.

In addition to demonstrating the usefulness of different features in distinguishing seizures, we have also shown that the introduction of the window-sum condition improves the classification skills of all investigated features. This can be attributed to the extra dimension that we add by looking at the past. This way, not every outlier is marked as a seizure, which increases the robustness of the feature as a classifier.

Based on our hypothesis regarding the characteristics of epileptic activity in EEG and previous studies, we expected all features to have value in the classification of epilepsy. Given its wide application in clinical settings, we hypothesized that the lower border of the aEEG would be the most informative feature, and our results confirmed this expectation. Although all features showed predictive capabilities higher than chance, many did not perform as well as anticipated.

One of the main reasons that could explain why no single feature, besides aEEG, shows classifying abilities relevant in PICU-setting, is the variability in the manifestation of seizures in EEGs. Research has demonstrated substantial inter- and intra-variability in seizure expression, making it challenging to

differentiate all seizures based on a single characteristic. The inclusion of paediatric patients of varying ages in our dataset may have contributed to even greater variability in the manifestation of epileptic activity in EEGs, due to changing background patterns during development. Even after normalising the data by the mean per patient, differences in the behaviour of EEG patterns may still vary significantly in this patient population. This could have made it more challenging to classify ictal periods based on generalised features.

All amplitude-related features, except the variance in aEEG, were to some extent able to identify seizures from non-seizures. This can be explained by the fact that a common characteristic of epilepsy is the synchronous firing of many neurons, resulting in high amplitude. We hypothesised that, due to the broad variability in the manifestation of seizures, the amplitude could not be used to capture all seizures perfectly due to its simplicity. However, it is possible that due to this simplicity and straightforward use, many seizures can be identified with it. The presence of high amplitude with significant variation in its values is a generally reliable indicator of epileptic activity.

The maximum frequency band powers were among the best classifying features, while all but one power ratio did not suffice as solo classifiers. The fact that the frequency band powers showed classifying abilities may be attributed partly to the fact that the powers are associated with the amplitude, which we already confirmed to be indicative of the presence of seizures. This also explains the high correlation between these features and the variation in amplitude. Additionally, it is interesting to note that a higher theta-total ratio is predictive of the presence of seizures, suggesting that power in the theta band increases relatively more compared to the other frequency bands. The theta ratio performs well in the optimisation group, while the determined threshold does not work for the larger test set. There is a difference in age between the two groups, with the optimisation set consisting mostly of younger patients, while the test set has a broader range of included ages. It is known that younger patients tend to have lower EEG background activity, which could explain the higher power in the lower frequency range, such as theta, that cannot be reproduced in a group with more, older patients. However, further research is needed to provide more conclusive answers. Furthermore, the variability of frequency power ratios during different phases of a seizure may be a reason why most of the power ratios did not work effectively in seizure detection [47, 52]. To address this issue, it may be beneficial to find a way to incorporate these changes in frequency power ratios during different phases of a seizure, which could potentially improve the performance of seizure detection features.

Autocorrelation is a feature related to the rhythmicity of a signal during seizures. Though used similarly in earlier work [33, 34], this feature did not show highly predictive capabilities to distinguish seizures in an EEG during our study. This could indicate that, though a common characteristic, rhythmicity might not always be present during seizures, or not in a fashion that is easily identified with the currently used method. Deburchgraeve et al. found that single patients without much rhythmicity could have a great influence on the performance of the classifier [33]. Therefore, it is possible that this feature is very patient-specific, but it may still be useful in a decision tree together with other features. Another possible explanation is the difference in method, as other studies have often used the ratio between zero-crossings of the autocorrelation, whereas we used the ratio between the peaks. A more elaborate study that identifies the flaws and challenges of this particular method is necessary to find conclusive answers.

The minimum entropy measure revealed that during seizures, there is a decrease in the level of chaos compared to regular periods, likely caused by rhythmicity during seizures. Although the classification ability was significant, it did not meet our expectations. This could be attributed to various factors, such as only a portion of the seizure periods in our dataset exhibiting rhythmic behaviour, as corroborated by the limited results of the autocorrelation feature. Additionally, Tsallis entropy might not have been the most suitable entropy measure for this purpose, as other forms such as permutation and sample entropy have been used in epilepsy classification before. Therefore, it may be beneficial to explore these measures in future studies.

Although coherence and PLV have been shown to be good indicators of epilepsy in other studies [34, 41–44, 55], our results did not show high performance values for these features. There could be several explanations for this discrepancy. One possible factor is our limited electrode set-up. During the calculation of coherence and PLV, we considered the maximum value between one of the six channels over which these features were computed. This implies that we assumed that at least two electrodes

forming that channel were situated in the area where a seizure occurred, resulting in higher coherence or PLV during seizure periods compared to non-seizure periods. Due to our limited electrode setup, it is possible that the seizure activity was located in the area of just one of the electrodes, which could have decreased the maximum coherence or PLV during seizures. However, other studies have also used limited channels for their analysis, so this cannot solely explain the low performance values obtained in our study, which calls for further exploration of this feature and its behaviour during seizures.

The performance of the individual analysis was better than that of the random forest classifier, even though the aEEG feature was also incorporated into the machine learning model. This might be explained by several things. Firstly, there were slight differences in the data used to establish the performance of the individual analyses and the random forest classifier, due to the 5-fold analysis, leading to variation in the results. Additionally, and more importantly, the incorporation of the minimum window-sum condition in the individual analysis may have contributed to the observed variation. Therefore, it is not feasible to make a direct comparison between this random forest model and the aEEG feature classifier. A possible solution would be to create a model that also takes into account the window-sum condition, however, it was outside the scope of this paper. The robustness of the random forest classifier was found to be suboptimal. This may be attributed to the fact that the ROC results of the various folds were each time, based on a small subset of the data. If these patients exhibit a unique expression of a seizure in their EEG, the results can vary greatly, resulting in either good or poor performance. A larger dataset would be able to reduce the effect of such outliers.

Some features demonstrated robust performance on the test set when using the determined threshold derived from the optimisation set. Nevertheless, for some features, these thresholds failed to exhibit such generalisability. A possible explanation for this discrepancy could be that the subset of data in the optimisation set exhibited a particular behaviour for that feature, which did not correspond to the behaviour of the feature in the test set. To ensure the development of more reliable thresholds, future research should consider employing a larger, more diverse optimization. This would increase the generalisability of the model and improve its performance on unseen data.

5.2. Limitations

One point we need to address is the fact that we are using only four electrodes. Several studies have shown that physicians find aEEG interpretation more difficult than EEG when it comes to epilepsy, partly because it is time-compressed, but also because there is much less information to obtain from four electrodes compared to a full setup [67]. The limited information could also be a restricting factor in technological classifiers. Additionally, it is possible that the localisation of the electrodes relative to the seizures is sub-optimal, which may result in the signal not clearly showing the occurrence of a seizure. This could also create problems for the features used in this study. It is possible that valuable information is lost due to the use of only four electrodes, which may result in the features working less effectively.

The bias in the included datasets, created through the data selection process can have influenced the results. The 127 EEGs included in the study were selected based on patients who first had an aEEG recorded, followed by a regular long-term EEG. These EEGs were likely recorded for patients in whom the PICU staff was uncertain about the presence of seizures, which prompted the EEG recording to seek a second opinion from neurophysiologists. As a result, the included EEGs can be stapled as containing hard-to-identify seizures or containing seizure-like events, making it harder to distinguish seizures from non-seizures. However, the study, therefore, does reflect the actual performance of our features on target (a)EEGs with hard-to-identify seizures based on aEEG inspection alone, representing the real-world scenario, making our findings relevant to clinical practice.

Another limitation was that our electrode montage was referenced to the average, which may have altered the overall synchronisation in the signal, making the difference in PLV or coherence between seizure and non-seizure periods less distinct. Other studies used different methods to determine coherence and PLV features, such as taking the mean coherence between all channels [34] or only examining the synchronisation between the left and right hemispheres [42]. Therefore, more research is necessary to determine the optimal usage of synchronisation methods for epilepsy classification.

The fact that we made a classification for each second of EEG will have influenced our results. Not every second during a seizure displays the same characteristics, causing missed events. Additionally, the beginning and end of seizures may not be sharply delineated, which can result in further misclassifications. However, our study was not aimed at accurately classifying every second of the EEG, but rather at identifying trends that can be visualized and provide more information to aid in the diagnosis of seizures.

The gold standard for seizure detection is based on the interpretation of neurophysiology staff, which can be prone to biases, variability and mistakes. Additionally, determining the precise boundaries of the beginning and end of a seizure can be challenging. Therefore, it must be taken into consideration that we compare our results to an imperfect gold standard. Since the interpretation of aEEG, even by neurophysiologists, is more difficult than EEG, we have determined the gold standard based on EEG. However, we acknowledge that errors can still occur in this process. Despite this possible limitation, we deemed this approach to be the best achievable standard for comparison in our study.

Another aspect to consider, regarding the thresholds set for each feature, is the method used to determine them. We chose the best threshold based on the maximum value of sensitivity plus specificity. However, it can be debated whether this is the best approach. The method of determining the threshold will depend on how the features will be used, either in combination or as a standalone analysis. In this study, sensitivity was more important than specificity, as we preferred to minimise the number of missed seizures, even if this resulted in a slightly higher false positive rate. F-beta scores were also explored, but they did not provide consistent thresholds for all features. Therefore, sensitivity plus specificity was deemed the best and most robust measure for all the features for now.

The method used to remove artefacts from EEG signals may have affected how seizures are detected in the recordings. We used artefact subspace reconstruction (ASR), which identifies non-brain activity based on signal variance and removes it while preserving the underlying EEG signal, using a piece of artefact-free EEG as a reference. One of the steps it takes is burst suppression, which removes high amplitudes that resemble artefacts. However, high amplitudes are also present in epileptic activity and might therefore be (partially) removed by the ASR method. This method has previously been used in an EEG study on epilepsy detection [68], and since the amplitude features perform well in this study, it appears the effect is fairly limited. However, it is important to conduct further research on the impact of ASR on epilepsy detection.

Due to the minimum window-sum condition used in the analysis, shorter seizures may have been missed. Each feature was optimised with time windows ranging from 5 to 60 seconds and with a maximum of 48 positive points. However, the higher number of required positive points can make it challenging to detect seizures of the same magnitude as the time window. To address this issue, window lengths of up to 60 seconds and minimum sums of 48 positive points or less were used in the study, trying to limit the amount of missed seizures. For the results of the performance evaluation, the influence of missing these short seizures will be limited, but it is important to keep in mind this limitation during future implementation. Another point to consider is that the window-sum condition introduces a delay in the detection of the signal. This is because a certain number of epochs need to be above a threshold to detect epilepsy, which only occurs after the ictal period has started. The maximum delay is less than a minute, so the effect on diagnosis and treatment is minimal, but it should still be taken into account.

In this study, we normalised the data by dividing the calculated feature value per second by the mean of the entire feature of that patient. This approach allowed us to set thresholds based on the patient's individual means of the feature, eliminating the need for a general threshold. However, it is worth noting that other normalisation methods, such as subtracting the mean of the value and dividing by the standard deviation, are more commonly used. It may be worthwhile for future studies to use this type of normalisation to better understand the effects of our chosen normalisation method.

One potential confounding factor in our study is the use of a band-pass filter with a lower cut-off frequency of 1.5 Hz. While this is a common practice in studies involving adult patients, it may have inadvertently filtered out more relevant information in our paediatric EEG data.

5.3. Future research recommendations

Several things must be considered before clinical implementation. One concern is the establishment of the mean of a certain time window per patient to use as a basis for the threshold. The behaviour of the feature within this time window can have implications on the overall classification performance. For instance, if the window contains seizures, the threshold will be based on the false assumption that the mean is from a regular period. One way to overcome this issue is to allow the staff to select a specific time window that is certain not to contain seizures, though this might increase workload. Another factor that might influence the threshold is the length of the time window. Careful consideration and further research is needed for related issues that might arise during implementation.

Further research could be conducted to investigate which features are capable of identifying specific phases of an ictal period. It is possible that certain features only exhibit changes after a seizure has already commenced, while others may show changes during, or even before the seizure suggesting predictive qualities. Identifying which features are capable of predicting and classifying different stages of seizures could be a next step in creating a more intricate and valuable neuromonitoring dashboard.

Another area of interest may be to explore the generalisability of the feature per age group. Since EEG patterns can vary significantly across different age groups, it might be valuable to investigate which feature combinations, with what thresholds, are most effective for each specific age group. The selection of feature combinations based on age group could potentially improve the generalisability of our findings and lead to more accurate seizure detection in a wider range of patients.

It would be interesting to find out the performance of the features or a combined model in identifying seizures of different duration. It might be possible that seizures were missed due to the employed minimum window-sum condition and to the fact that shorter seizures might not exhibit the full range of epileptic behaviour as larger-sized seizures would. As the results of the test set showed, a great range of seizure lengths is included in the study. Determining the performances of the features on datasets with seizures of different duration could provide insight into which features can identify which length of seizure.

Regarding the limitation of having four electrodes, it might be useful to evaluate the performances of the features on standard EEG recordings. It might reveal that when using more electrodes, the features show better performances, indicating that the information provided by four electrodes is not sufficient. A future study could focus on evaluating the performances of all the features using various amounts of electrodes, determining the optimal setup, with as few electrodes as possible.

The primary objective is to enhance neuromonitoring in the PICU by incorporating supplementary EEG features that aid in identifying seizures. One possible strategy is to visualise the trends of other features on the neuromonitor, in addition to the aEEG signal. A suggestion might be the minimal entropy, based on the lack of correlation it showed with the aEEG feature. Alternatively, one could use the aEEG feature as the foundation of a prediction model, combined with a feature with high sensitivity confirming seizure occurrence, and another feature with high specificity providing reassurance of the absence of seizure activity. This model could provide a probability score that the PICU staff could take into consideration during diagnosis. Future research is necessary to determine how the results of diagnosis based on visual inspection of the aEEG compare to the diagnosis using the enhanced neuromonitoring tool. If the effectiveness of the new monitoring tool can be confirmed, it has the potential to decrease the uncertainty surrounding certain patients, particularly those without clear seizure patterns on the aEEG.

5.4. Conclusion

Our results show that various observer-based features can aid in the identification of seizures in 4-electrode EEG. These additional features could increase the accuracy and decrease the delay and uncertainty in diagnosing epilepsy in aEEG by PICU staff, reducing the need for full setup EEG monitoring and improving patient outcomes in paediatric critical care.

Bibliography

[1] World Health Organization, International League Against Epilepsy, and International Bureau for Epilepsy. *Epilepsy: a public health imperative*. Tech. rep. Geneva: World Health Organization, 2019.

[2] P. Camfield and C. Camfield. "Incidence, prevalence and aetiology of seizures and epilepsy in children". In: *Epileptic Disorders* 17.2 (June 2015), pp. 117–123. DOI: 10.1684/EPD.2015.0736.

[3] N. S. Abend et al. "Electroencephalographic Monitoring in the Pediatric Intensive Care Unit". In: *Current neurology and neuroscience reports* 13.3 (Feb. 2013), p. 330. DOI: 10.1007/S11910-012-0330-3.

[4] N. Jette et al. "Frequency and predictors of nonconvulsive seizures during continuous electroencephalographic monitoring in critically ill children". In: *Archives of neurology* 63.12 (Dec. 2006), pp. 1750–1755. DOI: 10.1001/ARCHNEUR.63.12.1750.

[5] A. Shahwan et al. "The prevalence of seizures in comatose children in the pediatric intensive care unit: a prospective video-EEG study". In: *Epilepsia* 51.7 (July 2010), pp. 1198–1204. DOI: 10.1111/J.1528-1167.2009.02517.X.

[6] K. Williams, R. Jarrar, and J. Buchhalter. "Continuous video-EEG monitoring in pediatric intensive care units". In: *Epilepsia* 52.6 (June 2011), pp. 1130–1136. DOI: 10.1111/J.1528-1167.2011.03070.X.

[7] D. M. Ferriero et al. "Management of Stroke in Neonates and Children: A Scientific Statement From the American Heart Association/American Stroke Association". In: *Stroke* 50.3 (Mar. 2019), E51–E96. DOI: 10.1161/STR.0000000000000183.

[8] K. G. Jordan. "Continuous EEG monitoring in the neuroscience intensive care unit and emergency department". In: *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society* 16.1 (1999), pp. 14–39. DOI: 10.1097/00004691-199901000-00002.

[9] M. L. Scheuer. "Continuous EEG monitoring in the intensive care unit". In: *Epilepsia* 43 Suppl 3.SUPPL. 3 (2002), pp. 114–127. DOI: 10.1046/J.1528-1157.43.S.3.7.X.

[10] D. Friedman, J. Claassen, and L. J. Hirsch. "Continuous electroencephalogram monitoring in the intensive care unit". In: *Anesthesia and analgesia* 109.2 (2009), pp. 506–523. DOI: 10.1213/ANE.0B013E3181A9D8B5.

[11] I. Sánchez Fernández et al. "Electrographic seizures after convulsive status epilepticus in children and young adults: a retrospective multicenter study". In: *The Journal of pediatrics* 164.2 (2014). DOI: 10.1016/J.JPEDS.2013.09.032.

