NTNU – Trondheim
Norwegian University of
Science and Technology

EEG Channel-Selection Method for Epileptic-Seizure Detection Using Machine Learning Techniques

Marie Øverby

December 2022

Specialisation project
Department of Engineering Cybernetics
Norwegian University of Science and Technology

Supervisor: Marta Molinas
Co-supervisor: Luis Alfredo Moctezuma

Abstract

This work presents two approaches for epileptic seizure detection. One patient-independent and one patient-dependent approach. Feature and channel reduction was done on the patient-independent approach.

Two datasets containing electroencephalographic (EEG) signals were used, CHB-MIT Scalp EEG Database [1] and Siena Scalp EEG Database [2]. The first one uses 22 patients and 22 to 38 channels for recording. The latter uses 14 patients and 22 to 28 channels. Both are from PyseoNet, [3, 4].

To detect epileptic seizure the EEG signals were first decomposed into different sub-bands using the Discrete Wavelet Transform (DWT). From these sub-bands sixteen different features were extracted. The obtained features were used as input for Random forest (RF), Gradient boosting (GB) and Support vector machine (SVM) in order to classify seizure and seizure-free periods.

For the patient-independent approach, the feature importance for each machine learning method was found, and the most important features were chosen. These were used when finding the channel importance. The performance of the three algorithms using a low number of features and an increasing number of channels is presented.

A clear conclusion on which machine learning method, features and channels gives the best performance is not presented due to variations in results for each run of the method. The standard deviation of the accuracy was about ±3%. Nevertheless, high-performance measures both for a patient-dependent and patient-independent approach are presented. An accuracy between 95.9% and 100% was obtained for the patient-dependent approach, depending on which machine learning method was used. An accuracy of 97.6%, 96.4% and 88.4% were obtained for the patient-independent approach using 1-3 features and one channel, also depending on which machine learning method is used.

Preface

This specialisation project was submitted as part of the five years Master's degree program Cybernetics and Robotics at NTNU Trondheim.

I would like to thank my supervisor Professor Marta Molinas and my co-supervisor Luis Alfredo Moctezuma for the help throughout the project.

The IDUN cluster was used for most of the computations done [5].

Contents

Chapter 1

Introduction

Epilepsy is a neurological disorder characterised by recurrent seizures. It affects almost 1% of the world's population, both genders and all ages [6]. The seizures are caused by abnormal electrical activity in the brain, which can manifest in various symptoms. Some common symptoms of epilepsy include loss of awareness or consciousness, muscle spasms, and vision, hearing, or taste sensations [7].

The seizures are sudden and unpredictable, causing unpredictability in everyday life and limiting the patient's ability to participate in activities such as driving or swimming. More severely, the risk of premature death in people with epilepsy is up to three times higher than in the general population [8]. In addition to the physical effects, epilepsy can cause psychological conditions such as anxiety and depression [7].

Seizures occur when a group of brain cells discharges excessive electrical signals. The severity of the seizure depends on where the focus is and how far it spreads in the brain. Despite advances in treatment, epilepsy can still significantly impact an individual's quality of life, and there is no cure for the condition [7].

1.1 Motivation

One approach to improve the quality of life for people with epilepsy is by predicting seizures. Using only a few electrodes placed on the scalp, patients can be notified on a portable device, such as a smartphone, before a seizure occurs. Injuries related to seizures can be prevented in addition to a more predictable everyday life.

The first step in seizure prediction is seizure detection. Using brain wave monitoring, such as EEG signals and classification using machine learning is a well-established practice and has been done with high accuracy [9, 10, 11]. However, most research done in the field uses training and test data from the same patient [8]. This patient-dependent approach is not proven to give the same accuracy when testing on other patients. In order to use the results from previous research for something that has value for other patients, a patient-independent approach needs to be developed. In addition to high

accuracy, the patient-independent model also has to combine few features and few channels in order to be both fast and easy to use for the patients to increase the performance of real-time prediction.

1.2 Problem Description

This work aims to propose a patient-independent method for channel and feature reduction based on their importance for epileptic seizure classification using EEG signals.

1.3 Related Works

To give a rough overview of previous research relevant for this work the survey "Automated Epileptic Seizure Detection in Pediatric Subjects of CHB-MIT EEG Database—A Survey" [8] from October 2021, together with the paper "EEG Channel-Selection Method for Epileptic-Seizure Classification Based on Multi-Objective Optimisation" [11] from June 2020 is presented.

Both patient-dependent and patient-independent approaches have been used, where patient-dependent methods tend to perform better [8]. Different methods for feature extraction were used, including Discrete Wavelet Transform (DWT), Empirical Mode Decomposition (EMD), Empirical Wavelet Transform (EWT) and Multilevel Wavelet Decomposition (MWD). Support Vector Machine (SVM), k-nearest neighbours, random forest, gradient boosting, Bayes net, C4.5 and naive Bayes are examples of classification methods. In all papers, the CHB-MIT dataset was used.

Both patient-dependent and patient-independent approaches are investigated in [8]. Different features from time, frequency, and time-frequency domain, in addition to nonlinear features, were extracted and fed into different classifiers. Three papers from the survey are especially relevant to this work. The first one is [12]. 22 patients were used, the sample rate was 256 Hz, and MWD was used to do decomposition. SVM was used for classification. A sensitivity of 97.98% and a specificity of 89.90% were obtained. The second one is [10]. 24 patients, 256 Hz as sampling rate and DWT for decomposition was used. SVM with a radial basis function (RBF) kernel was used for classification. An average accuracy of 96.87%, a sensitivity of 72.99% and a specificity of 98.13% for the patient-dependent approach and an accuracy of 95.79% for the patient-independent approach was obtained. The third one is [9]. 24 patients with a patient-dependent approach were used. EWT was used for decomposition, and random forest, k-nearest neighbours, naive Bayes, Bayes net and the C4.5 algorithm was used for classification. This gave a sensitivity of 97.91%, a specificity of 99.57% and an accuracy of 99.41%. The paper that presents the best results for a patient-dependent approach is [11]. Accuracy up to 100% with an average over all patients of 97% $\pm$ 5% was obtained using only one channel and DWT for decomposition. When using EMD and 1 or 2 channels, an accuracy up to 98% was obtained. SVM, k-nearest neighbours, random forest, and naive Bayes were used for classification. All 24 patients were used.

1.4 Structure of the Report

Chapter 2 describes the relevant theory, from signal collection to feature extraction and classification using different machine learning methods. Previous research on epilepsy detection is also presented. In chapter 3, the datasets are presented, in addition to the method used and the performance measures to evaluate the results. The results are presented in chapter 4. A short summary, a discussion and some suggestions for further work are found in 5.

Chapter 2

Theory

2.1 Electroencephalography

EEG is a noninvasive tool that can measure electrical signals from the brain. Brain cells communicate via electrical impulses. These impulses can be measured by placing several electrodes on the scalp. EEG signals can be used to detect epileptic seizures, as the seizures are caused by a disruption in the brain signals. This disruption is visible in an EEG signal and is used to distinguish between seizure and seizure-free periods. To give an example, 10 seconds of the pure EEG signal with and without a seizure is shown in Figure 2.1. The signal is from CHB-MIT Scalp EEG Database [1], patient 13, recording number 19, channel F3. For the seizure-free signal, the amplitude ranges between -100 to 100, with some exceptions. The signal is different and the amplitude ranges from -300 to 200 for the signal with seizure.

2.2 The Discrete Wavelet Transform for Decomposition

DWT is a method of decomposition used to represent a signal in a more redundant form [13]. It gives information about both time and frequency. Feature extraction based on DWT coefficients makes the analysis less computationally heavy.