[12] I. Sánchez Fernández et al. "Time to electroencephalography is independently associated with outcome in critically ill neonates and children". In: *Epilepsia* 58.3 (Mar. 2017), pp. 420–428. DOI: 10.1111/EPI.13653.

[13] N. S. Abend et al. "Electroencephalographic monitoring during hypothermia after pediatric cardiac arrest". In: *Neurology* 72.22 (June 2009), pp. 1931–1940. DOI: 10.1212/WNL.0B013E3181A82687.

[14] M. Hunfeld et al. "A Systematic Review of Neuromonitoring Modalities in Children Beyond Neonatal Period After Cardiac Arrest". In: *Pediatric critical care medicine : a journal of the Society of Critical Care Medicine and the World Federation of Pediatric Intensive and Critical Care Societies* 21.10 (2020), E927–E933. DOI: 10.1097/PCC.0000000000002415.

[15] S. Bandyopadhyay, M. Z. Koubeissi, and N. J. Azar. "Physiologic Basis of EEG and Epilepsy". In: *Epilepsy Board Review*. Springer, New York, NY, 2017, pp. 3–13. DOI: 10.1007/978-1-4939-6774-2{_}1.

[16] L. J. Hirsch and R. P. Brenner. "Atlas of EEG in Critical Care". In: *Atlas of EEG in Critical Care* (Jan. 2010). DOI: 10.1002/9780470746707.

[17] S. M. Sanchez et al. "Pediatric ICU EEG monitoring: current resources and practice in the United States and Canada". In: *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society* 30.2 (Apr. 2013), pp. 156–160. DOI: 10.1097/WNP.0B013E31827EDA27.

[18] R. A. Shellhaas et al. "The American clinical neurophysiology society's guideline on continuous electroencephalography monitoring in neonates". In: *Journal of Clinical Neurophysiology* 28.6 (Dec. 2011), pp. 611–617. DOI: 10.1097/WNP.0B013E31823E96D7.

[19] D. Maynard, P. F. Prior, and D. F. Scott. "Device for continuous monitoring of cerebral activity in resuscitated patients." In: *British Medical Journal* 4.5682 (Nov. 1969), p. 545. DOI: 10.1136/BMJ.4.5682.545-A.

[20] L. Hellström-Westas. "Amplitude-integrated electroencephalography for seizure detection in newborn infants". In: *Seminars in fetal & neonatal medicine* 23.3 (June 2018), pp. 175–182. DOI: 10.1016/J.SINY.2018.02.003.

[21] W. Liu et al. "Prognostic Value of Clinical Tests in Neonates With Hypoxic-Ischemic Encephalopathy Treated With Therapeutic Hypothermia: A Systematic Review and Meta-Analysis". In: *Frontiers in Neurology* 0 (Feb. 2020), p. 133. DOI: 10.3389/FNEUR.2020.00133.

[22] N. Bruns et al. "aEEG Use in Pediatric Critical Care-An Online Survey". In: *Frontiers in pediatrics* 8 (Jan. 2020). DOI: 10.3389/FPED.2020.00003.

[23] N. Bruns, U. Felderhoff-Müser, and C. Dohna-Schwake. "aEEG as a useful tool for neuromonitoring in critically ill children - Current evidence and knowledge gaps". In: *Acta paediatrica (Oslo, Norway : 1992)* 110.4 (Apr. 2021), pp. 1132–1140. DOI: 10.1111/APA.15676.

[24] A. T. Tzallas et al. "Automated Epileptic Seizure Detection Methods: A Review Study". In: *Epilepsy - Histological, Electroencephalographic and Psychological Aspects* (2012). DOI: 10.5772/1194.

[25] U. R. Acharya et al. "Automated EEG analysis of epilepsy: A review". In: *Knowledge-Based Systems* 45 (June 2013), pp. 147–165. DOI: 10.1016/J.KNOSYS.2013.02.014.

[26] L. Orosco et al. "Review: A Survey of Performance and Techniques for Automatic Epilepsy Detection". In: *Journal of Medical and Biological Engineering* 33.6 (2013), pp. 526–537. DOI: 10.5405/jmbe.1463.

[27] A. Sharmila. "Epilepsy detection from EEG signals: a review". In: *Journal of Medical Engineering & Technology* 42.5 (Nov. 2018), pp. 368–380. DOI: 10.1080/03091902.2018.1513576.

[28] T. Kim et al. "Epileptic Seizure Detection and Experimental Treatment: A Review". In: *Frontiers in Neurology* 11 (July 2020), p. 701. DOI: 10.3389/FNEUR.2020.00701/XML/NLM.

[29] A. A. de Hond et al. "Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review". In: *npj Digital Medicine 2022 5:1* 5.1 (Jan. 2022), pp. 1–13. DOI: 10.1038/s41746-021-00549-7.

[30] E. B. Bromfield, J. E. Cavazos, and J. I. Sirven. *An introduction to epilepsy*. West Hartford (CCT): American Epilepsy Society, 2006.

[31] A. Aarabi, F. Wallois, and R. Grebe. "Automated neonatal seizure detection: A multistage classification system through feature selection based on relevance and redundancy analysis". In: *Clinical Neurophysiology* 117.2 (Feb. 2006), pp. 328–340. DOI: 10.1016/J.CLINPH.2005.10.006.

[32] B. R. Greene et al. "A comparison of quantitative EEG features for neonatal seizure detection". In: *Clinical Neurophysiology* 119.6 (June 2008), pp. 1248–1261. DOI: 10.1016/J.CLINPH.2008.02.001.

[33] W. Deburchgraeve et al. "Automated neonatal seizure detection mimicking a human observer reading EEG". In: *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology* 119.11 (Nov. 2008), pp. 2447–2454. DOI: 10.1016/J.CLINPH.2008.07.281.

[34] M. C. Cloostermans, C. C. De Vos, and M. J. Van Putten. "A novel approach for computer assisted EEG monitoring in the adult ICU". In: *Clinical Neurophysiology* 122.10 (Oct. 2011), pp. 2100–2109. DOI: 10.1016/J.CLINPH.2011.02.035.

[35] J. Li et al. "Using Permutation Entropy to Measure the Changes in EEG Signals During Absence Seizures". In: *Entropy* 16.6 (2014), pp. 3049–3061. DOI: 10.3390/E16063049.

[36] E. Ferlazzo et al. "Permutation entropy of scalp EEG: A tool to investigate epilepsies: Suggestions from absence epilepsies". In: *Clinical Neurophysiology* 125.1 (Jan. 2014), pp. 13–20. DOI: 10.1016/J.CLINPH.2013.06.023.

[37] T. Yang, W. Chen, and G. Cao. "Automated classification of neonatal amplitude-integrated EEG based on gradient boosting method". In: *Biomedical Signal Processing and Control* 28 (July 2016), pp. 50–57. DOI: 10.1016/J.BSPC.2016.04.004.

[38] T. Kalayci and O. özdamar. "Wavelet Preprocessing for Automated Neural Network Detection of EEG Spikes". In: *IEEE Engineering in Medicine and Biology Magazine* 14.2 (1995), pp. 160–166. DOI: 10.1109/51.376754.

[39] S. G. Mallat. "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11.7 (1989), pp. 674–693. DOI: 10.1109/34.192463.

[40] S. J. Schiff et al. "Fast wavelet transformation of EEG". In: *Electroencephalography and clinical neurophysiology* 91.6 (1994), pp. 442–455. DOI: 10.1016/0013-4694(94)90165-1.

[41] S. Abdullateef et al. "Quantitative detection of seizures with minimal-density EEG montage using phase synchrony and cross-channel coherence amplitude in critical care". In: *Proceedings of the 44th International Conference of the IEEE Engineering in Medicine and Biology Society*. Glasgow: IEEE Xplore, July 2022, pp. 259–262. DOI: 10.1109/EMBC48229.2022.9871595.

[42] D. Ravish, S. Devi, and S. Krishnamoorthy. "Wavelet analysis of EEG for seizure detection: Coherence and phase synchrony estimation". In: *Biomedical Research* 26.3 (Jan. 2015), pp. 514–524.

[43] I. Wijayanto, R. Hartanto, and H. Nugroho. "Quantitative Analysis of Inter- and Intrahemispheric Coherence on Epileptic Electroencephalography Signal". In: *Journal of Medical Signals and Sensors* 12.2 (Apr. 2022), p. 145. DOI: 10.4103/JMSS.JMSS{_}63{_}20.

[44] J. Romaine et al. "EEG—Single-Channel Envelope Synchronisation and Classification for Seizure Detection and Prediction". In: *Brain Sciences* 11.4 (Apr. 2021), p. 516. DOI: 10.3390/brainsci11040516.

[45] C. Stam, G. Nolte, and A. Daffertshofer. "Phase lag index: assessment of functional connectivity from multi channel EEG and MEG with diminished bias from common sources". In: *Human brain mapping* 28.11 (Nov. 2007), pp. 1178–1193. DOI: 10.1002/HBM.20346.

[46] L. Hellström-Westas, L. De Vries, and I. Rosén. *An Atlas of Amplitude-Integrated EEGs in the Newborn*. 2nd ed. Boca Raton: CRC Press, Aug. 2008. DOI: 10.3109/9781439813898.

[47] L. V. Marcuse, M. C. Fields, and J. J. Yoo. "Rowan's Primer of EEG: Second edition". In: *Rowan's Primer of EEG: Second Edition* (Jan. 2015), pp. 1–206. DOI: 10.1097/00004691-900000000-99503.

[48] C. M. Lommen et al. "An algorithm for the automatic detection of seizures in neonatal amplitude-integrated EEG". In: *Acta paediatrica (Oslo, Norway : 1992)* 96.5 (May 2007), pp. 674–680. DOI: 10.1111/J.1651-2227.2007.00223.X.

[49] Y. Wang et al. "Classification of neonatal amplitude-integrated EEG using random forest model with combined feature". In: *Proceedings - 2013 IEEE International Conference on Bioinformatics and Biomedicine, IEEE BIBM 2013* (2013), pp. 285–290. DOI: 10.1109/BIBM.2013.6732504.

[50] C. Satirasethawong, A. Lek-Uthai, and K. Chomtho. "Amplitude-integrated EEG processing and its performance for automatic seizure detection". In: *IEEE 2015 International Conference on Signal and Image Processing Applications, ICSIPA*. Institute of Electrical and Electronics Engineers Inc., Feb. 2015, pp. 551–556. DOI: 10.1109/ICSIPA.2015.7412252.

[51] D. Wu et al. "Automatic epileptic seizures joint detection algorithm based on improved multi-domain feature of cEEG and spike feature of aEEG". In: *IEEE Access* 7 (2019), pp. 41551–41564. DOI: 10.1109/ACCESS.2019.2904949.

[52] D. Gajic et al. "Classification of eeg signals for detection of epileptic seizures based on wavelets and statistical pattern recognition". In: *Biomedical Engineering: Applications, Basis and Communications* 26.02 (Apr. 2014), p. 1450021. DOI: 10.4015/S1016237214500215.

[53] Z. Iscan, Z. Dokur, and T. Demiralp. "Classification of electroencephalogram signals with combined time and frequency features". In: *Expert Systems with Applications* 38.8 (Aug. 2011), pp. 10499–10505. DOI: 10.1016/J.ESWA.2011.02.110.

[54] J. Pyrzowski et al. "Interval analysis of interictal EEG: pathology of the alpha rhythm in focal epilepsy". In: *Scientific Reports 2015 5:1* 5.1 (Nov. 2015), pp. 1–10. DOI: 10.1038/srep16230.

[55] M. J. Van Putten. "The colorful brain: visualization of EEG background patterns". In: *Journal of clinical neurophysiology : official publication of the American Electroencephalographic Society* 25.2 (Apr. 2008), pp. 63–68. DOI: 10.1097/WNP.0B013E31816BDF85.

[56] J.-P. Lachaux et al. "Measuring phase synchrony in brain signals". In: *Human Brain Mapping* 8.4 (1999), p. 194. DOI: 10.1002/(sici)1097-0193(1999)8:4<194::aid-hbm4>3.0.co;2-c.

[57] A. M. Bastos and J. M. Schoffelen. "A tutorial review of functional connectivity analysis methods and their interpretational pitfalls". In: *Frontiers in Systems Neuroscience* 9.JAN2016 (Jan. 2016), p. 175. DOI: 10.3389/FNSYS.2015.00175/BIBTEX.

[58] D. Cho et al. "EEG-Based Prediction of Epileptic Seizures Using Phase Synchronization Elicited from Noise-Assisted Multivariate Empirical Mode Decomposition". In: *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25.8 (Aug. 2017), pp. 1309–1318. DOI: 10.1109/TNSRE.2016.2618937.

[59] MATLAB. *Matlab version: 9.14.0 (R2023a)*. Natick, Massachusetts, 2023.

[60] A. Delorme and S. Makeig. "EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis". In: *Journal of Neuroscience Methods* 134.1 (Mar. 2004), pp. 9–21. DOI: 10.1016/j.jneumeth.2003.10.009.

[61] R. Oostenveld et al. "FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data". In: *Computational intelligence and neuroscience* 2011 (2011). DOI: 10.1155/2011/156869.

[62] T. R. Mullen et al. "Real-time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG". In: *IEEE transactions on bio-medical engineering* 62.11 (Nov. 2015), p. 2553. DOI: 10.1109/TBME.2015.2481482.

[63] S. Blum et al. "A riemannian modification of artifact subspace reconstruction for EEG artifact handling". In: *Frontiers in Human Neuroscience* 13 (Feb. 2019), p. 141. DOI: 10.3389/FNHUM.2019.00141/BIBTEX.

[64] Z. A. Vesoulis et al. "WU-NEAT: A clinically validated, open-source MATLAB toolbox for limited-channel neonatal EEG analysis". In: *Computer Methods and Programs in Biomedicine* 196 (Nov. 2020), p. 105716. DOI: 10.1016/J.CMPB.2020.105716.

[65] A. Liu et al. "Detection of neonatal seizures through computerized EEG analysis". In: *Electroencephalography and clinical neurophysiology* 82.1 (1992), pp. 30–37. DOI: 10.1016/0013-4694(92)90179-L.

[66] F. Pedregosa et al. "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12.85 (2011), pp. 2825–2830.

[67] M. Kadivar et al. "A Comparison Of Conventional Electroencephalography With Amplitude-Integrated EEG In Detection Of Neonatal Seizures". In: *Medical devices (Auckland, N.Z.)* 12 (2019), pp. 489–496. DOI: 10.2147/MDER.S214662.

[68] G. S. Drenthen et al. "Predictive value of functional MRI and EEG in epilepsy diagnosis after a first seizure". In: *Epilepsy & Behavior* 115 (Feb. 2021), p. 107651. DOI: 10.1016/J.YEBEH.2020.107651.

A

Patient characteristics

Figure A.1: Distribution of the ages of the patients included in the test set (light) and in the optimisation set (dark).

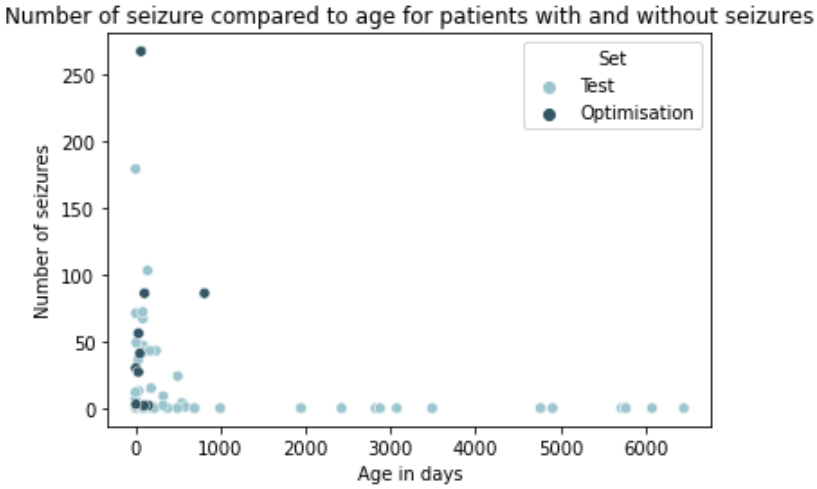


Figure A.2: The distribution of the number of seizures versus the age in days of the patient in the test (light) and optimisation (dark) set.

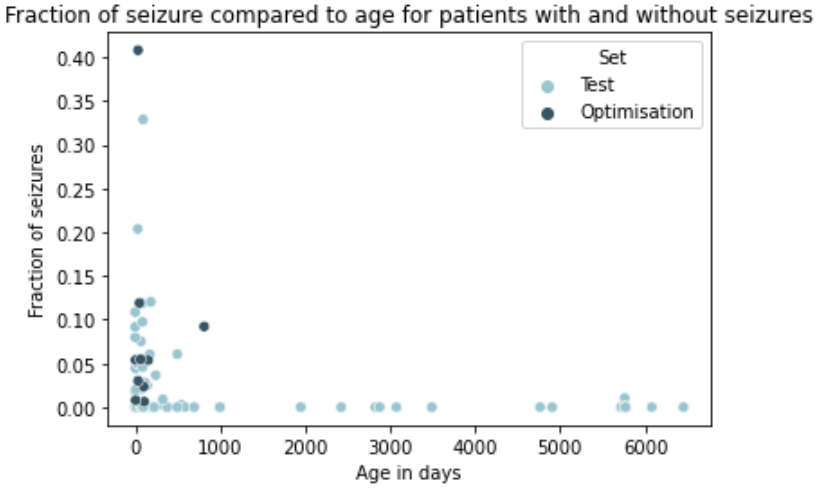


Figure A.3: The distribution of the fraction of seizure in the recording versus the age in days of the patient in the test (light) and optimisation (dark) set.

B

Schematic overview code

The source code is available at github.com/SanneLange/Epilepsy-detection

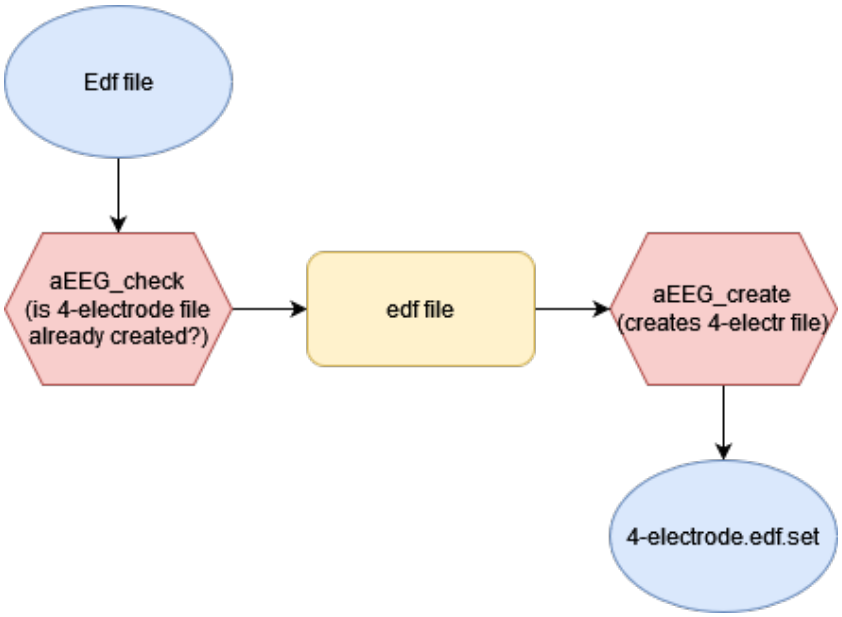


Figure B.1: Schematic overview of loading and transforming the data. The input data (blue) is the EDF (European Data Format) file with the EEG recording. The first function (red) checks whether a file already exists of that EEG with four electrodes, and otherwise, it creates a 4-electrode .set file. All functions are presented in red, while the variables are shaded in yellow and the input and output files are in blue.

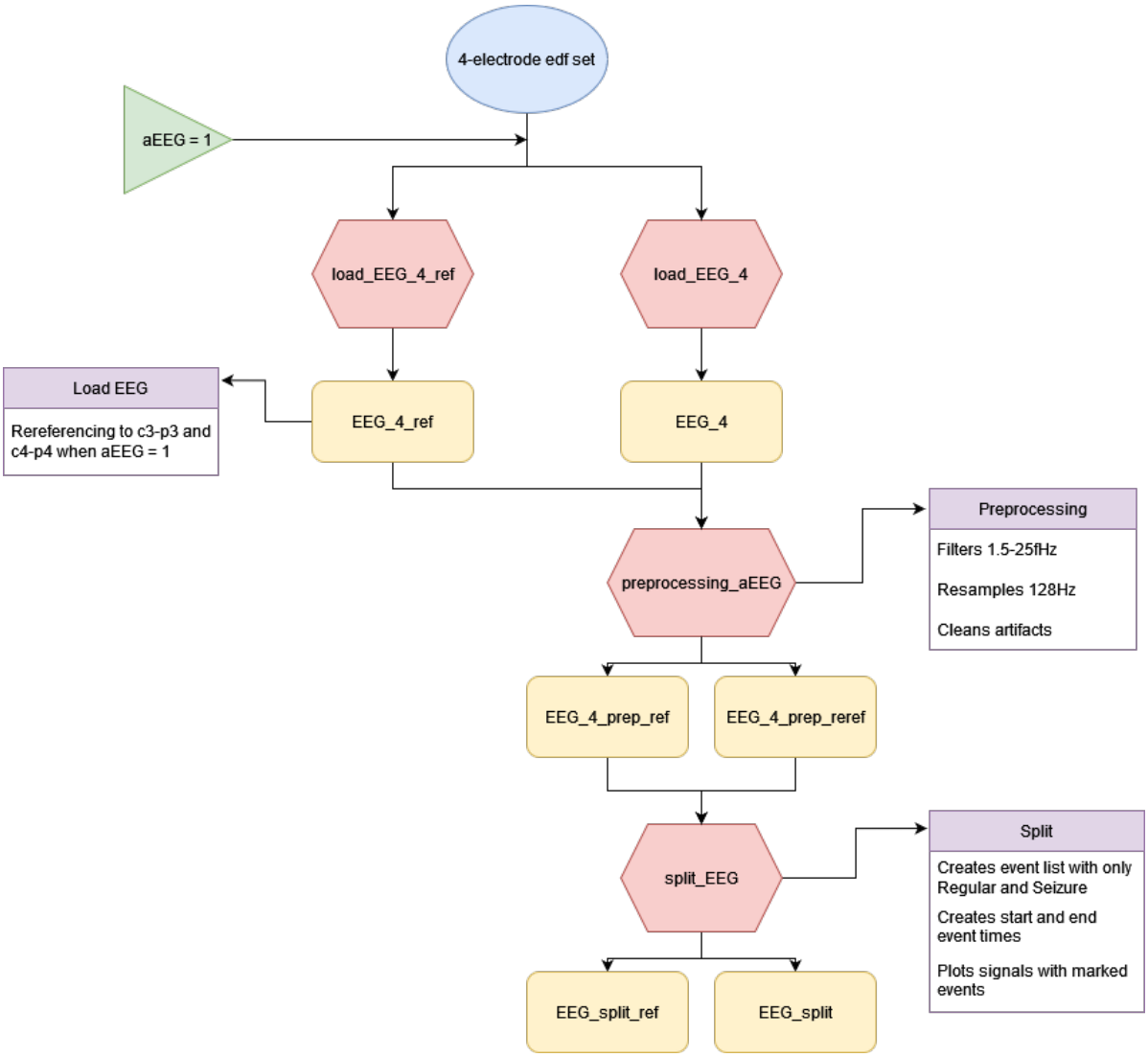


Figure B.2: Schematic overview of preprocessing steps of the data. The input data (blue) is loaded, and for the feature method aEEG, the signals are re-referenced. The data is preprocessed and the seizures are identified. All functions are presented in red, while the variables are shaded in yellow and the input and output files are in blue. Purple boxes provide more information about the functions.

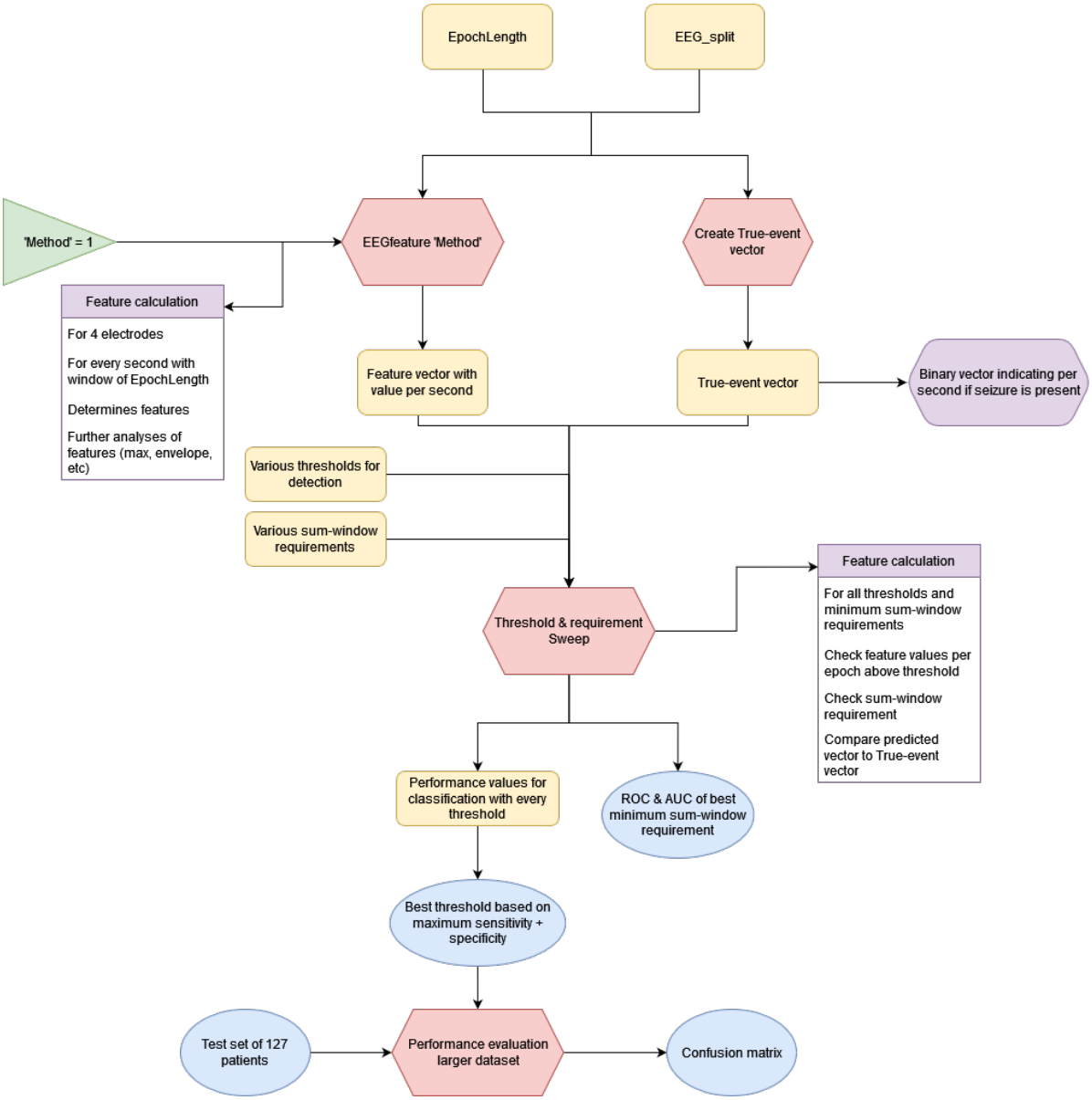
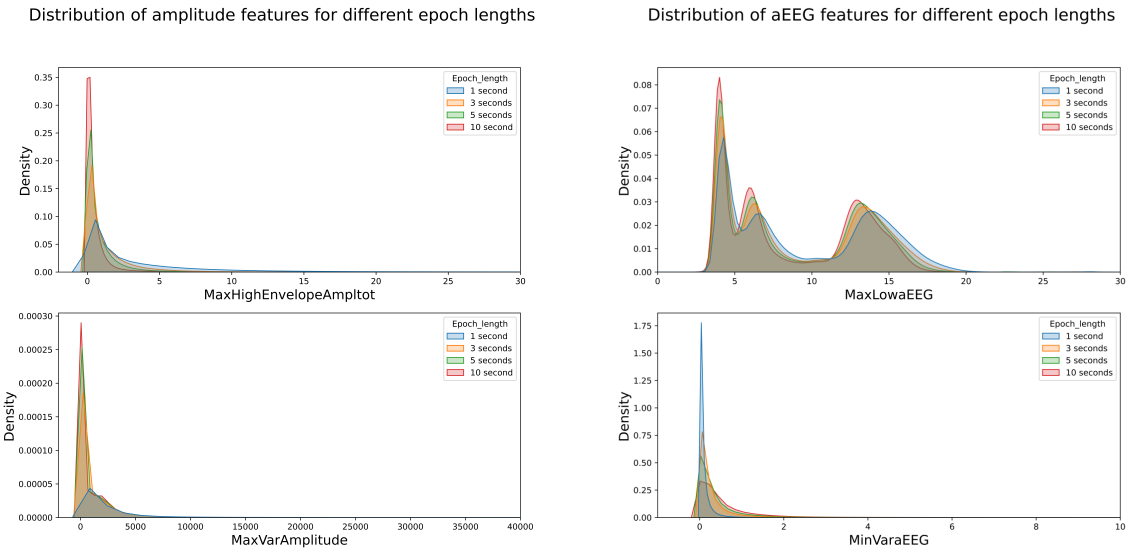


Figure B.3: Schematic overview of the steps taken in the individual method analysis. The EEG signal is used to create the feature values and the vector indicating true events. The threshold & window-sum requirement sweep finds the best window-sum combination and threshold. This threshold is then evaluated with the test set. All functions are presented in red, while the variables are shaded in yellow and the input and output files are in blue. Purple boxes provide more information about the functions.

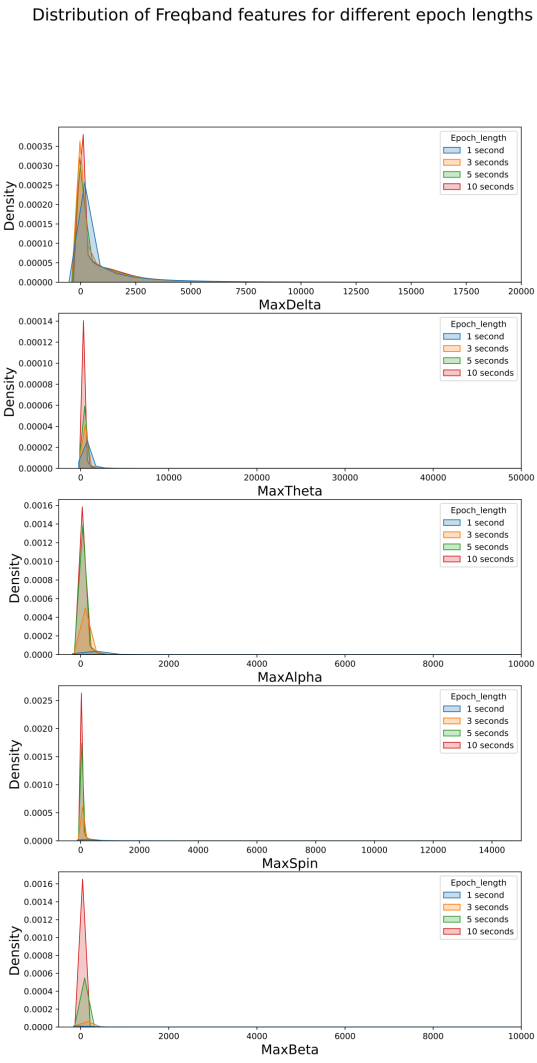
C

Distribution plots of different epoch lengths

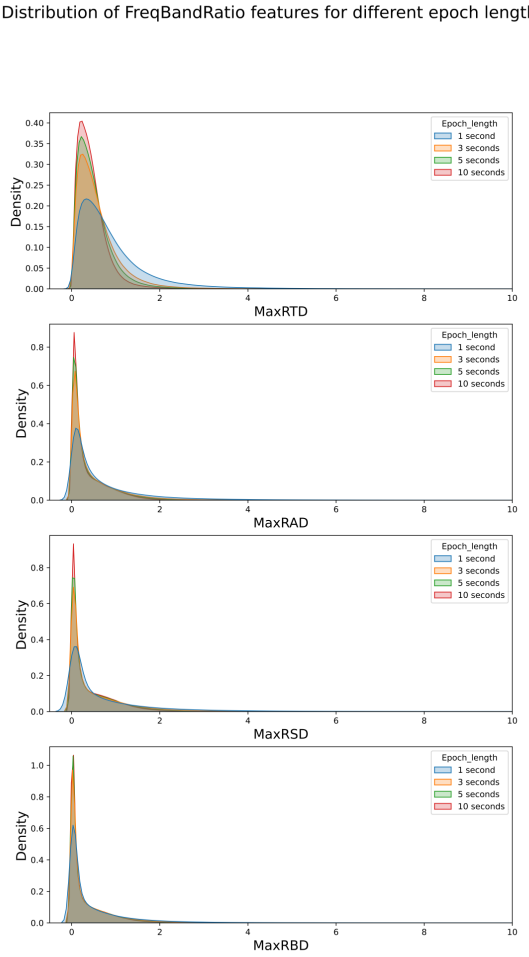


(a) **(b)**

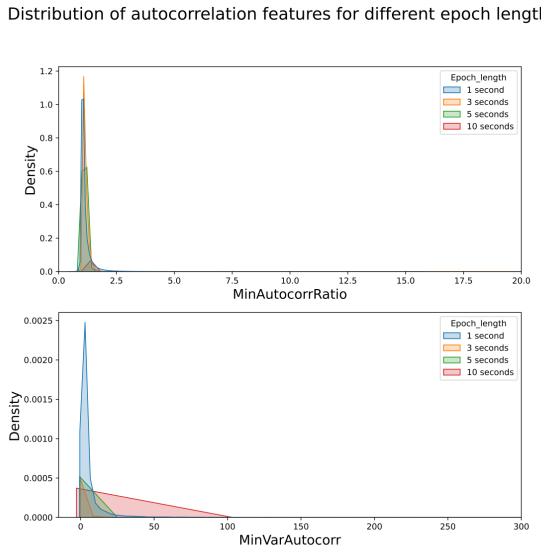
Figure C.1: Distribution of feature values calculated for window-length of 1 (blue), 3 (orange), 5 (green), and 10 (red) seconds.