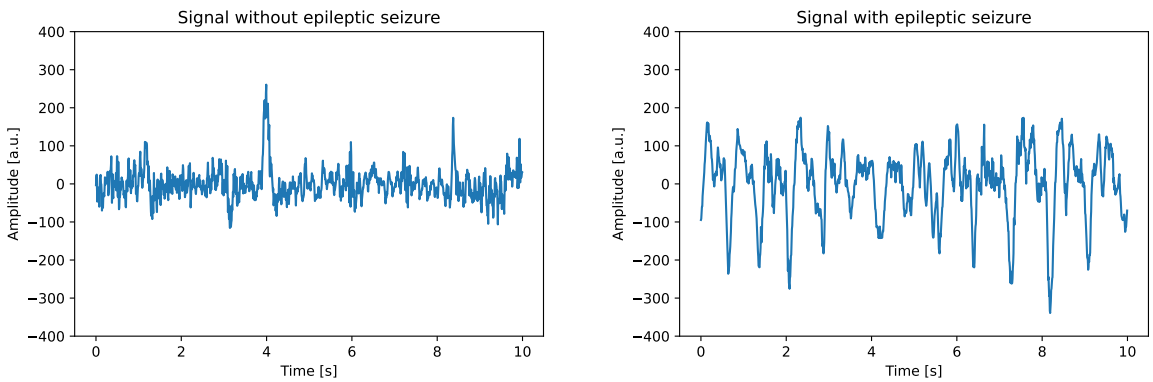


Figure 2.1: Pure EEG signals.

DWT approximates the signal using superposition of basis functions. The output from DWT is a set of coefficients describing the signal as a linear combination of these basis functions, called wavelets. They are formed by scaling and translating a mother wavelet. The mother wavelet must be chosen depending on the signal, either using trial and error or optimisation algorithms [11].

The wavelets are decomposed n times, where n is the level of decomposition. This results in detail coefficients ($D_1, ...D_N$), from high pass filtering and approximation coefficient (A_N), from low pass filtering. The low pass coefficients are again decomposed into low pass and high pass coefficients, and so on. It is these coefficients that are the output of the DWT.

2.3 Feature Extraction

Features from decomposition techniques such as DWT can be used to find similarities in the signal across different patients and recordings. Sixteen different features were used in this work, they are listed below. Most of these are well-known, the others that are most relevant are described below.

- Mean value
- Root mean square value
- Peak-to-peak amplitude
- Standard deviation
- Variance
- Skewness
- Kurtosis
- Line length
- Instantaneous energy
- Teager energy
- Hjort mobility
- Hjort complexity
- Selvik fractal dimension
- Higuchi fractal dimension
- Katz fractal dimension
- Petrosian fractal dimension

Petrosian Fractal Dimension

Petrosian Fractal Dimension is a fast way of estimating the fractal dimension by using the binary sequence of the signal. It was first proposed by Petrosian [14]. It can be expressed as:

$$FD_{Petrosian} = \frac{log_{10}n}{log_{10}n + log_{10}\left(\frac{n}{n+0.4N_\nabla}\right)} \tag{2.1}$$

n is the number of samples, and N_∇ is the number of sign changes in the binary sequence.

Katz Fractal Dimension

The Katz Fractal Dimension is derived directly from the waveform, not the binary sequence as Petrosian fractal dimention. Katz proposed a method for finding the fractal dimension by summing and averaging over the maximum distance d between the starting point and any other point and the euclidean distances between successive points [15]. It can be expressed as:

$$FD_{Katz} = \frac{log_{10}n}{log_{10}(\frac{d}{L}) + log_{10}n} \tag{2.2}$$

n is the number of samples, L is the length of the curve, and d is the diameter [16].

2.4 Classification Using Machine Learning

Machine learning algorithms can be trained on features extracted from EEG data from patients with epilepsy in order to identify patterns that indicate the presence of a seizure. This can replace the time-consuming and monotonous task earlier performed by a neurologist [8] as the machine learning can be done without human interaction.

2.4.1 Tree-based Methods

Tree-based machine learning methods are based on decision trees. These are tree-like structures used to make decisions using a test. The outcome of each test (tree) is weighed by a probability. The importance of the data points in a tree-based algorithm can be found by either computing the mean of accumulation of the impurity decrease within each tree or by finding the feature permutation [17]. The first one, Mean Decrease in Impurity (MDI), is calculated by summing over the number of splits that includes the data point across all trees, divided by the number of samples it splits. The latter one, Mean Decrease in Accuracy (MDA), is defined as the decrease in a model score when a single feature is shuffled randomly [18]. MDA helps break the relationships between the data points and their labels. The drop in the score indicates how much the model depends on the data point [17, 19]. MDA is the most popular of the two [20].

Random Forest

To create a Random Forest (RF) [18] classifier, a bootstrapped dataset is created. This is done by taking random samples out of the original dataset. Multiple decision trees are created using a random subset of data points from this bootstrapped dataset at each step. This results in a wide variety of trees and makes the random forest more efficient than single decision trees. The variety trees are then used to vote for which class a random sample from the dataset belongs to. The sample is assigned to the class with the most votes.

Gradient Boosting

To do classification using gradient boosting (GB) [21], the probability given that a sample belongs to one or the other class is calculated and compared to a threshold. The probability is calculated by finding the initial prediction for every sample. The prediction is converted into a probability by using a logistic function. Residuals are calculated by taking the difference between the observed and predicted values. A decision tree is built using the data points from the dataset to predict the residuals. The output values for the leaves in the decision trees are then calculated using the previous probability. Predictions are updated by combining the initial leaf with the tree and a learning rate. The same process is repeated for the following samples, but the residuals are now calculated as the difference between the observed data and the predicted from the leaf and previous trees. Running a new data sample down these trees, summing the probabilities will give a number that can be compared with the threshold to decide which class it belongs to.

2.4.2 Kernel based Methods

Support Vector Machines

The support vector machine (SVM) algorithm is a linear, kernel-based machine learning method based on calculating the decision boundary, called a hyperplane, in the feature space.

Given data in dimension N, calculating the hyperplane in the same dimension can be problematic as the data might not be linearly separable in N. To make this calculation possible, the SVM algorithm projects the data from a dimension N to a higher dimension, M. A hyperplane is found that separates the two classes linearly in dimension M. This projection is made using the kernel trick [22], which makes it possible to work with low dimensional data but with the advances of a higher dimension.

When deciding on a decision boundary, the trade-off between the largest possible margin between the two classes and the smallest possible number of errors is evaluated. The margin is the distance between the hyperplane and the classes. Errors are data points classified wrongly. The importance of each data point is given by the weight assigned to it. It is relative to the decision boundary.

Chapter 3

Materials and Method

3.1 Datasets

3.1.1 CHB-MIT Scalp EEG Database

The data used is from Children's Hospital Boston (CHB) and the Massachusetts Institute of Technology (MIT) [1], and was obtained from PhysioNet [3]. It contains long-term EEG records from 22 patients (ages 1.5 to 22), two of which has been recorded twice at different times, resulting in 24 instances. The instances from the same patients are treated as individual subjects, and thus 24 patients are referred to throughout this work. The patient's brain waves were recorded up to several days after withdrawal of anti-seizure medication.

The sample rate is 256 Hz. The files contain EEG signals from 22 to 38 electrodes, the number varying for each recording and patient. The extended international 10-20 EEG electrode positions and nomenclature system was used for these recordings, the placement is shown in Figure 3.1. Patient 04, 09, 12, 13, and 15 has channels that are not named after the 10-20 system, e.g. patient 13 with channels "LUE-RAE" and "EKG1-EKG2". These are disregarded. For some patients, the electrode in position $T8-P8$ is listed twice, e.g. for patient 12, it is both channel 21 and 28. The Euclidean distance between the signals on these channels is 0, so the duplicate (channel 28) was ignored. In some cases, up to five dummy signals (named "-" or ".") were interspersed among the EEG signals to obtain an easy-to-read display format. These dummy signals were ignored. All the other channels were used for this work.

3.1.2 Siena Scalp EEG Database

Another dataset from the Unit of Neurology and Neurophysiology of the University of Siena [2] was used to verify the results. It was obtained from PysioNet [4]. The dataset contains long-term EEG records from 14 patients, ages 20 to 71.

The sample rate is 512Hz. The files contain 23 to 28 EEG signals. The international 10-20 EEG electrode positions and nomenclature system was used. All patients also have one or two EKG signals. These were ignored.

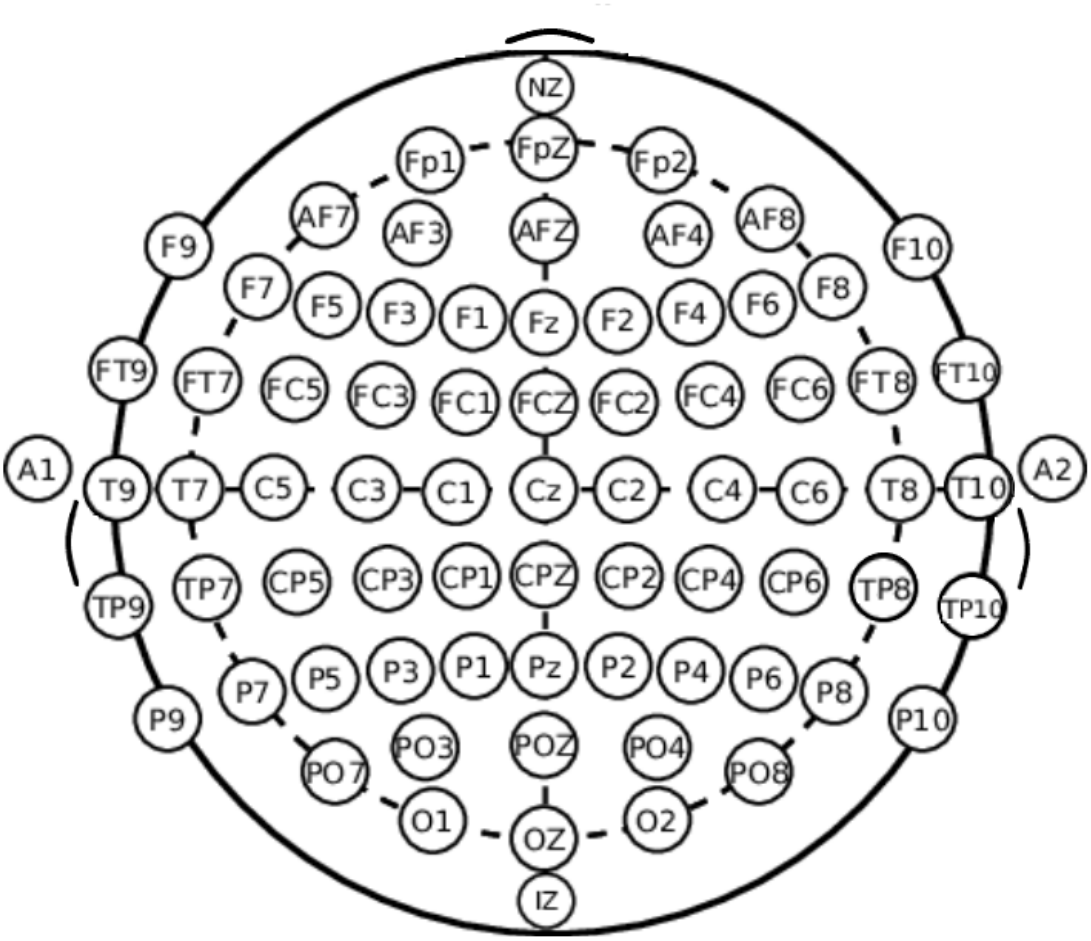


Figure 3.1: Placement of the electrodes in the extended international 10-20 system. The curve on the top marks the nose, and the curves on each side mark the ears. Adapted from [23].

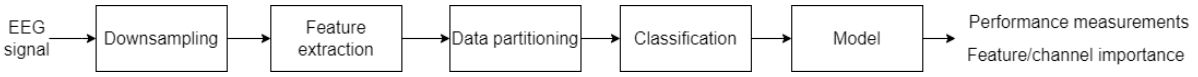


Figure 3.2: The steps resulting in classification of the signal.

The results presented in this work was obtained using the CHB-MIT dataset unless otherwise noted.

3.2 Feature Extraction and Classification

Figure 3.2 gives an overview of how the process from signal to classification was done. The process started with extracting the signal and downsampling it from 256 Hz to 128 Hz, as research has shown that this reduces computational time without affecting the results [24]. Feature extraction was done by decomposing the signal using DWT with four decomposition levels. For DWT, the mother wavelet was the bi-orthogonal 2.2. These choices were based on the results from previous research, [24, 25, 26]. Sixteen features were calculated based on the coefficients from the decomposition. The features were organised in a matrix. The matrix set-up differed based on whether feature or channel importance should be found.

The rows in the matrix was split into training and test set. 33% was used for testing and 67% for training. These were fed into three machine-learning algorithms: random forest, gradient boosting and SVM. The gradient boosting classifier used a learning rate of 0.75. SVM used a linear kernel. The importance of the features in the classification was found for each machine learning algorithm using MDA for the tree-based classifiers and weights for SVM. Using only the most important features, the importance of each channel was found using MDI for the tree-based classifiers, and weights for SVM.

Both a patient-dependent and a patient-independent method were used. The patient-dependent method was a starting point for the patient-independent method and was used to verify the results from the patient-independent method. Only performance measures are reported, and feature and channel importance are not. The patient-dependent approach was tested on both CHB-MIT and the Siena dataset. Different runs of the method gave different performance, the standard deviation of the accuracy was about ±3%. Feature importance and channel importance also changed. The results presented are after running the algorithm once.

3.2.1 Feature Importance

To determine the feature importance, a matrix with features as columns had to be created. For illustrative purposes, the architecture of the feature matrix X is shown for two patients, two recordings and two channels in Figure 3.3. To construct the matrix, the signal from each recording is split into intervals. If the recording contains one epileptic seizure, there will be three intervals: one seizure-free, which is the beginning of the recording (no recordings start with a seizure), one with the epileptic seizure and one without seizure which is the end of the recording. If there are two seizures in a recording, there will be five intervals, one without seizure, then one with seizure, one without, one

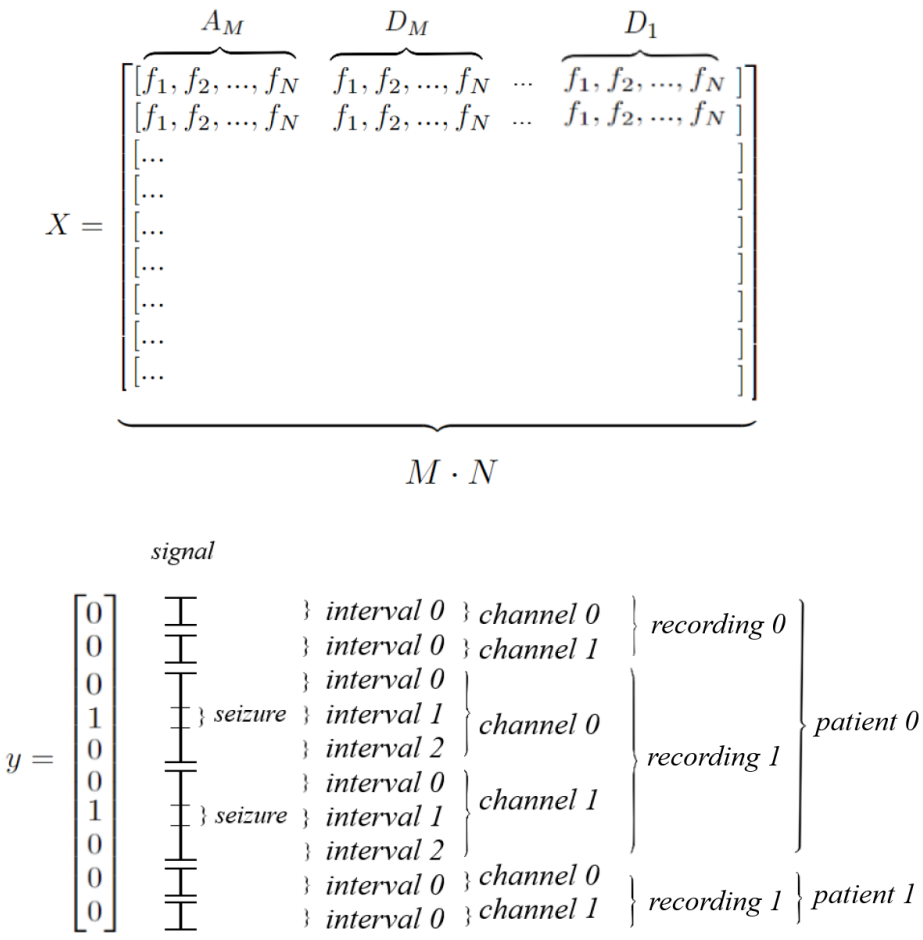


Figure 3.3: Setup for finding feature importance.

with seizure and one without, etc.