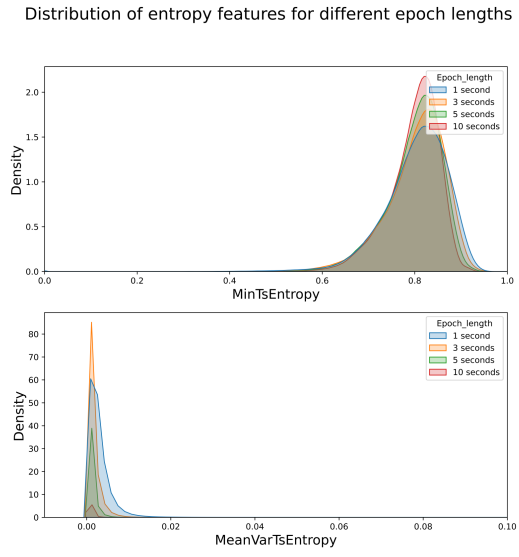
(c)



(d)



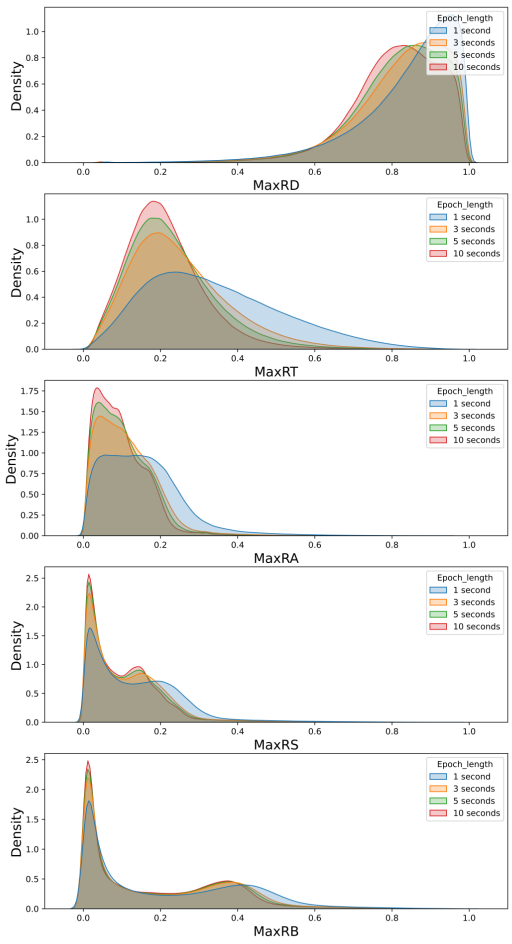
(e)



(f)

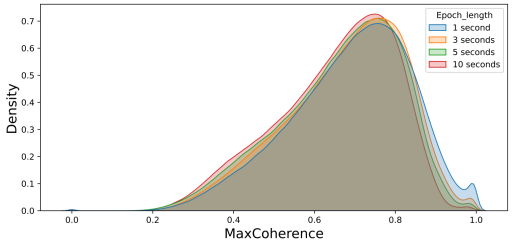
Figure C.1: Continuation: Distribution of feature values calculated for a window-length of 1 (blue), 3 (orange), 5 (green), and 10 (red) seconds.

Distribution of FreqbandRatioTotal features for different epoch lengths



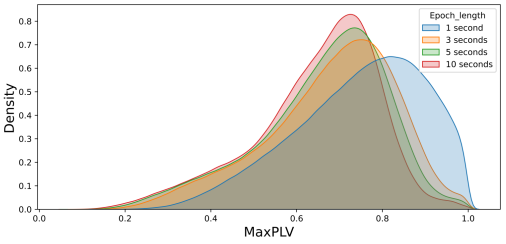
(g)

Distribution of coherence features for different epoch lengths



(h)

Distribution of PLV features for different epoch lengths



(i)

Figure C.1: Continuation: Distribution of feature values calculated for window-length of 1 (blue), 3 (orange), 5 (green), and 10 (red) seconds.

D

Statistical measures

Sensitivity

Sensitivity is a measure of how well a classifier can identify true positives. It is defined as the proportion of actual positives that are correctly identified as such by the classifier. It can also be called sensitivity or true positive rate. The formula for sensitivity is:

$$Sensitivity = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{D.1}$$

Specificity

Specificity is a measure of how well a classifier can identify true negatives. It is defined as the proportion of actual negatives that are correctly identified as such by the classifier. The formula for specificity is:

$$Specificity = \frac{True\ Negatives}{True\ Negatives + False\ Positives} \tag{D.2}$$

Accuracy

Accuracy is a measure of how well a classifier can correctly classify both positive and negative cases. It is defined as the proportion of correct classifications made by the classifier. The formula for accuracy is:

$$Accuracy = \frac{True\ Positives + True\ Negatives}{True\ Positives + False\ Positives + True\ Negatives + False\ Negatives} \tag{D.3}$$

Precision

Precision is a measure of the ability of a classifier to correctly identify positive cases among all classified positive cases. It is defined as the proportion of true positive cases among all cases classified as positive by the classifier. The formula for precision is:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{D.4}$$

F1-score

The F1-score is a measure that combines precision and sensitivity into a single metric. It is defined as the harmonic mean of precision and sensitivity. The formula for F1-score is:

$$F1\text{-}score = 2 \cdot \frac{Precision \cdot Sensitivity}{Precision + Sensitivity} \tag{D.5}$$

F_β-score

The F-beta score is defined as:

$$F_\beta = (1 + \beta^2) \cdot \frac{\text{Precision} \cdot \text{Sensitivity}}{(\beta^2 \cdot \text{Precision}) + \text{Sensitivity}} \tag{D.6}$$

where β is a parameter that controls the weight given to precision versus sensitivity. A value of $\beta = 1$ gives equal weight to precision and sensitivity (i.e., the F1 score), while higher values of β favour sensitivity over precision.

AUC

The AUC is the area under the ROC curve, which is a plot of the true positive rate (TPR) against the false positive rate (FPR) at different threshold values. The TPR is plotted on the y-axis and represents the proportion of true positives correctly identified by the classifier, while the FPR is plotted on the x-axis and represents the proportion of false positives incorrectly identified by the classifier. The AUC is a single number that summarises the classifier's performance over all possible thresholds. The AUC ranges from 0 to 1, with 0.5 indicating a random classifier and 1 indicating a perfect classifier.

Mann-Whitney U Test

The Mann-Whitney U test is a non-parametric statistical test used to compare two independent samples. It is used when the assumptions of normality and equal variance are not met. The test statistic U is calculated as follows:

$$U = R_1 - \frac{n_1(n_1 + 1)}{2} \tag{D.7}$$

where R_1 is the sum of the ranks of the observations in sample 1, n_1 is the sample size of sample 1, and $\frac{n_1(n_1+1)}{2}$ is a constant that represents the sum of the ranks of a sample of size n_1.

E

Distribution of feature values

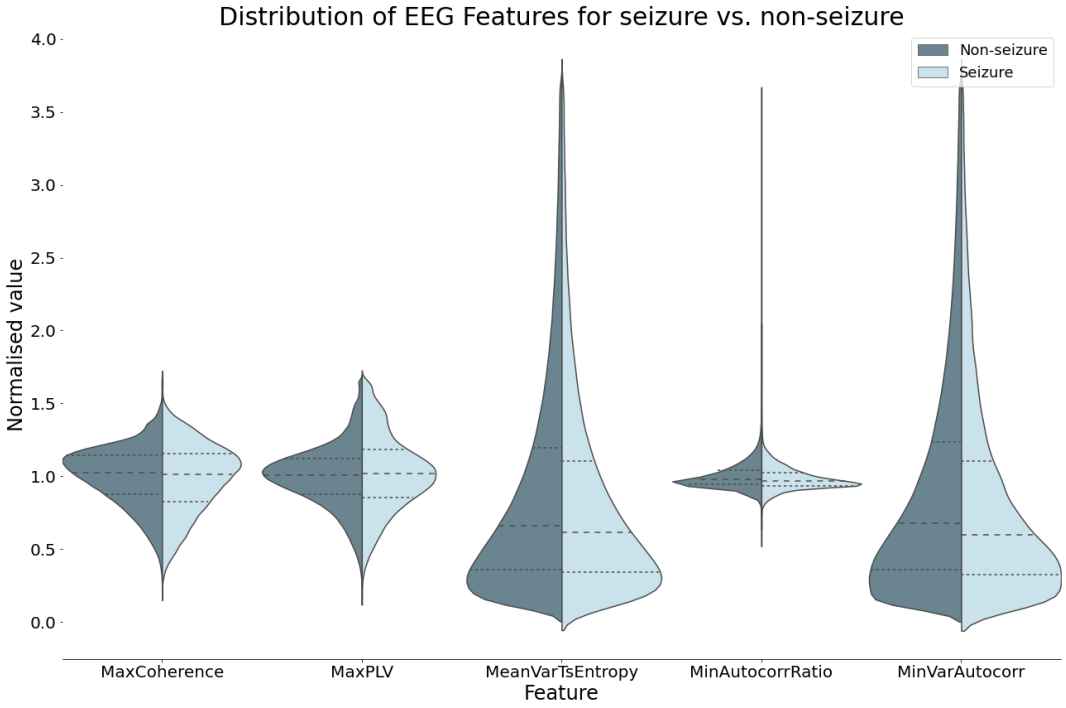


Figure E.1: Violin plots of the coherence, PLV, variance in entropy and autocorrelation features, with upper and lower 1% outliers removed. The distributions of the values of these features are shown for both the samples of the non-seizure (dark blue) and seizure (light blue) groups. The striped lines present the medians and the dotted lines indicate the interquartile ranges.

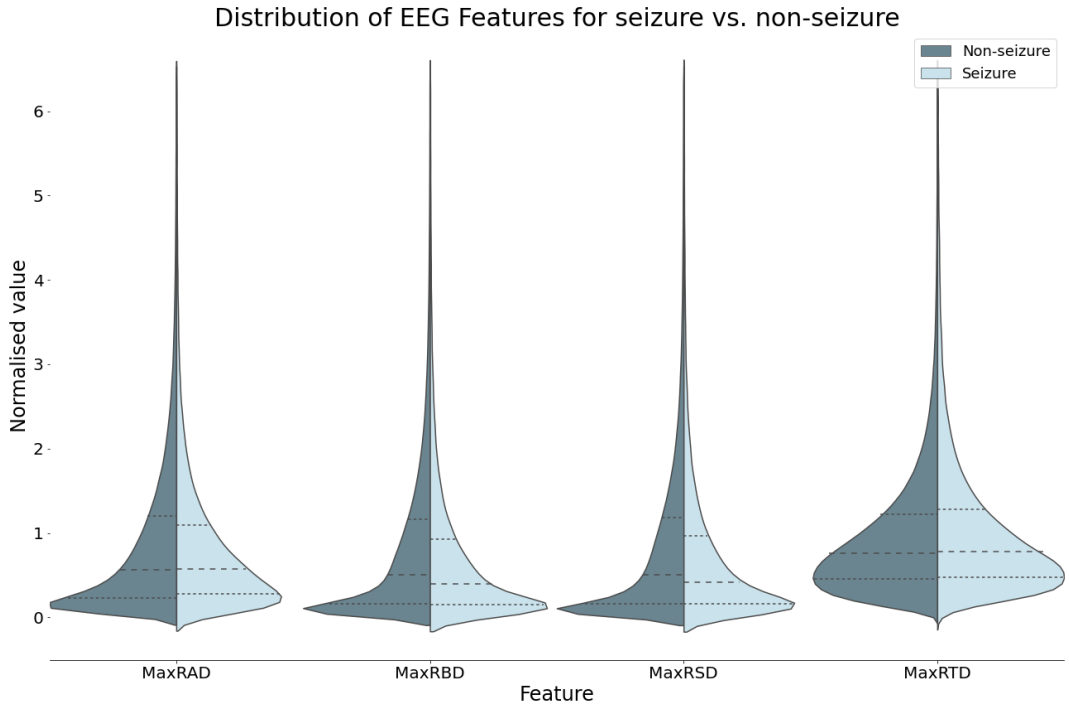


Figure E.2: Violin plots of the maximum ratio-to-delta features, with upper and lower 1% outliers removed. The distributions of the values of these features are shown for both the samples of the non-seizure (dark blue) and seizure (light blue) groups. The striped lines present the medians and the dotted lines indicate the interquartile ranges.

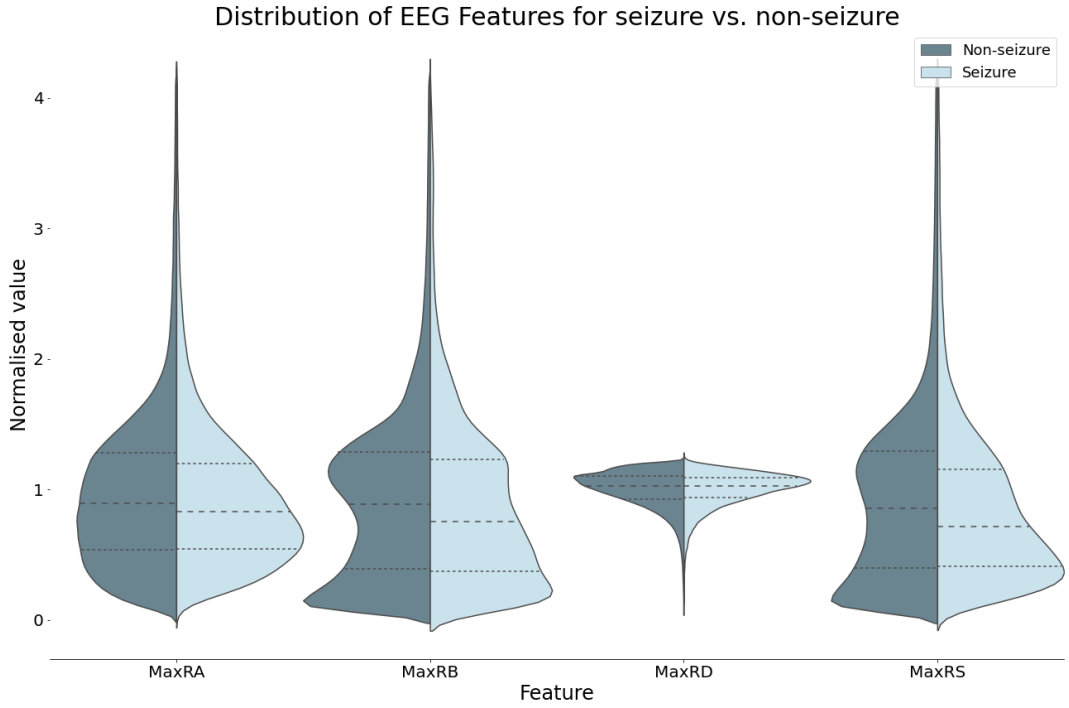


Figure E.3: Violin plots of the maximum ratio-to-total features, with upper and lower 1% outliers removed. The distributions of the values of these features are shown for both the samples of the non-seizure (dark blue) and seizure (light blue) groups. The striped lines present the medians and the dotted lines indicate the interquartile ranges.

F

Statistics individual analysis

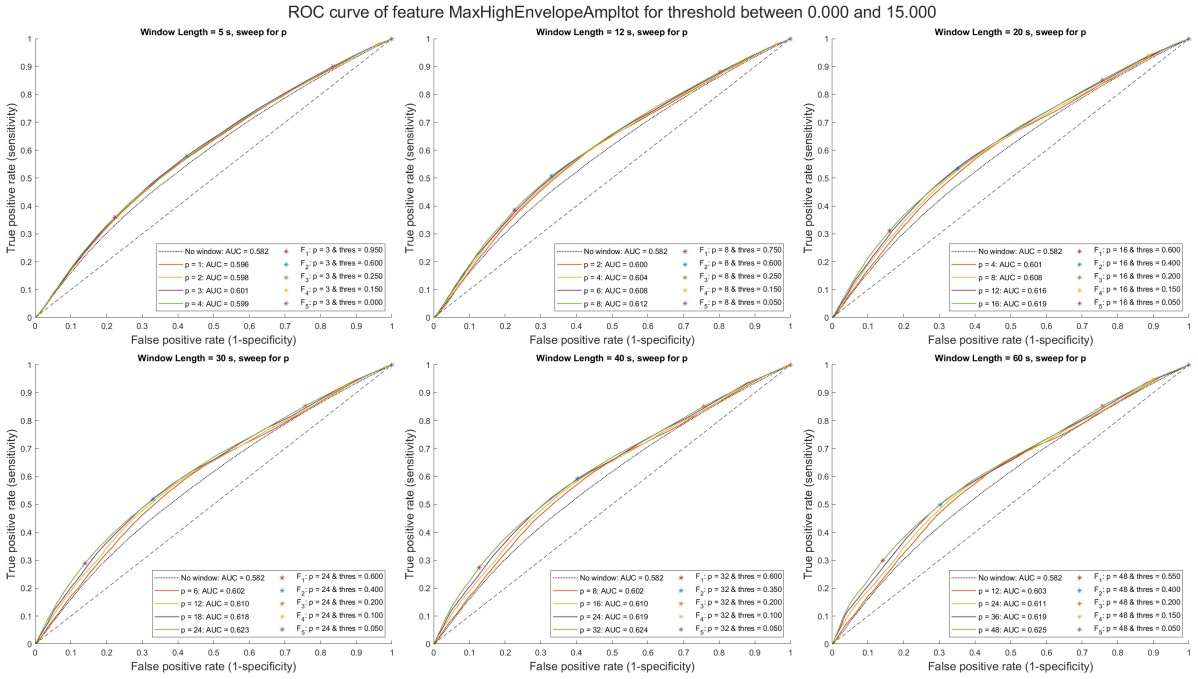
Table F.1: Statistical comparison of feature values between seizures and non-seizures

Feature	**Median (Q1-Q3)** Non-event	**Median (Q1-Q3)** Event	**p-value**
Maximum amplitude	0.59 (0.30-1.17)	0.79 (0.39-1.58)	<0.001
Variance in amplitude	0.64 (0.38-1.1)	1.08 (0.51-2.17)	<0.001
aEEG	0.97 (0.91-1.05)	1.13 (1.01-1.27)	<0.001
Variance in aEEG	0.54 (0.26-1.09)	0.57 (0.30-1.12)	<0.001
Maximum power			
Delta band	0.53 (0.25-1.12)	1.02 (0.41-2.28)	<0.001
Theta band	0.57 (0.32-0.99)	1.01 (0.48-2.25)	<0.001
Alpha1 band	0.60 (0.33-1.01)	0.82 (0.52-2.01)	<0.001
Alpha2 band	0.64 (0.33-1.07)	0.86 (0.54-1.87)	<0.001
Beta band	0.62 (0.28-1.01)	0.97 (0.57-1.69)	<0.001
Power ratio to delta			
Theta-delta	0.76 (0.45-1.23)	0.78 (0.48-1.31)	<0.001
Alpha1-delta	0.57 (0.23-1.23)	0.58 (0.28-1.11)	<0.001
Alpha2-delta	0.49 (0.15-1.20)	0.43 (0.16-1.00)	<0.001
Beta-delta	0.50 (0.15-1.18)	0.41 (0.15-0.95)	<0.001
Power ratio to total			
Delta-total	1.03 (0.93-1.11)	1.03 (0.94-1.09)	<0.001
Theta-total	0.91 (0.64-1.26)	1.01 (0.73-1.39)	<0.001
Alpha1-total	0.90 (0.54-1.29)	0.84 (0.55-1.21)	<0.001
Alpha2-total	0.86 (0.38-1.31)	0.73 (0.41-1.17)	<0.001
Beta-total	0.89 (0.37-1.3)	0.75 (0.37-1.24)	<0.001
Autocorrelation peak-ratio	0.98 (0.94-1.04)	0.97 (0.94-1.03)	<0.001
Variance in autocorr. peak-ratio	0.67 (0.35-1.27)	0.58 (0.30-1.11)	<0.001
Entropy	1.01 (0.96-1.06)	0.99 (0.93-1.04)	<0.001
Variance in entropy	0.67 (0.35-1.24)	0.62 (0.33-1.14)	<0.001
Coherence	1.03 (0.88-1.14)	1.01 (0.82-1.16)	<0.001
Phase locking value	1.01 (0.88-1.12)	1.02 (0.86-1.19)	<0.001

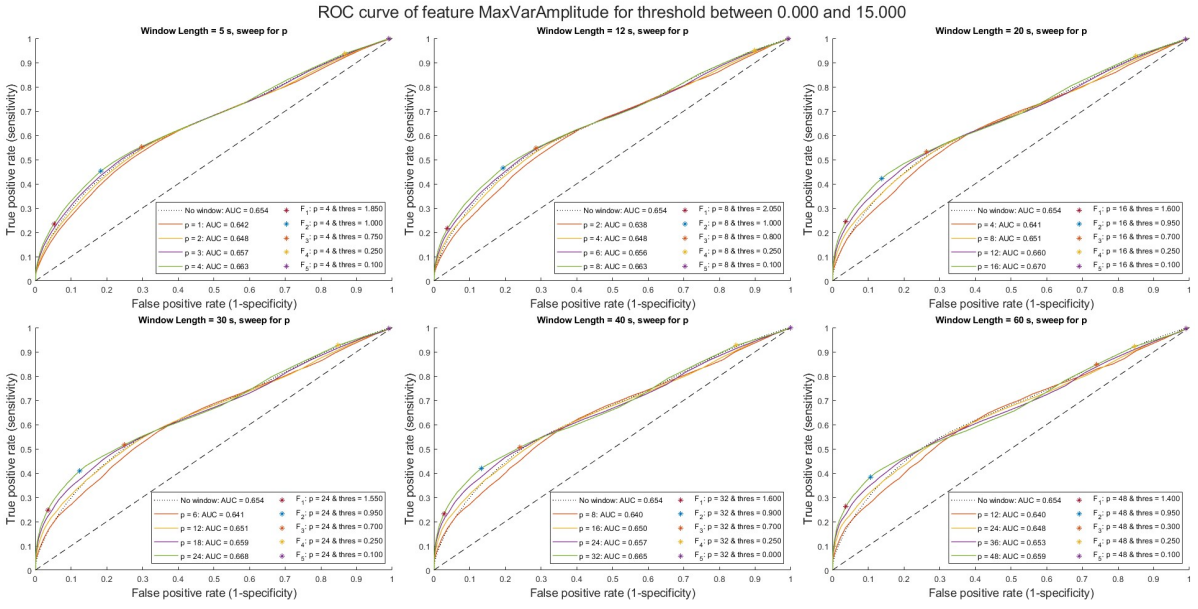
aEEG = Amplitude-integrated electroencephalography, Q1-Q3 = interquartile range, $p<0.05$ is deemed significant

G

ROC for threshold and window-sum condition sweep

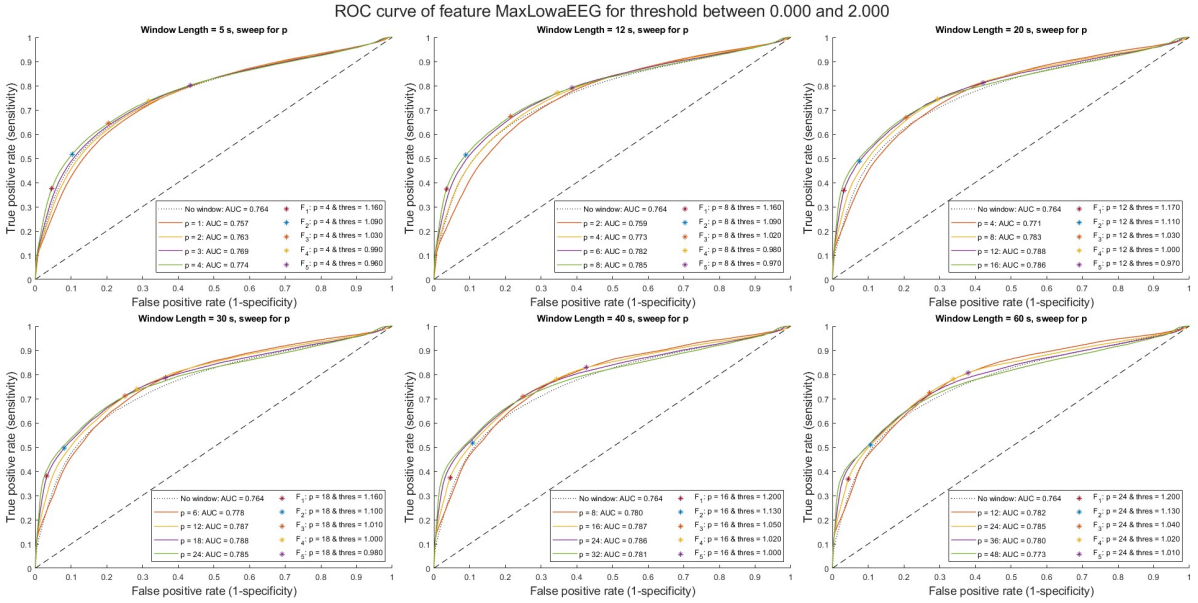


(a)

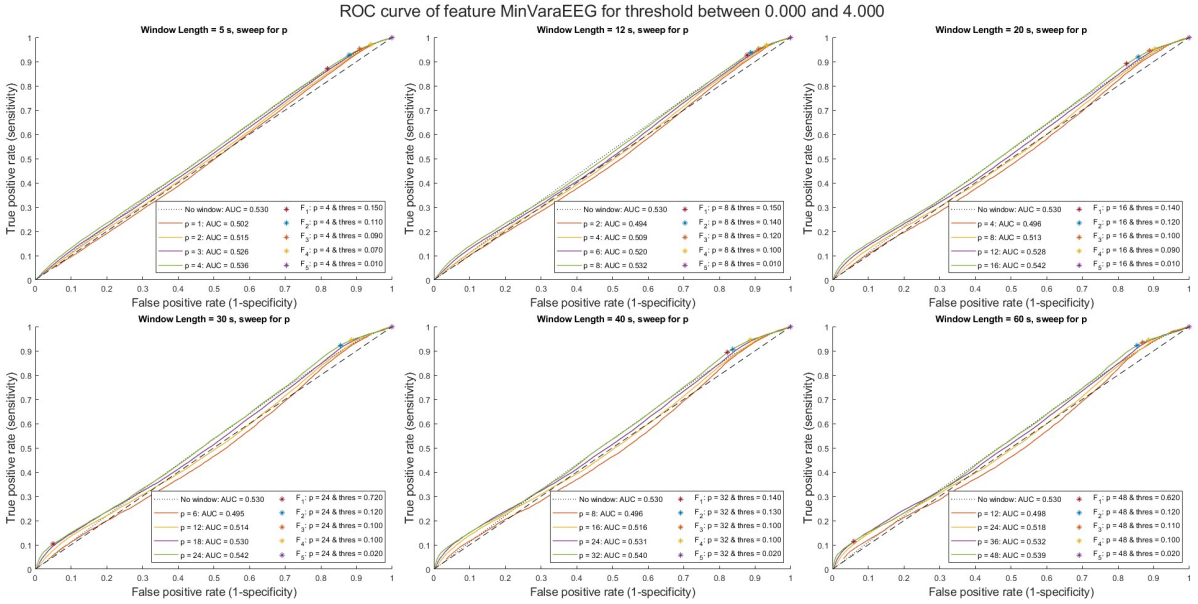


(b)

Figure G.1: ROC curves of a) the maximum amplitude and b) variance in amplitude feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

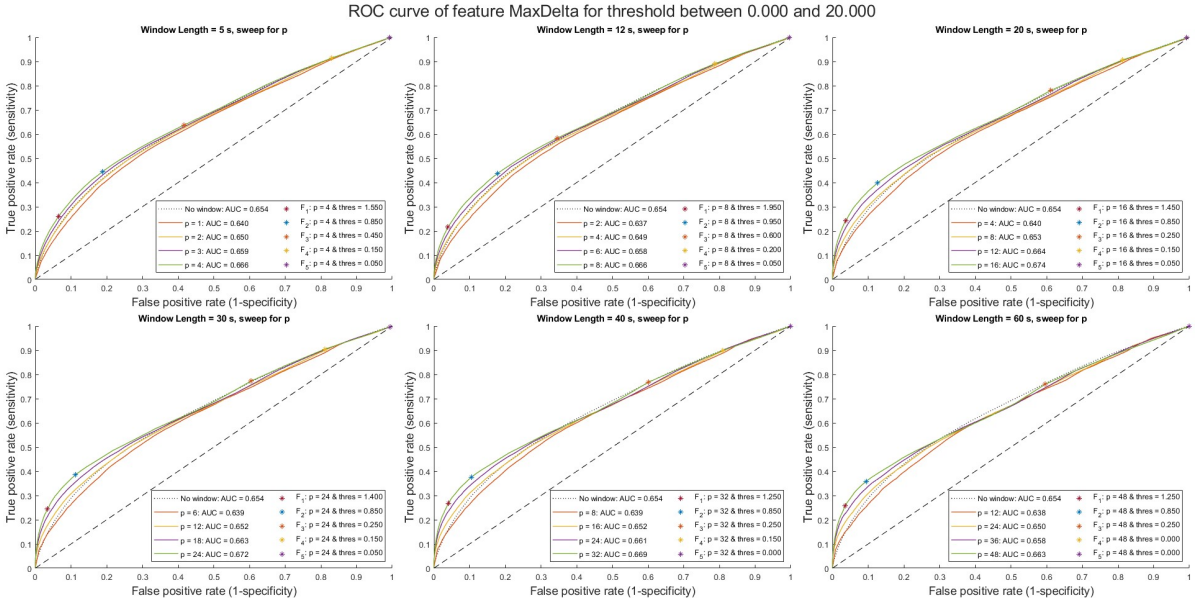


(c)

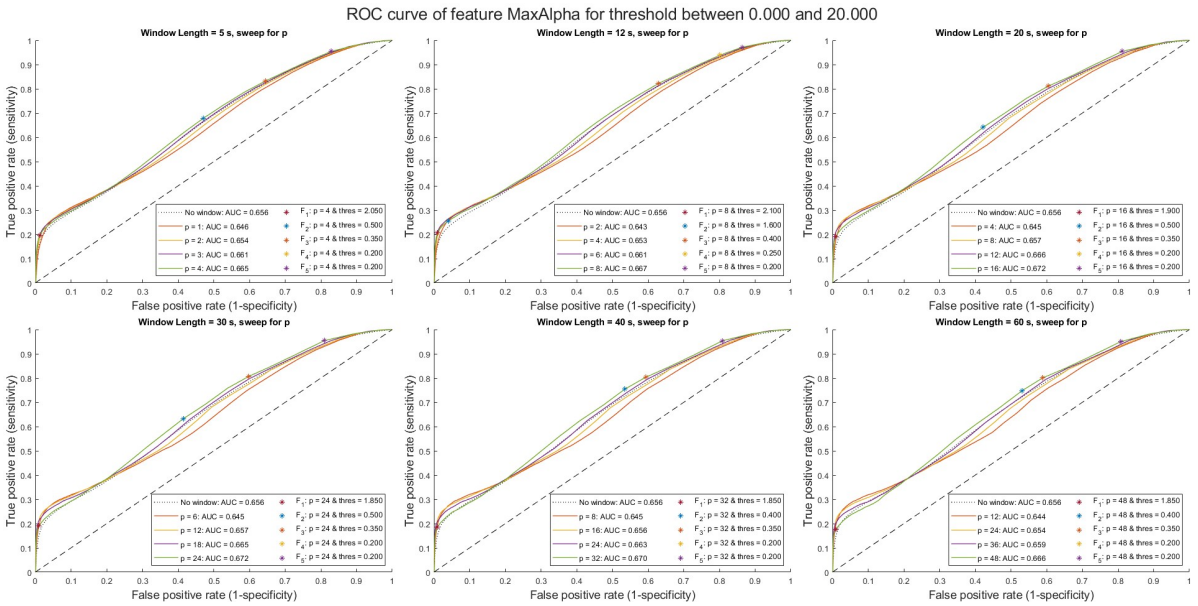


(d)

Figure G.1: Continuation: ROC curves of c) the maximum lower border and d) variance in aEEG feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

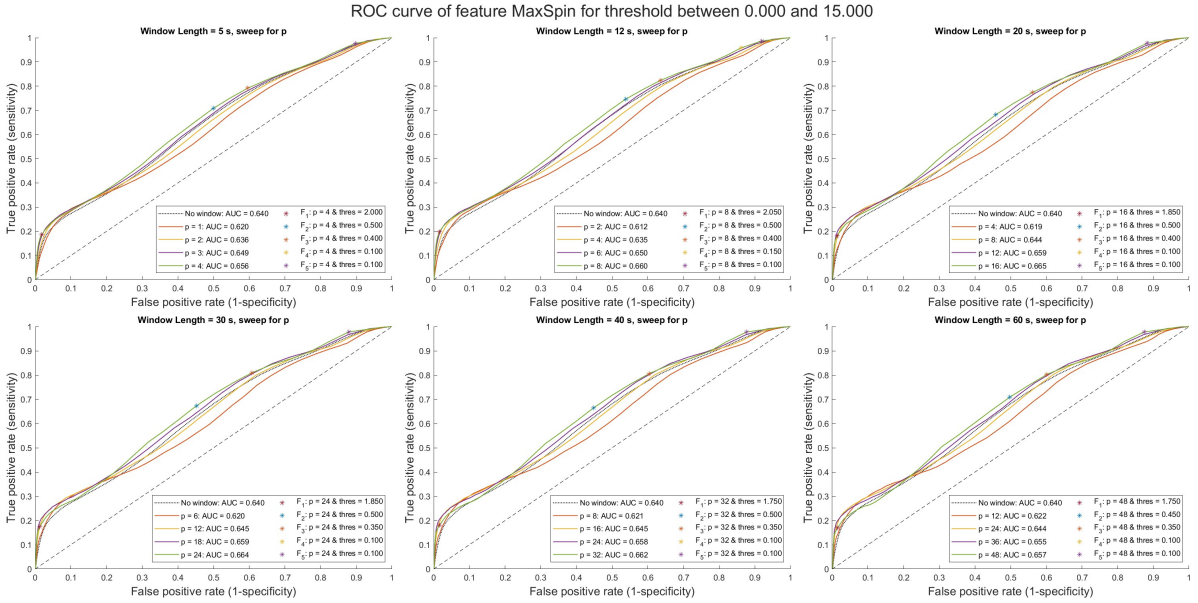


(e)