The rows in X are the number of features, N, from each decomposition level. There are M levels in total. These are extracted from each channel from each interval in each recording for each patient. These features are concatenated into one feature vector of length $N \cdot (M+1)$. The matrix height depends on the number of patients, how many recordings there are for each patient, how many seizures there are in each recording (resulting in several intervals), and how many channels are used for that recording.

The rows in the feature matrix were rearranged into random order to make sure the training and test set contained a random number of each class. The labels were rearranged in the same way to correspond to the correct feature row.

3.2.2 Channel Importance

To determine the channel importance, a matrix with channels as columns had to be created. The matrix used to find the most important channels has features as rows and channels as columns. Fig-

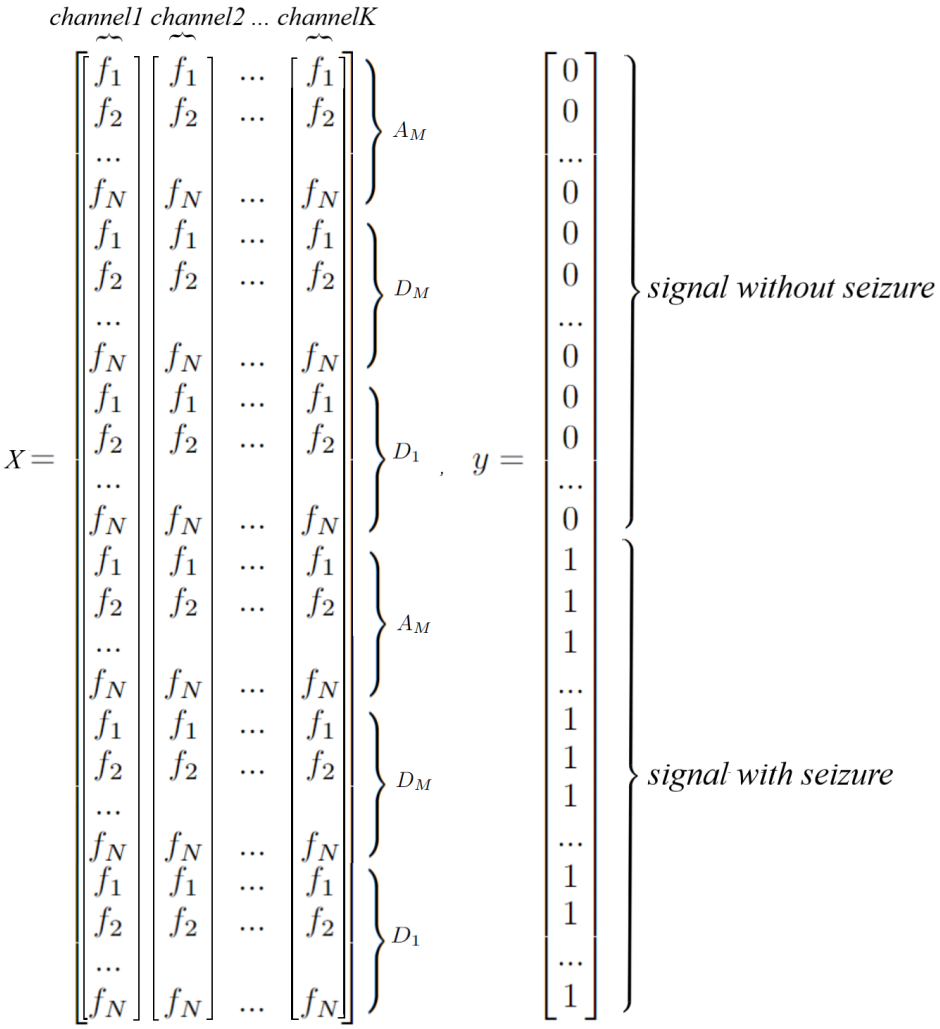


Figure 3.4: Setup for finding channel importance.

ure 3.4 shows how the feature matrix (X) and label vector (y) was set up. N is the number of features, M the number of decomposition levels and K the number of channels. To construct the feature matrix, the signal from each recording is split into intervals. The signal from intervals without seizure from the same channel (over multiple recordings and patients) was concatenated. So was the signal from intervals with seizures. The features for the seizure-free periods are calculated and added as columns. The features for seizure periods are calculated and added as an extension to existing columns. Corresponding labels are added to the label vector y. The rows in the feature matrix were rearranged into random order to make sure the training and test set contained a random number of each class. The labels were rearranged in the same way to correspond to the correct feature row.

3.3 Performance Measures

Different performance measures can be used to determine the classifier's performance.

Accuracy

Accuracy is an expression of how many correct predictions the algorithm made. It can be expressed as

$$Accuracy = \frac{Number\ of\ correct\ predictions}{Total\ number\ of\ predictions\ made} \tag{3.1}$$

The range is [0, 1], the closer to one, the better. However, accuracy does not always tell the whole truth about the results. If the dataset is unbalanced with a higher ratio of class 1 than class 0, then the accuracy will be high if all samples are labelled as belonging to class 1 [27].

Recall

Recall is the ratio between the number of true positive samples and all samples that should have been classified as positive, it can be expressed as:

$$Recall = \frac{True\ positives}{True\ positives + false\ negatives} \tag{3.2}$$

The range is [0, 1], the closer to one, the better.

Precision

Precision is the ratio between the number of true positives and the total number of positives predicted by the model. It can be expressed as:

$$Precision = \frac{True\ positives}{True\ positives + false\ positives} \tag{3.3}$$

The range is [0, 1], the closer to one, the better.

F1 score

The F1 score is based on precision and recall and tells how many samples are correctly classified, but also how robust it is, meaning that it does not misclassify samples. It can be expressed as:

$$F1 = 2 \cdot \frac{1}{\frac{1}{precision} + \frac{1}{recall}} \tag{3.4}$$

High precision and low recall give high accuracy but a low F1 score because it misclassifies samples which are hard to classify. The range is [0,1], the closer to one, the better.

Sensitivity

Sensitivity is a measurement of the true positive rate. It can be expressed as:

$$Sensitivity = \frac{True\ positives}{True\ positives + false\ negatives} \tag{3.5}$$

The range is [0,1]. The closer to 1, the better.

Specificity

Specificity is a measurement of the true negative rate. It can be expressed as:

$$Specificity = \frac{True\ negatives}{True\ negatives + false\ positives} \tag{3.6}$$

The range is [0,1]. The closer to 1, the better.

Confusion Matrix

A confusion matrix gives an overview of classifications and misclassification. For a binary classifier, it has four categories: true positives, true negatives, false positives and false negatives. Given a positive and a negative class, the true positives occur when the model correctly predicts that a sample belongs to the positive class. True negatives occur when the model correctly predicts that a sample belongs to the negative class. False positives are the samples incorrectly classified as belonging to the positive class. False negatives are the samples incorrectly classified as belonging to the negative class. The greater the value of true positives and true negatives and the lower the value of false negatives and false positives, the better the performance of the model.