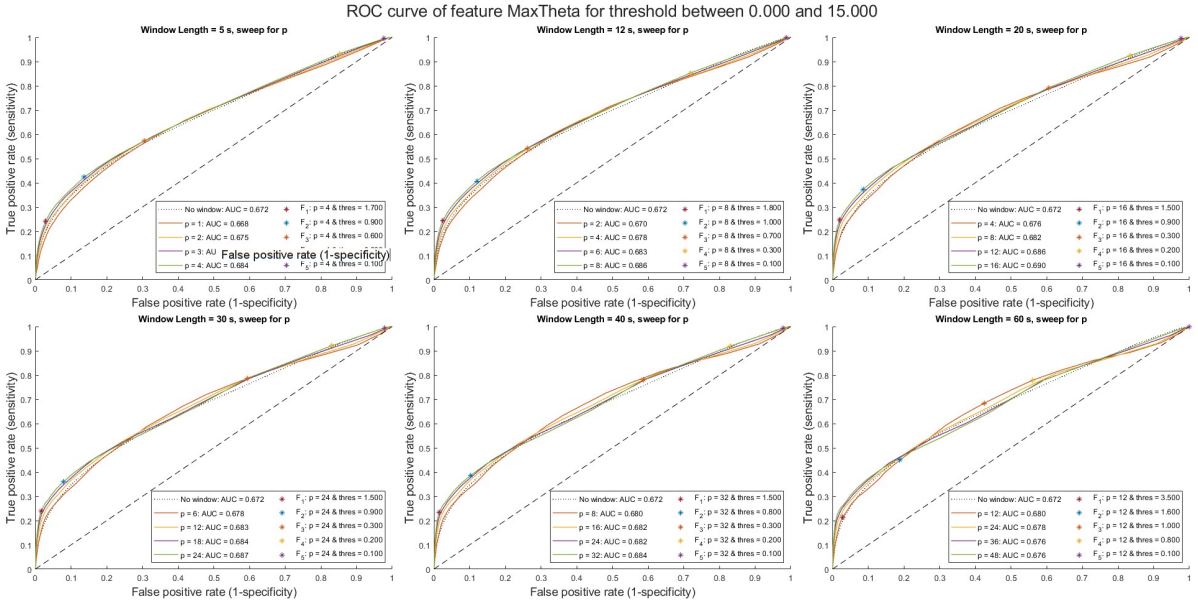


(f)

Figure G.1: Continuation: ROC curves of e) the maximum delta power and f) alpha1 power feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

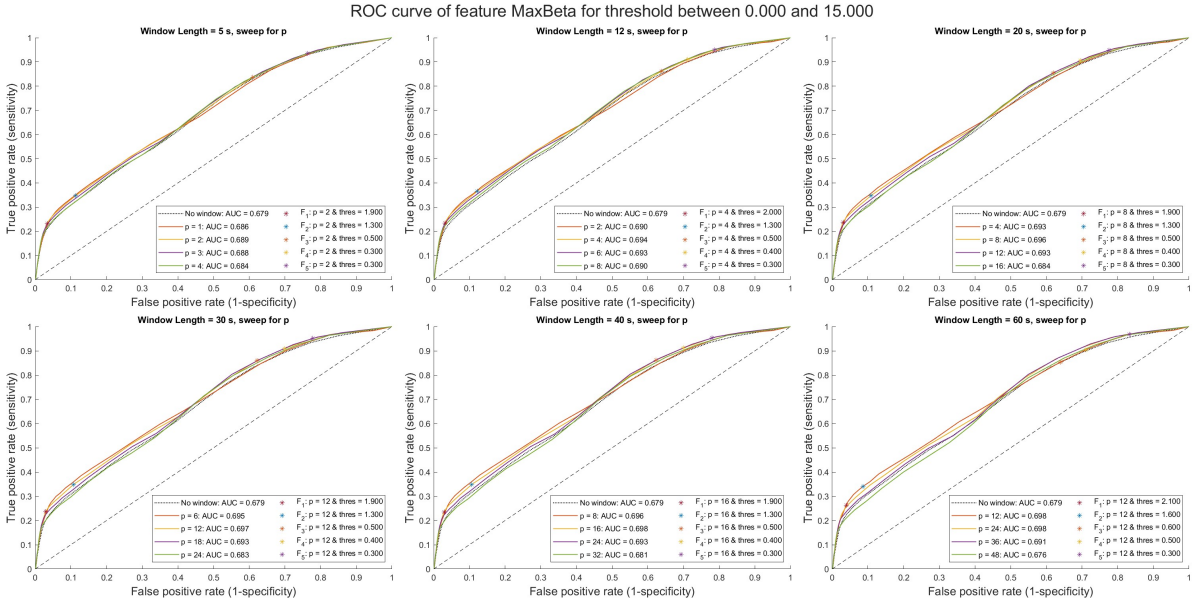


(g)

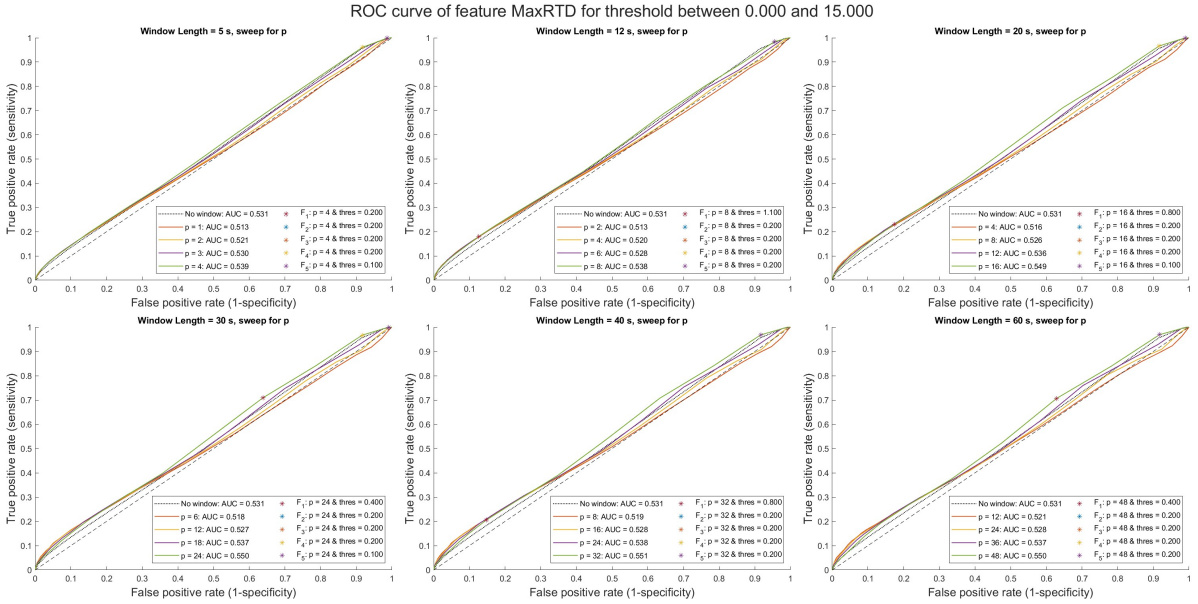


(h)

Figure G.1: Continuation: ROC curves of g) the maximum alpha2 power and h) theta power feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

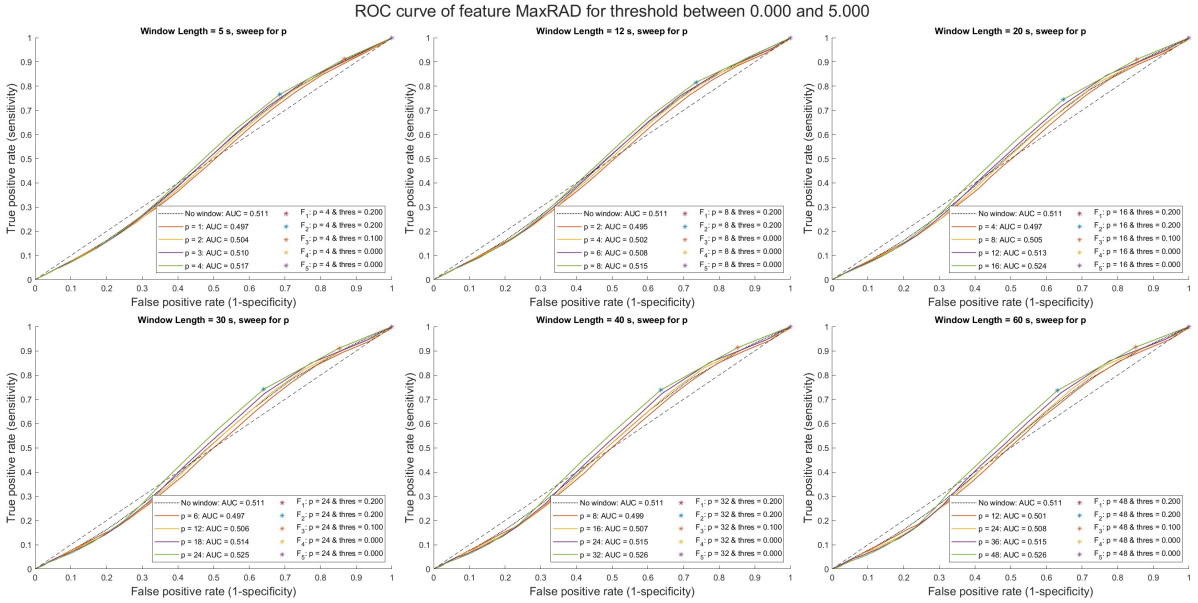


(i)

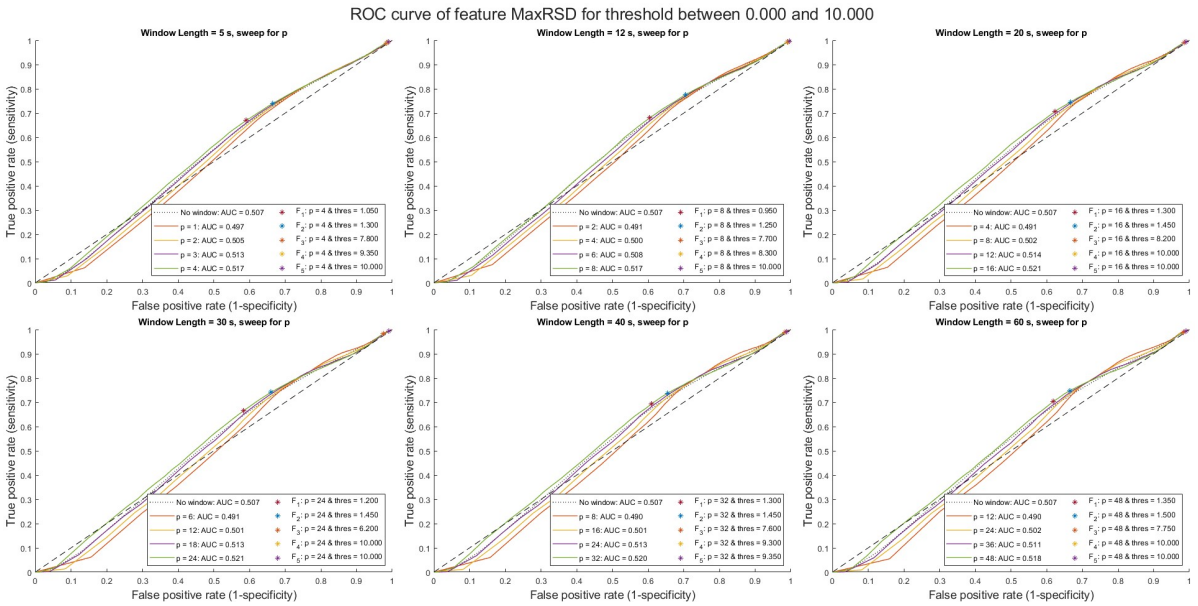


(j)

Figure G.1: Continuation: ROC curves of i) the maximum beta power and j) theta-delta power ratio feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

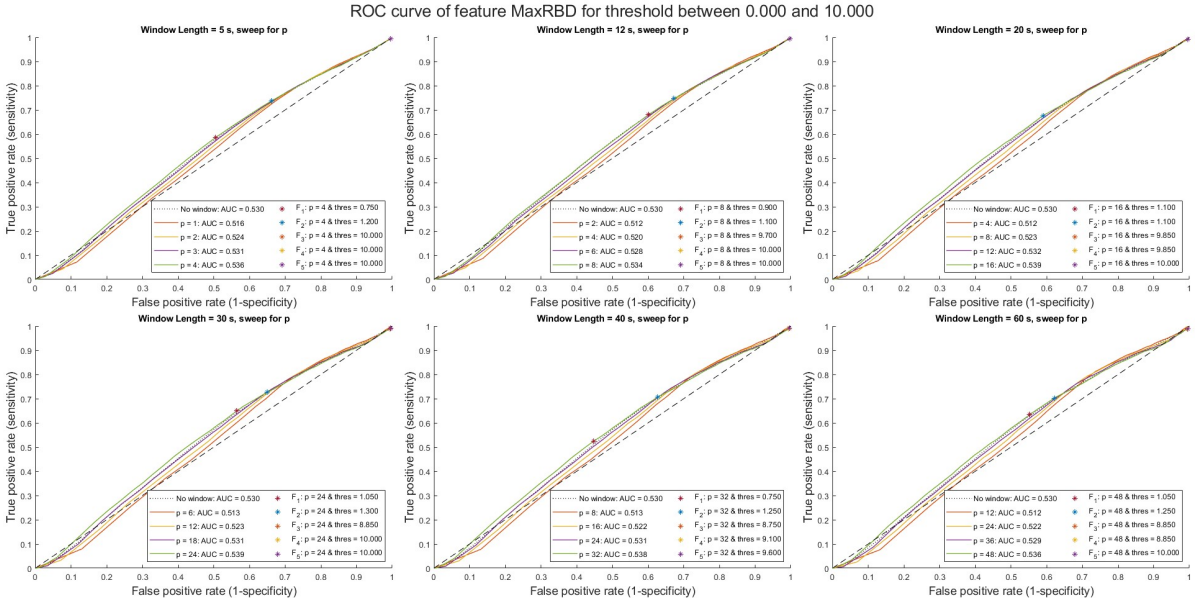


(k)

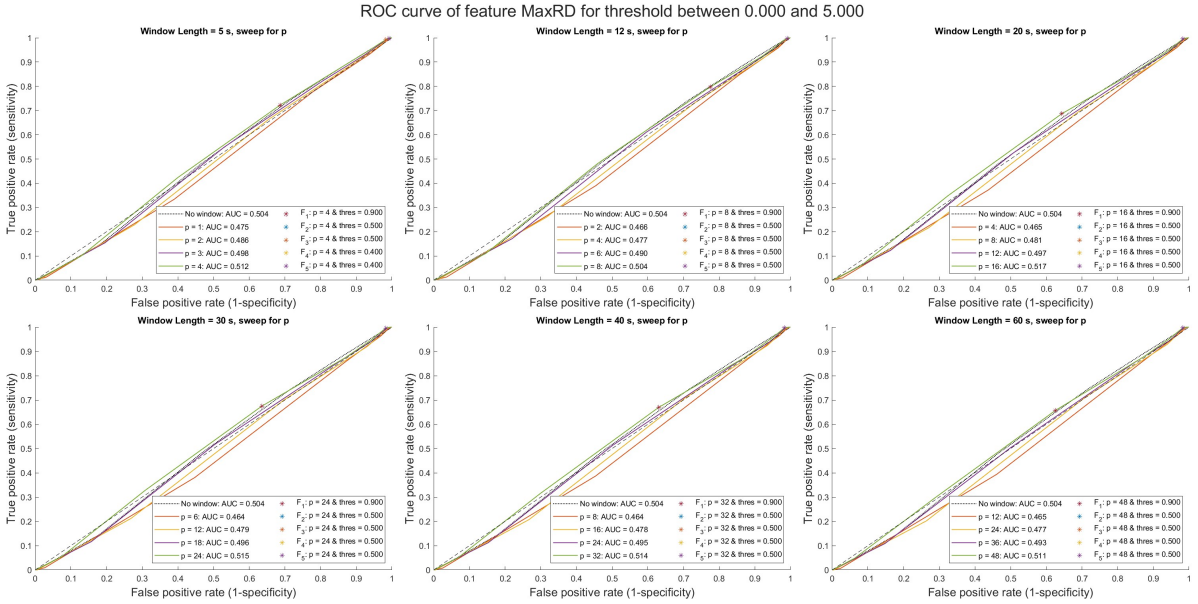


(l)

Figure G.1: Continuation: ROC curves of k) the maximum alpha-delta and l) alpha2-delta power ratio feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

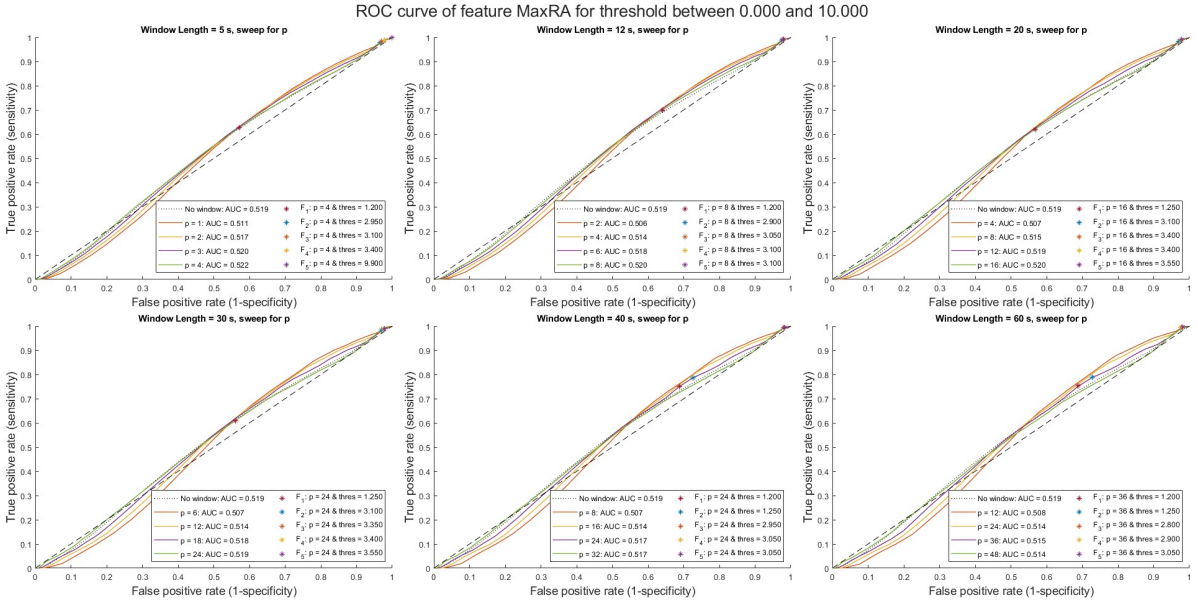


(m)

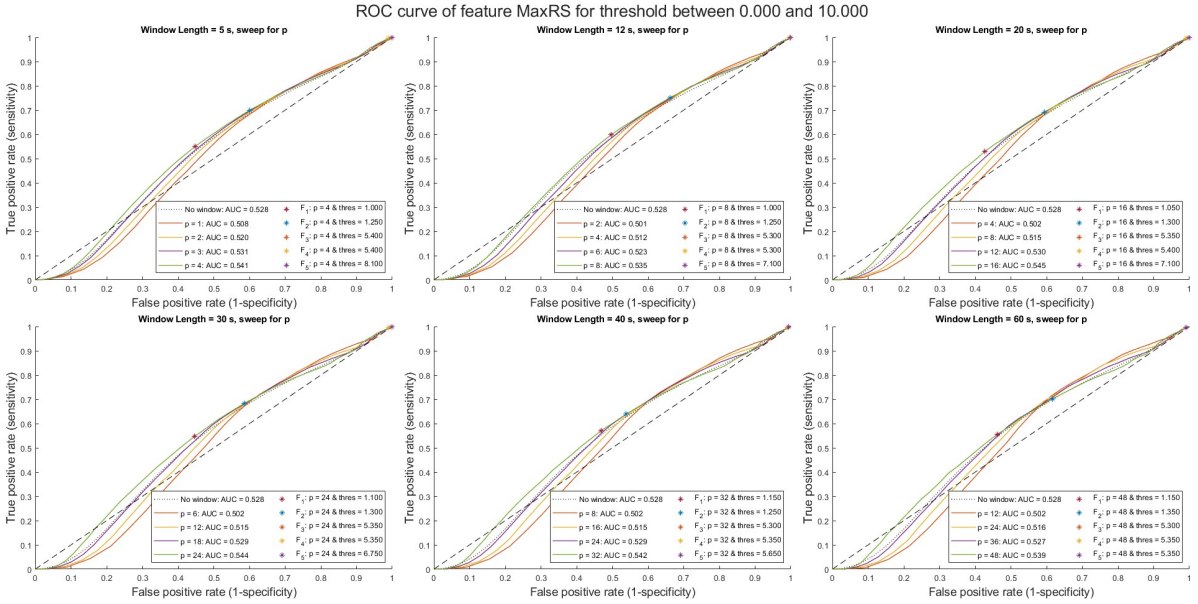


(n)

Figure G.1: Continuation: ROC curves of m) the maximum beta-delta and n) delta-total power ratio feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

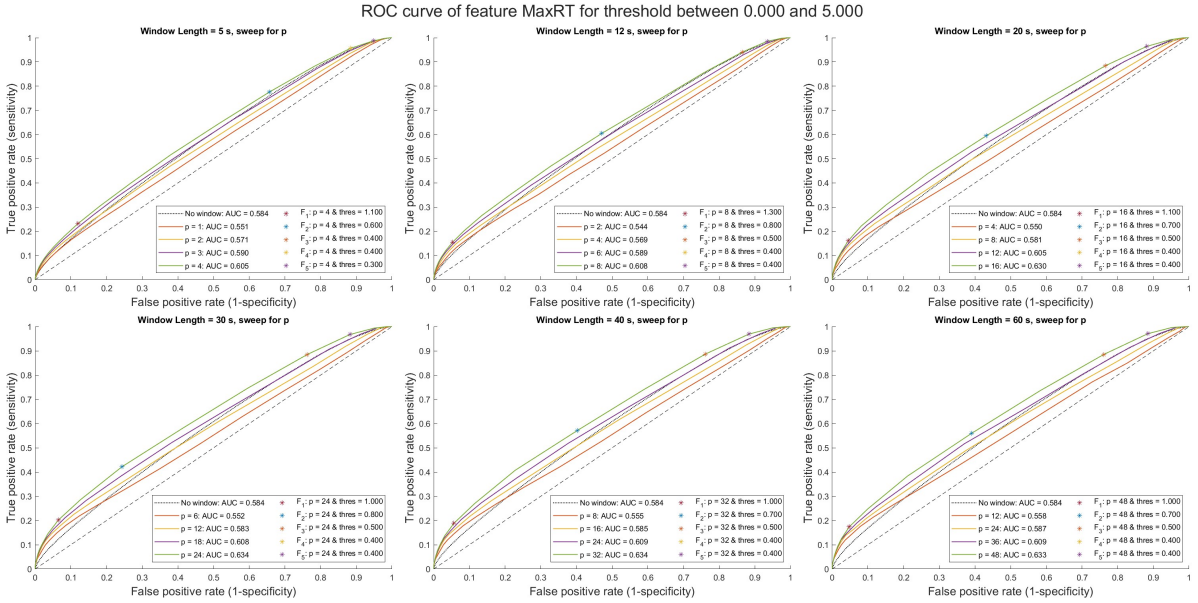


(o)

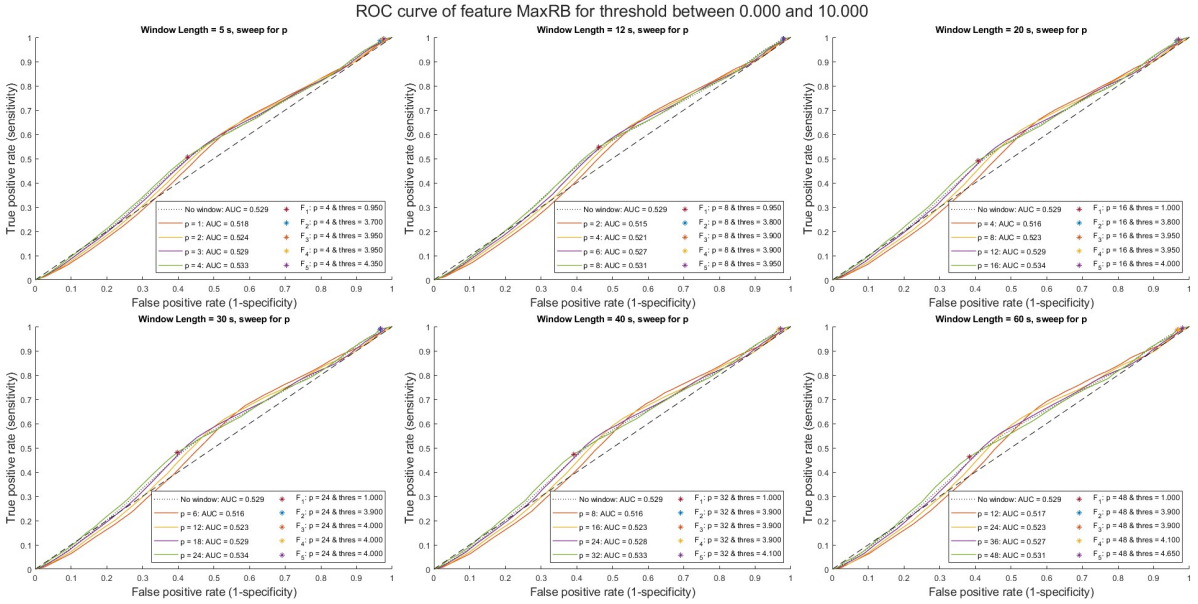


(p)

Figure G.1: Continuation: ROC curves of o) the maximum alpha-total and p) alpha2-total power ratio feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

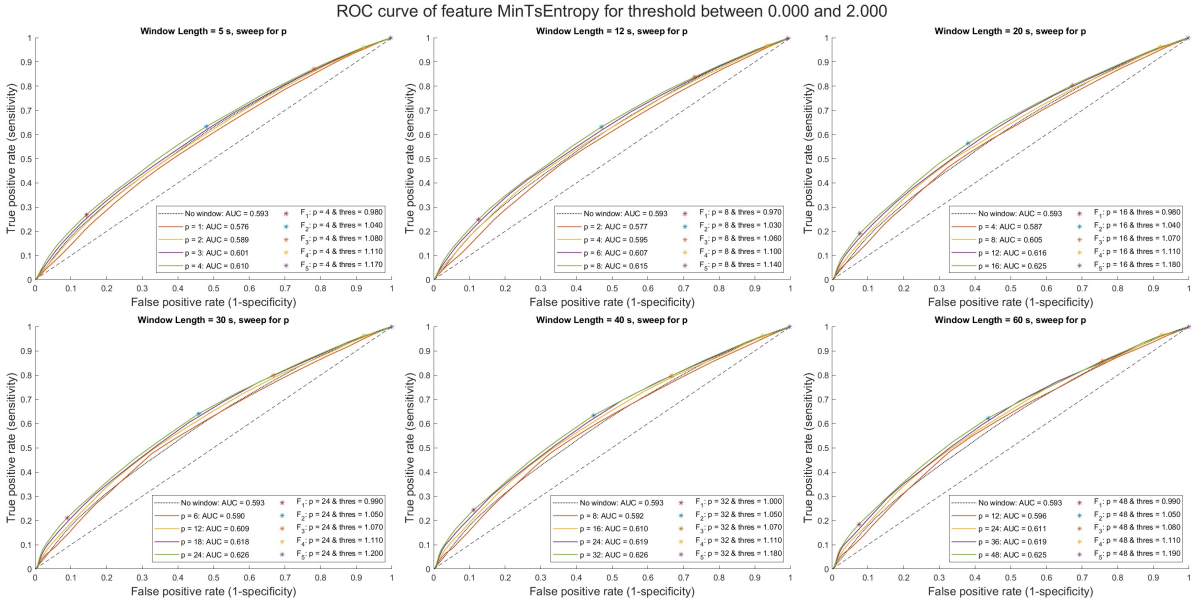


(q)

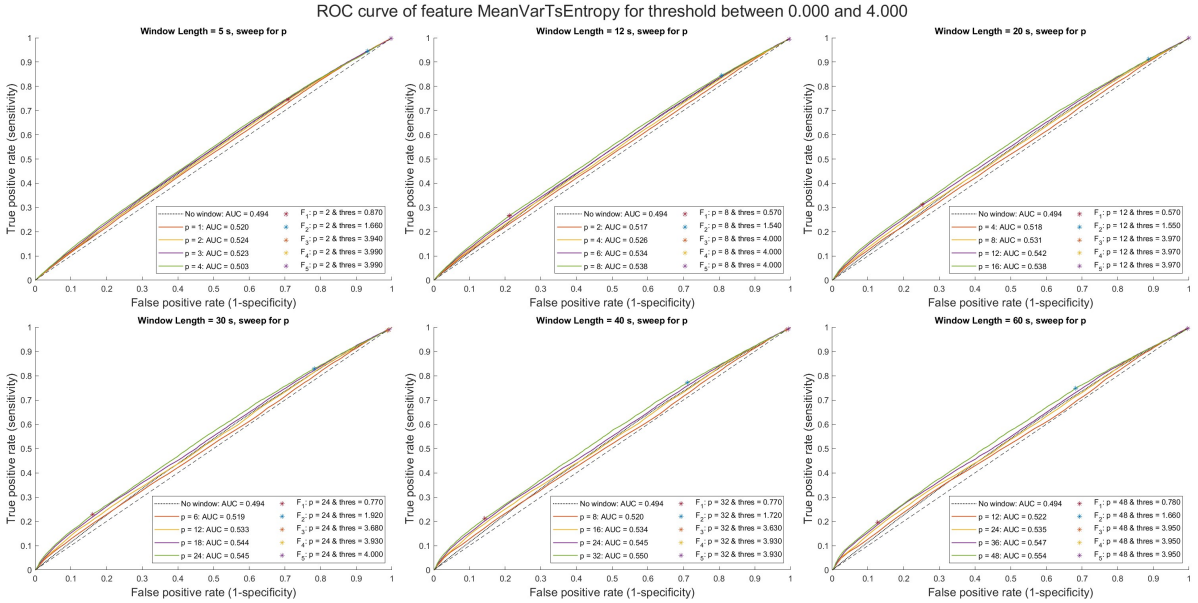


(r)

Figure G.1: Continuation: ROC curves of q) the maximum theta-total and r) beta-total power ratio feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

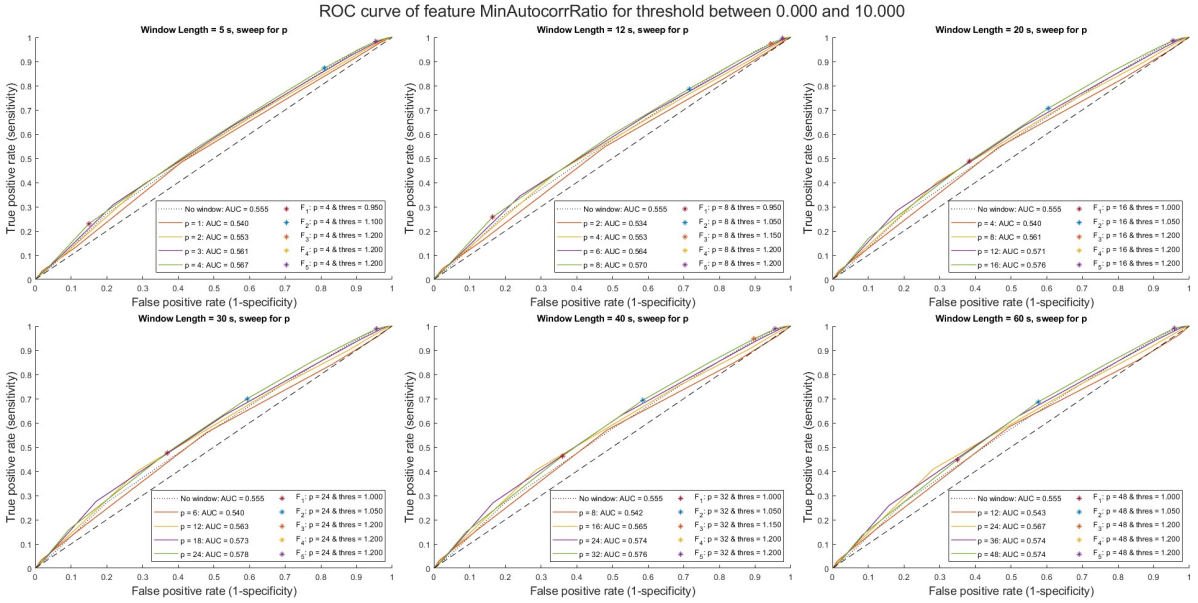


(s)