Area under the ROC Curve

Area Under the ROC curve (AUROC) is an efficient algorithm based on the area under a Receiver Operating Characteristic curve (ROC). ROC is found by plotting the two parameters *true positive rate* and *false positive rate* against each other for different classification thresholds. The true positive rate is the same as recall, while the false positive rate is expressed as:

$$\textit{False positive rate} = \frac{\textit{False positives}}{\textit{False positives} + \textit{true negatives}} \tag{3.7}$$

AUROC gives a measurement of performance across all classification thresholds. The range is $[0,1]$, the closer to one, the better the performance. It is a helpful measurement if the dataset is unbalanced because the true positives and true negatives are calculated for each class separately.

Chapter 4

Results

4.1 Patient Dependent Method

The objective of this experiment was to find the performance of the methods using a patient-dependent approach. The performance measures after running the model for each patient are shown in Table A.1, Table A.2, Table A.3. The average accuracy was 99.9% for random forest, 99.6% for gradient boosting and 99.2% for SVM.

The methods presented below were used to see the effect on the performance and verify the high performance measures. The methods were run on different patients, but only patient 15 is used here as an example for better comparison.

4.1.1 Balancing the Dataset

When running the dataset for only one patient, the data can become unbalanced depending on the patient's number of seizures and recordings. This can result in misleading performance measures [27].
The dataset was often imbalanced towards more seizure-free periods than seizure periods. To prevent this, seizure periods were split into segments of 6 seconds, as done in [11]. Performance was not much affected by the imbalance, however. For patient 15 with a ratio of seizure to seizure-free instances of 85%:15% versus 22%:78%, the performance was similar, shown in Table 4.1. All channels were used.

	RF	GB	SVM
Accuracy	**1.0**	**1.0**	0.945
F1 score	1.0	1.0	0.871
Sensitivity	1.0	1.0	0.8
Specificity	1.0	1.0	0.964
AUROC	1.0	1.0	0.943

(a) 85%:15% seizure to seizure-free.

	RF	GB	SVM
Accuracy	**1.0**	**1.0**	0.833
F1 score	1.0	1.0	0.798
Sensitivity	1.0	1.0	0.938
Specificity	1.0	1.0	0.625
AUROC	1.0	1.0	0.781

(b) 22%:78% seizure to seizure-free.

Table 4.1: Performance measures for an imbalanced dataset using sixteen features and all channels. From patient 15.

	RF	GB	SVM
Accuracy	0.997	**1.0**	0.828
F1 score	0.994	1.0	0.694
Sensitivity	0.980	1.0	0.588
Specificity	1.0	1.0	0.867

Table 4.2: Performance measures using 97% of data for testing. Sixteen features, all channels. From patient 15.

Patient 6	RF	GB	SVM
Accuracy	**1.0**	**1.0**	**1.0**
F1 score	1.0	1.0	1.0
Sensitivity	1.0	1.0	1.0
Specificity	1.0	1.0	1.0

Patient 5	RF	GB	SVM
Accuracy	**1.0**	**1.0**	**1.0**
F1 score	1.0	1.0	1.0
Sensitivity	1.0	1.0	1.0
Specificity	1.0	1.0	1.0

Table 4.3: Performance on the Siena dataset.

4.1.2 Training and Testing

Reducing the number of samples in the training set reduced the performance, as expected. For patient 15 with 16 features and all channels and 97% of the data used for testing resulted in the performance shown in Table 4.2.

4.1.3 Results on the Siena Scalp EEG Dataset

The accuracy obtained with the CHB-MIT dataset was verified when running the same algorithm on the dataset from Siena. Running the algorithm with all channels, four features which proved good results in [11] (mean value, instantaneous energy, Teager energy and Petrosian fractal dimension) for patient 5 and 6, the results obtained are shown in Table 4.3.

4.2 Patient Independent Method

This experiment aims to find the most important features and use these to find the most important channels for all patients. The performance of the models was measured using all features and all channels, then using only the most important features and all channels. Lastly, only accuracy was measured using the most important features and an increasing number of channels. The channels used in the last scenario were ordered according to their importance. First, calculating the performance using the most important channel, then using the two most important, three most important and so on.

Different machine learning methods yield different results. Their performances are shown in Table 4.4, Figure 4.1a, Figure 4.1b and Figure 4.1c. Random forest was the one that performed best, SVM second best and gradient boosting worst, though all three perform well. The most important feature

	RF	GB	SVM
Accuracy	**0.990**	0.989	0.961
F1 score	0.984	0.981	0.936
Sensitivity	0.993	0.984	0.972
Specificity	0.977	0.964	0.912
AUROC	0.985	0.964	0.942

Table 4.4: Performance using all features and all channels for a patient-independent approach.

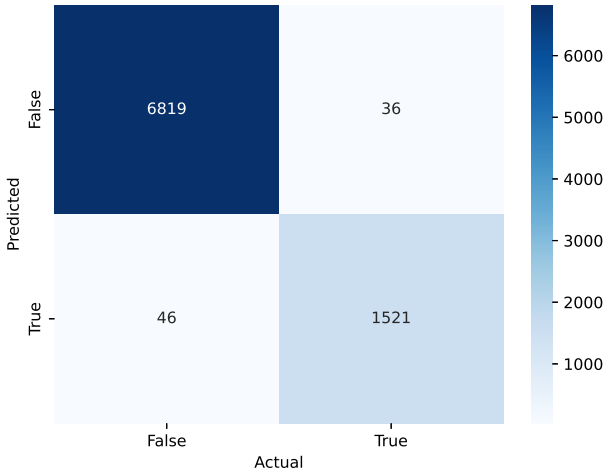
	RF	GB	SVM
Accuracy	**0.992**	0.902	0.986
F1 score	0.988	0.824	0.977
Sensitivity	0.994	0.962	0.99
Specificity	0.984	0.639	0.97
AUROC	0.989	0.8	0.98

Table 4.5: Performance using the most important features and all channels for a patient-independent approach.

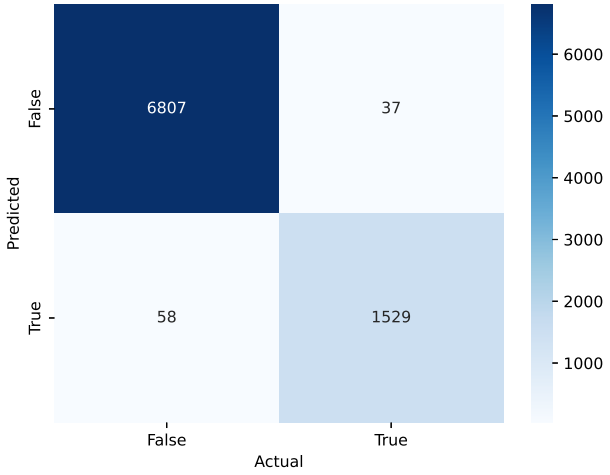
for the random forest was Petrosian fractal dimension, as shown in Figure 4.3. The most important feature for gradient boosting was also Petrosian fractal dimension, as shown in Figure 4.4. The most important features for SVM were standard deviation, root-mean-square and Katz fractal dimension, as shown in Figure 4.5.

Performance for random forest using only Petrosian fractal dimension and all channels is shown in Table 4.5 and Figure 4.2a. Performance for gradient boosting using only Petrosian fractal dimension and all channels is shown in Table 4.5 and Figure 4.2b. Performance for SVM using only standard deviation, root-mean-square, Katz fractal dimension and all channels is shown in Table 4.5 and Figure 4.2c.

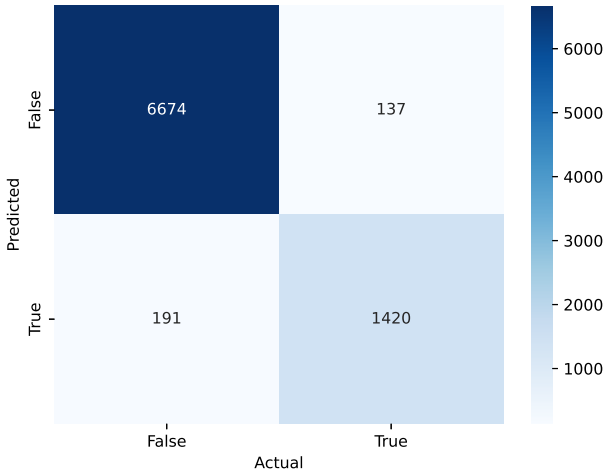
The channel importance using these features are shown in Figure 4.6. The accuracy for an increasing number of channels is plotted in Figure 4.9. The accuracy was high for all three classifiers, even when using a small number of channels. Random forest and SVM outperform gradient boosting when the number of channels was low.