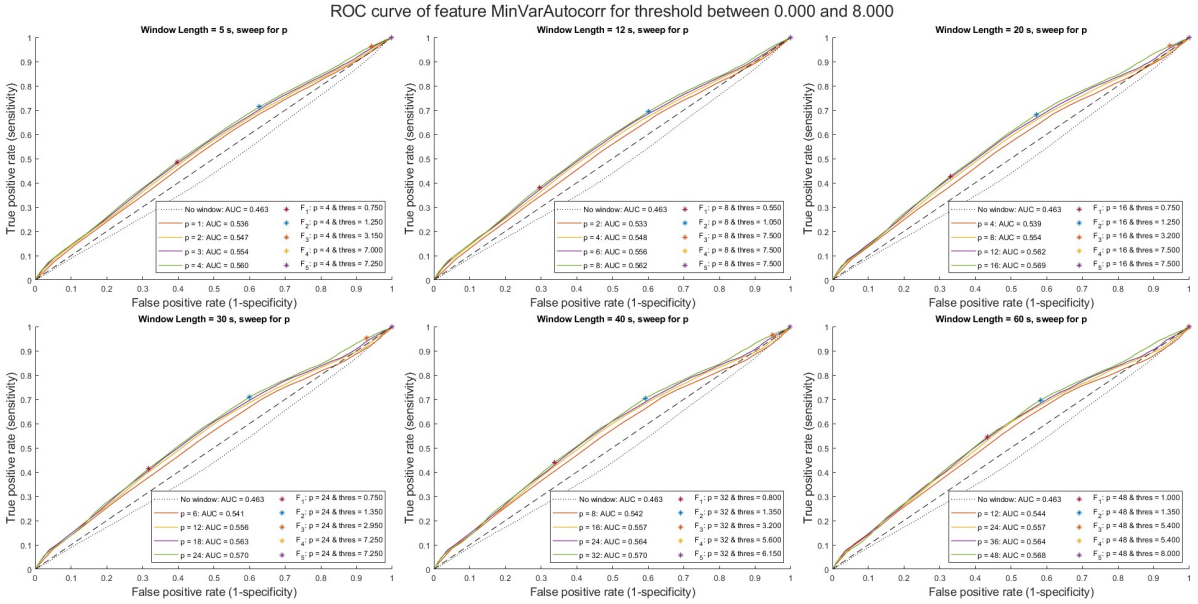


(t)

Figure G.1: Continuation: ROC curves of s) the minimum entropy and t) mean variance in entropy feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

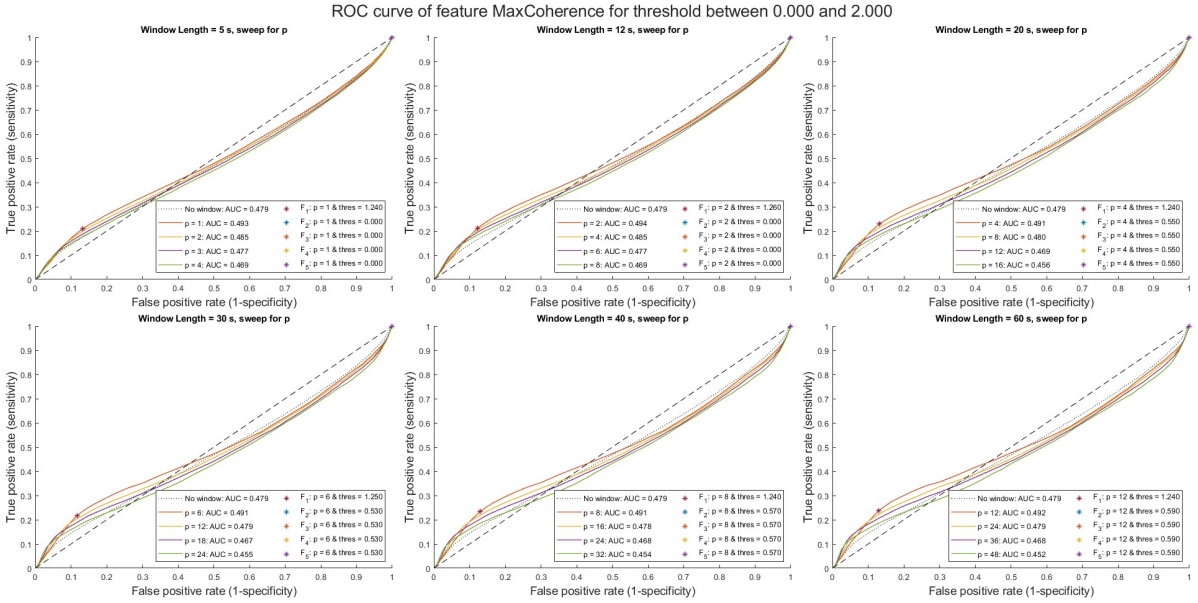


(u)

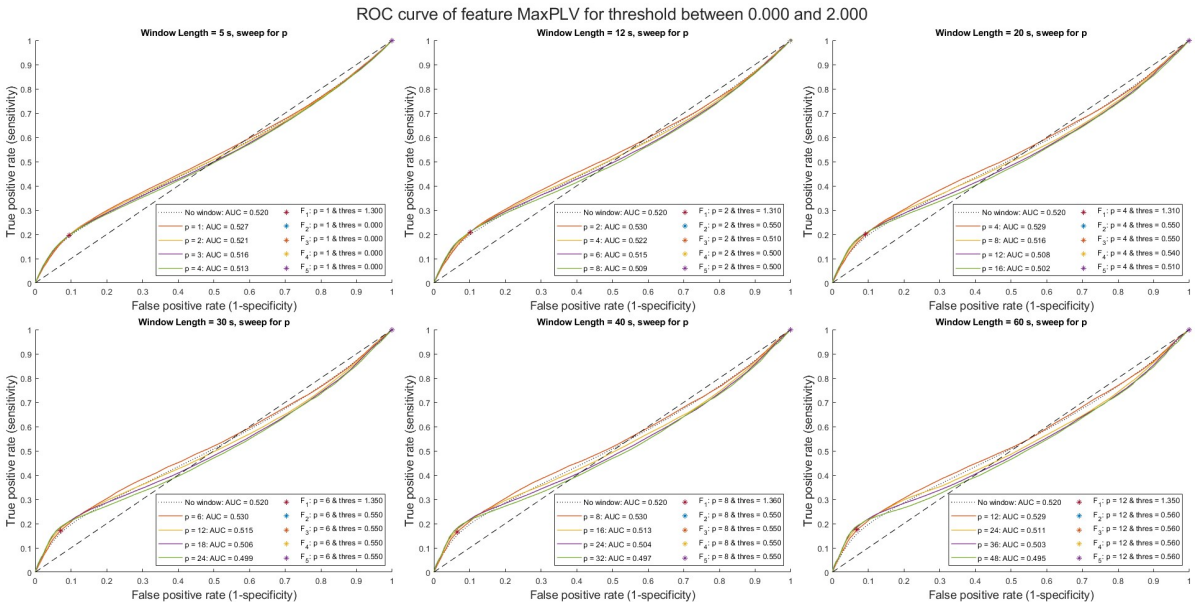


(v)

Figure G.1: Continuation: ROC curves of u) the minimum autocorrelation and v) variance in autocorrelation feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.



(w)



(x)

Figure G.1: Continuation: ROC curves of w) the maximum coherence and x) PLV feature, for a variety of window-sum criteria, in which a range of thresholds is exploited. For these results, the threshold was based on the mean of the entire EEG. The stars indicate various F-beta scores, located in the ROC with the relative best AUC score.

H

Information on random forest classifier

The random forest classifier is a popular machine learning algorithm used for classification tasks. It is an ensemble algorithm that creates numerous decision trees and combines their predictions to produce a final classification result.

Hyperparameters

The hyperparameters that were optimised to improve the performance of the random forest classifier are stated below:

n_estimators: This hyperparameter specifies the number of decision trees in the forest. Increasing this value can improve the accuracy of the model, but can also make it slower and more prone to overfitting.

max_depth: This specifies the maximum depth of each decision tree in the forest. Increasing the depth could improve the accuracy, but can also cause overfitting.

min_samples_split: It determines the minimum number of samples required to split a node.

min_samples_leaf: This hyperparameter defines the minimum number of samples required to be at a leaf node. Increasing the value for $min_samples_split$ and $min_samples_leaf$ can both make the model more conservative and reduce overfitting, but can also make it less sensitive to smaller differences in the data.

bootstrap: Bootstrapping is a technique that involves randomly sampling the data with replacement to create multiple datasets of equal size to the original data. Each decision tree is then trained on one of these bootstrap datasets, and their predictions are combined to make a final prediction.

criterion: The criterion-hyperparameter determines the function used to measure the quality of a split in a decision tree. The two options for this hyperparameter are 'gini 'and 'entropy'. Gini measure the probability of incorrectly classifying a randomly chosen sample. Entropy measures the randomness in the split.

Stratified k-fold

The StratifiedGroupKFold function is a variant of K-fold cross-validation. It splits the data into K (in our case 5) test and training sets, ensuring that each fold contains a representative sample the data, without overlapping samples in the 5 folds. In addition, a patient can either be in the test or in the train set and not split between them.

Randomised search tuning hyperparameters

Provided with a range of values for the various hyperparameters, the randomised search picks a random combination of these variables and fits the model the data to the model. This process is repeated for a predefined number of iterations, with each time a different combination of parameters.

Feature importance

The importance of each feature is determined by first calculating a baseline metric, which is defined by the scoring method, in our case the impurity and accuracy of the classifier. Next, one feature column is selected from the validation set and its values are permuted (i.e., randomly shuffled). Then, the metric is recalculated using the permuted feature column. By comparing the difference in the metric values between the original and permuted feature column, the permutation importance of the feature can be determined. A larger difference in the metric values indicates greater importance of the feature in the model's performance.

Results rf-model based on 10 patients

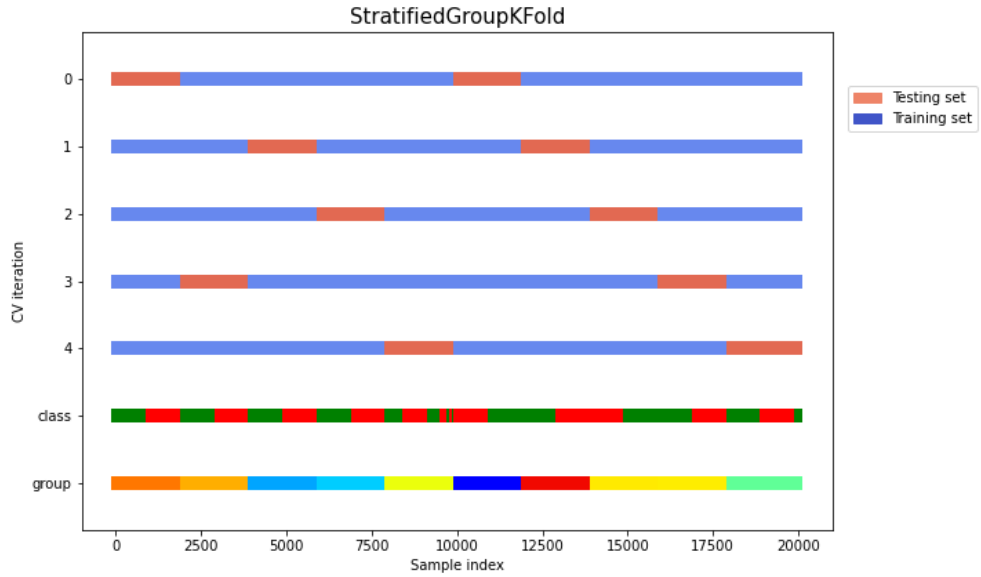


Figure H.1: Figure visualising the split of the data using a stratified 5-fold split. The first five rows represent the resampled data set with in orange the test set and in blue the training set for each fold. The sixth row represents the distribution of the seizures versus non-seizures in the data set. The group variable portrays all different patients. In each fold, each patient either belongs to the test or training set.

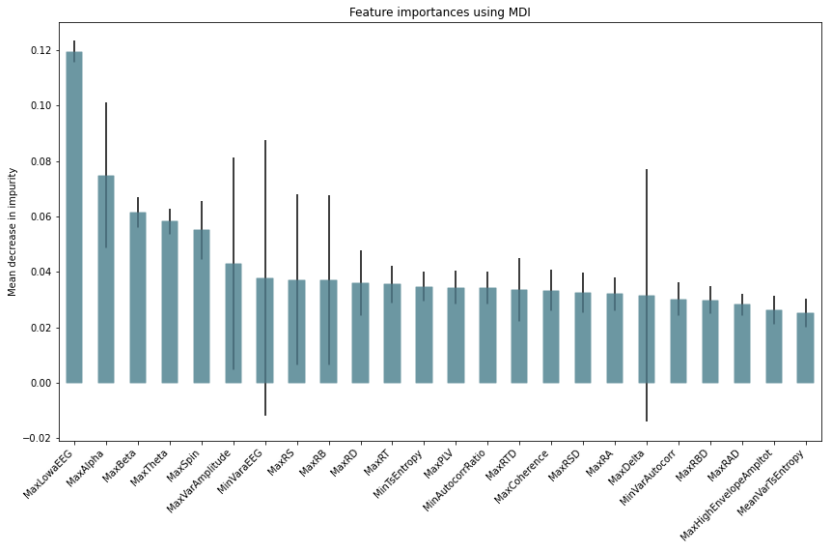


Figure H.2: Bar graph with the importance of each feature in the random classifier. On the x-axis all features are displayed and on the y-axis the mean decrease in impurity in performance between when the features are included in a regular way and in a permuted manner.

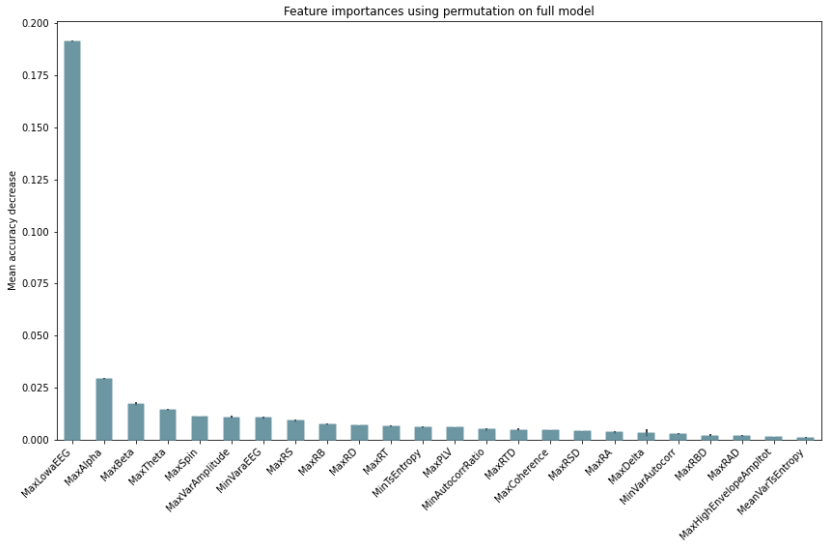


Figure H.3: Bar graph with the importance of each feature in the random classifier. On the x-axis all features are displayed and on the y-axis the mean decrease in accuracy in performance between when the features are included in a regular way and in a permuted manner.

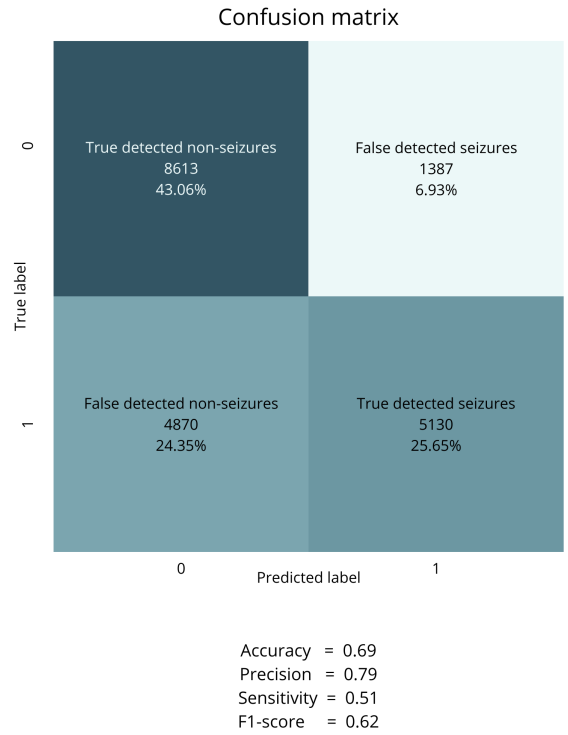


Figure H.4: Confusion matrix of the random forest classifier, showing the predicted class labels versus the actual class labels for the test data set. 0 indicates a non-seizure and 1 a seizure. Further statistical values are shown beneath the table.