(a) Random forest performance.

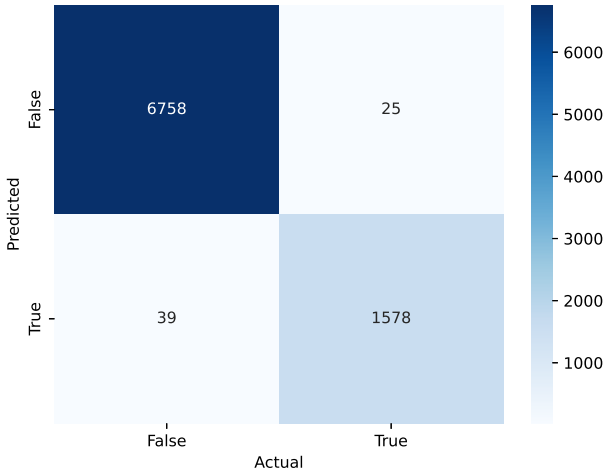


(b) Gradient boosting performance.

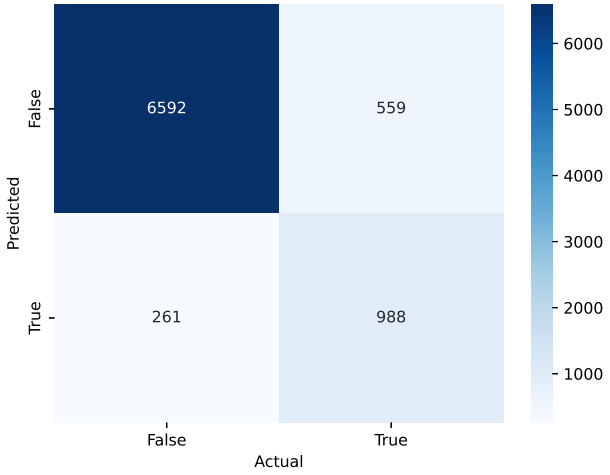


(c) SVM performance.

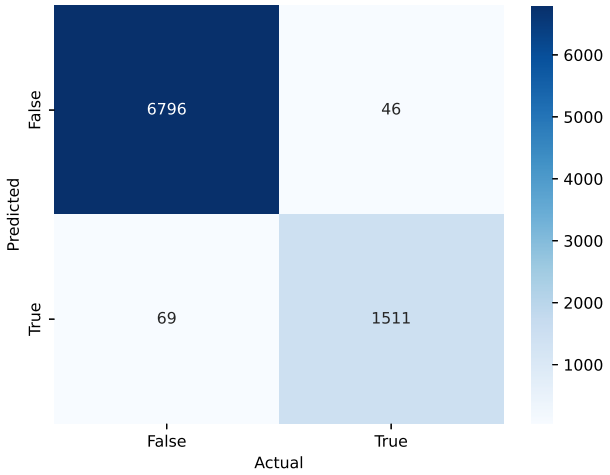
Figure 4.1: Confusion matrices using all features and all channels.



(a) Random forest performance.



(b) Gradient boosting performance.



(c) SVM performance.

Figure 4.2: Confusion matrices using the most important features and all channels.

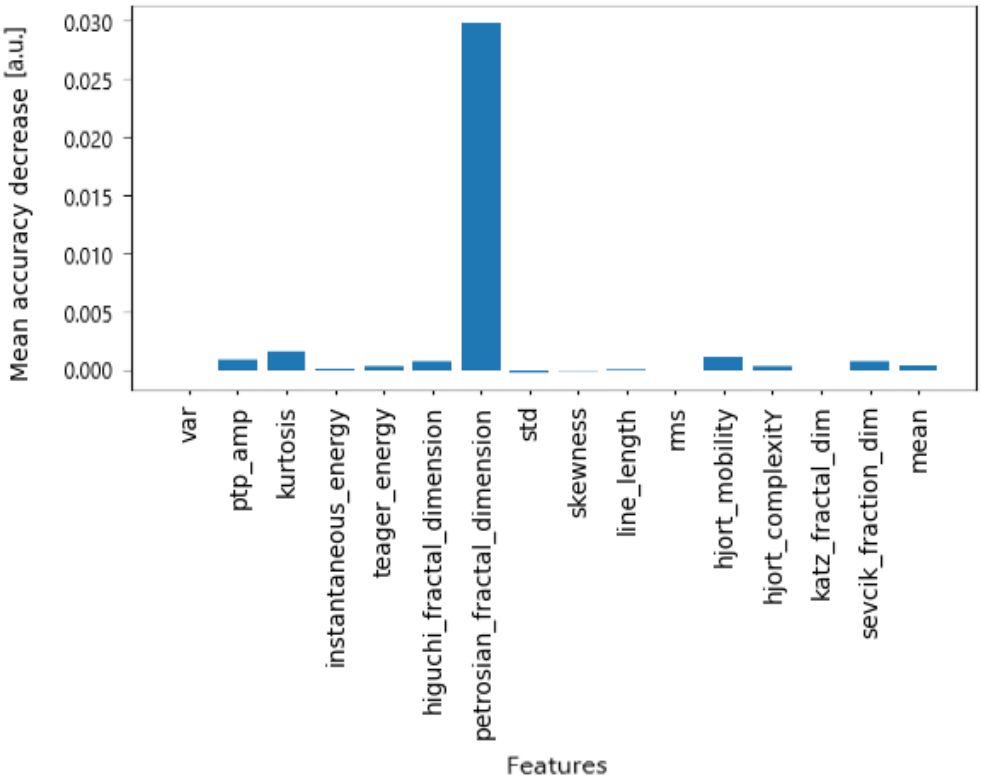


Figure 4.3: Feature importance for random forest using all features and all channels. Found using MDA.

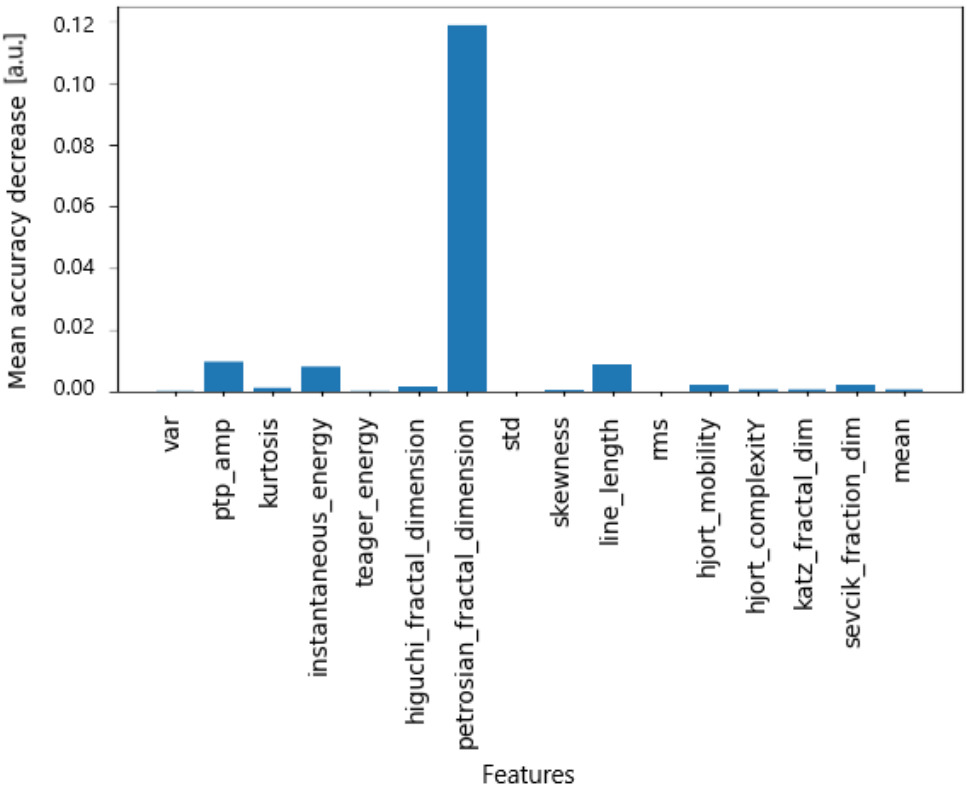


Figure 4.4: Feature importance for gradient boosting using all features and all channels. Found using MDA.

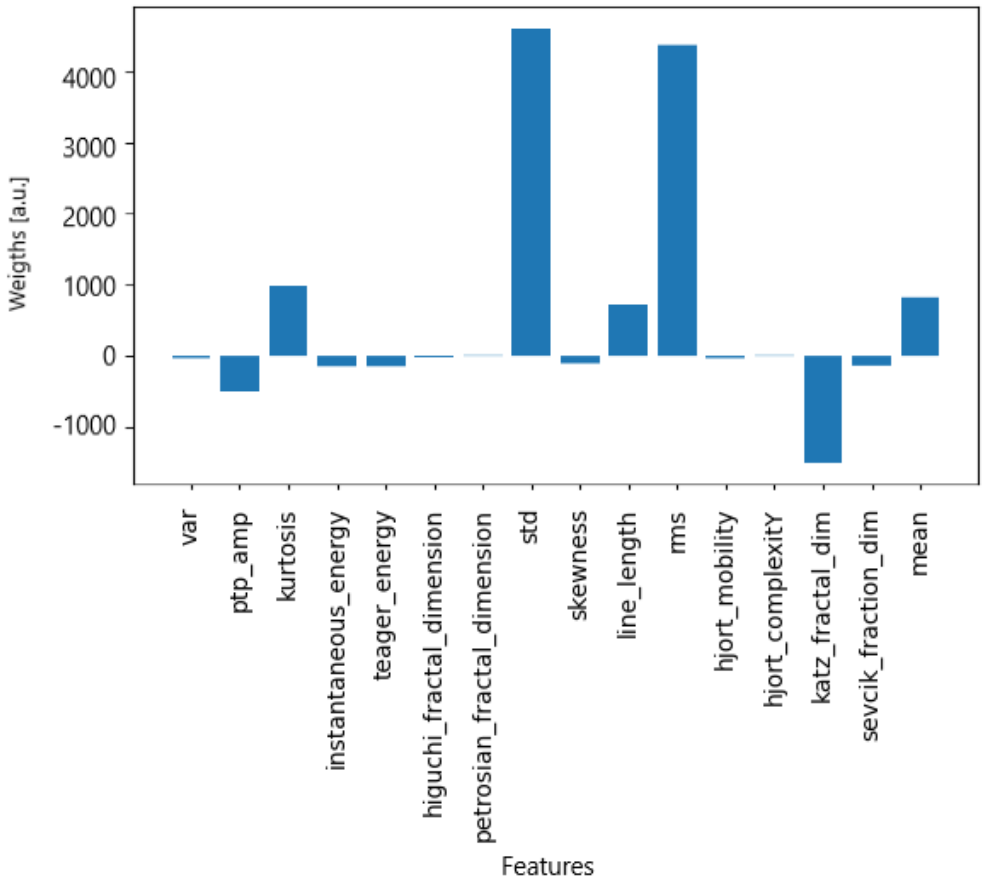
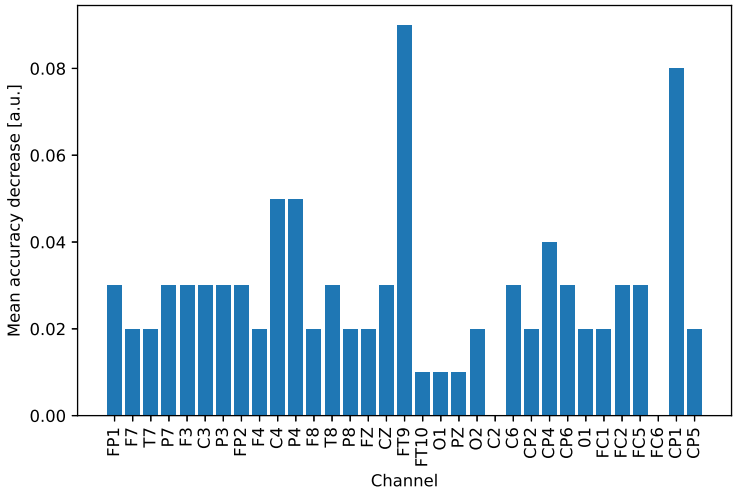
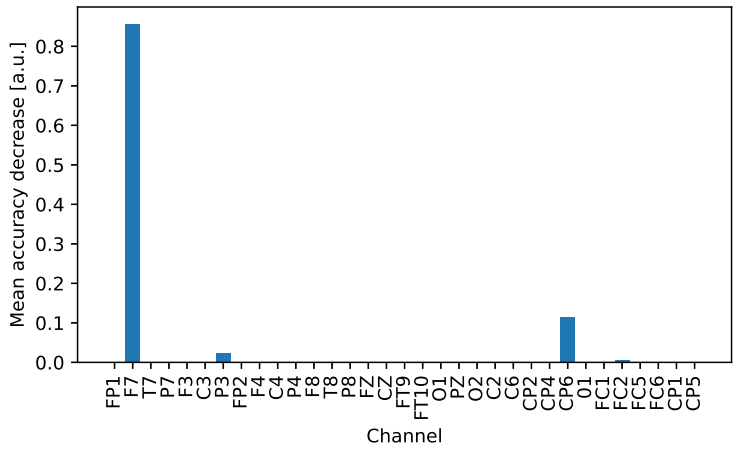


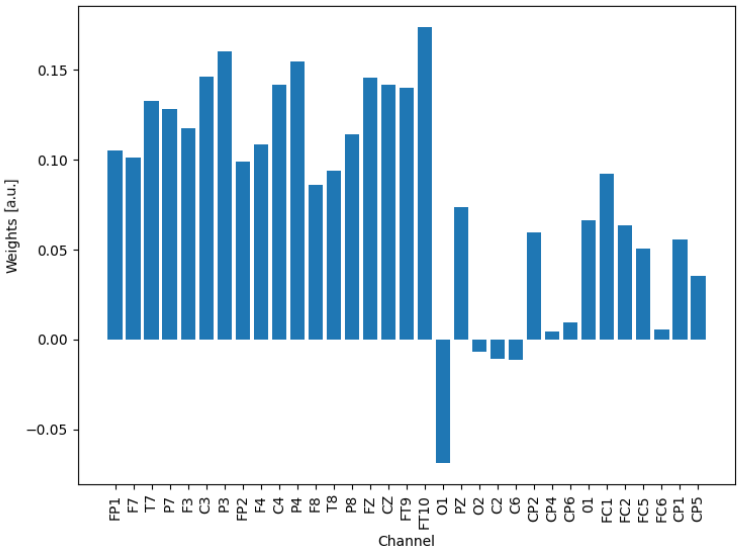
Figure 4.5: Feature importance for SVM using all features and all channels. Found using the weights.



(a) Channel importance for random forest using MDI.

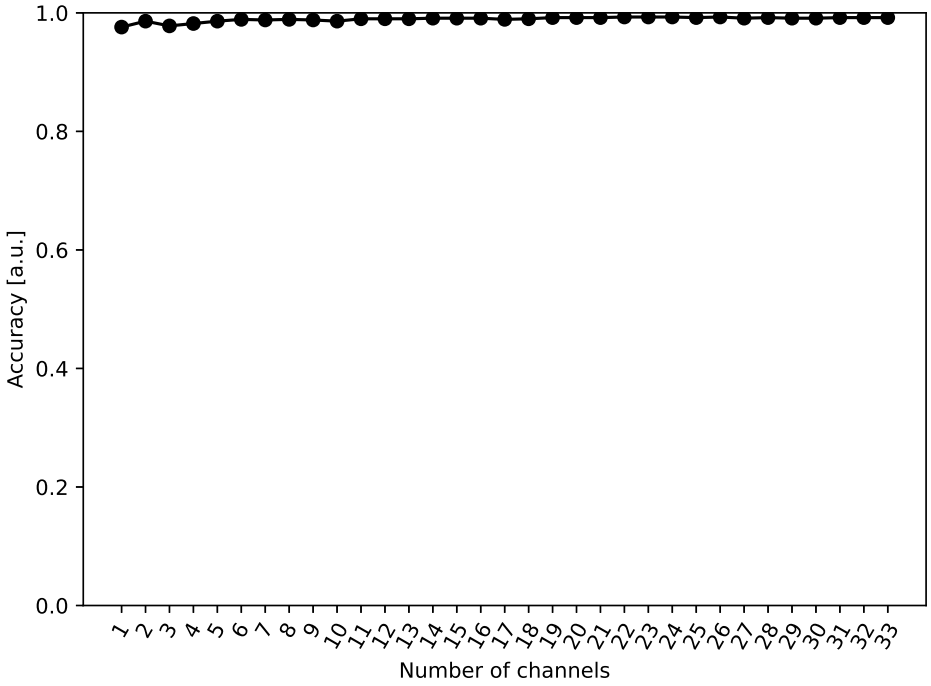


(b) Channel importance for gradient boosting using MDI.

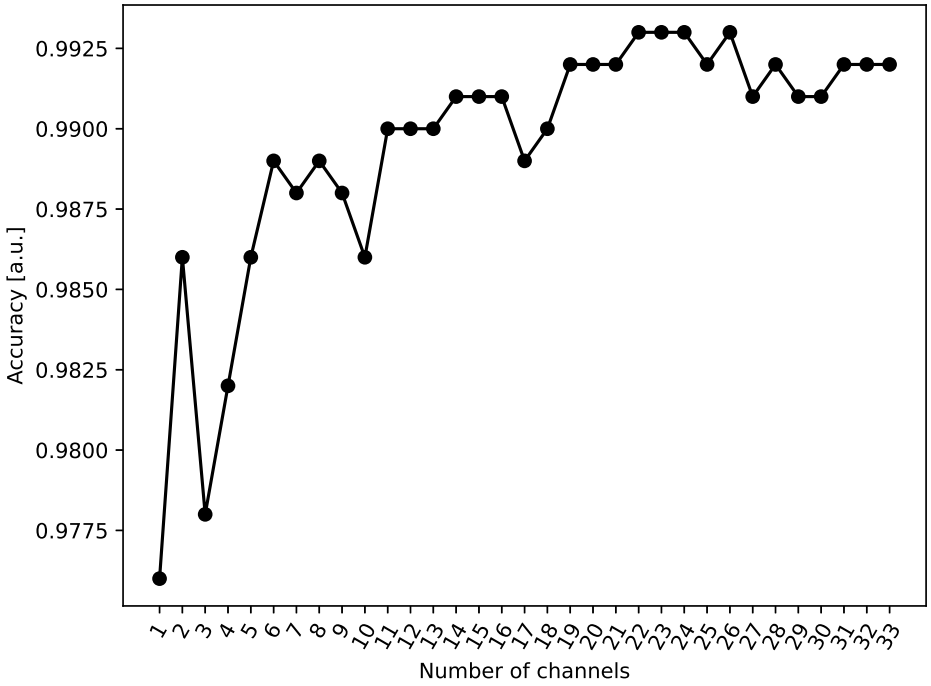


(c) Channel importance for SVM using weights.

Figure 4.6: Channel importance for the machine learning methods.

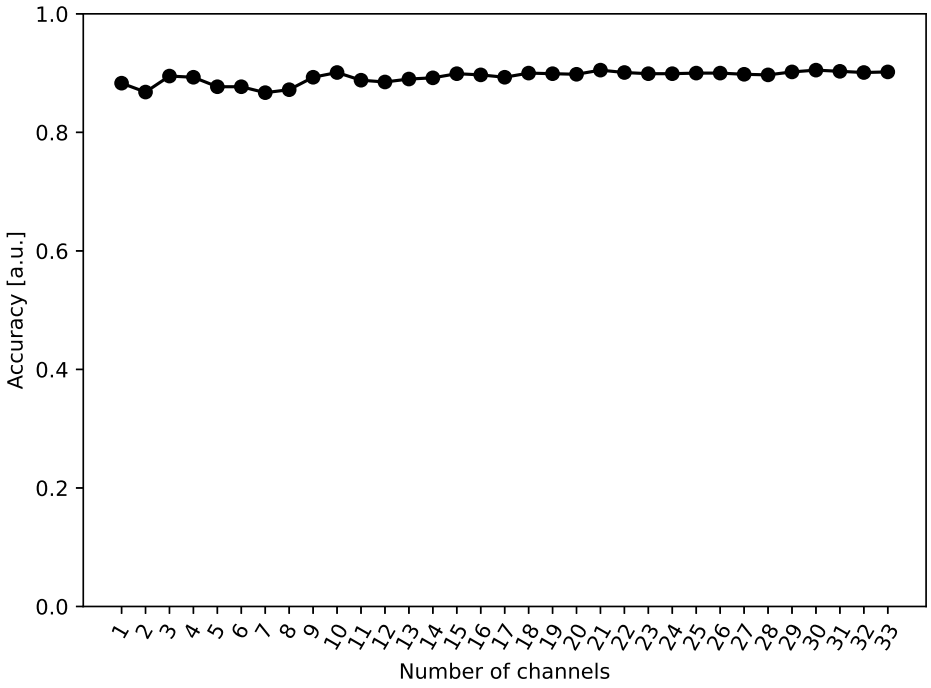


(a) Accuracy random forest.

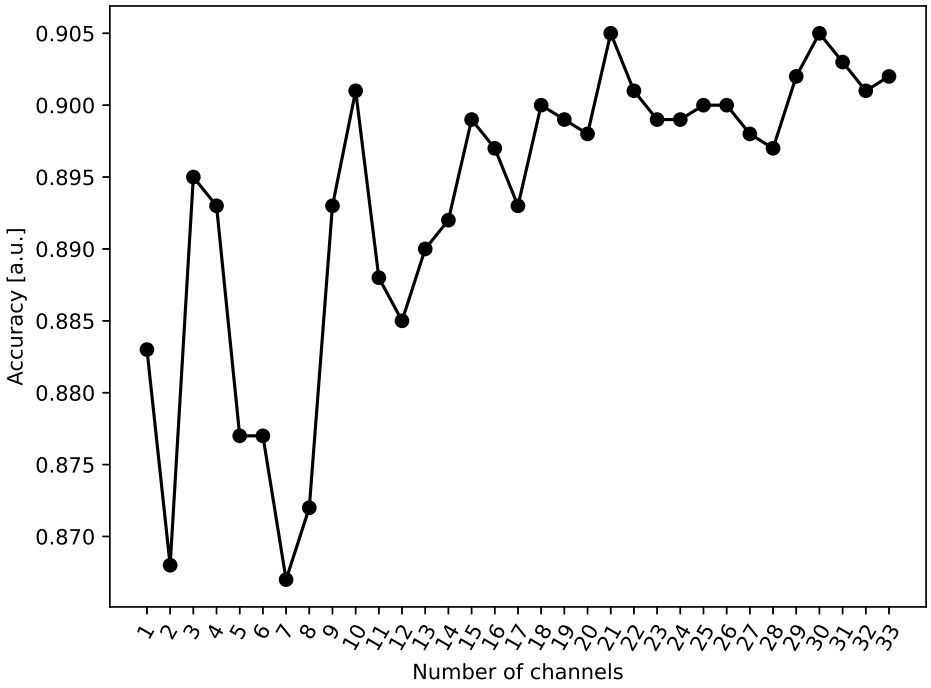


(b) Accuracy random forest, scaled.

Figure 4.7: Accuracy for random forest with an increasing number of channels.

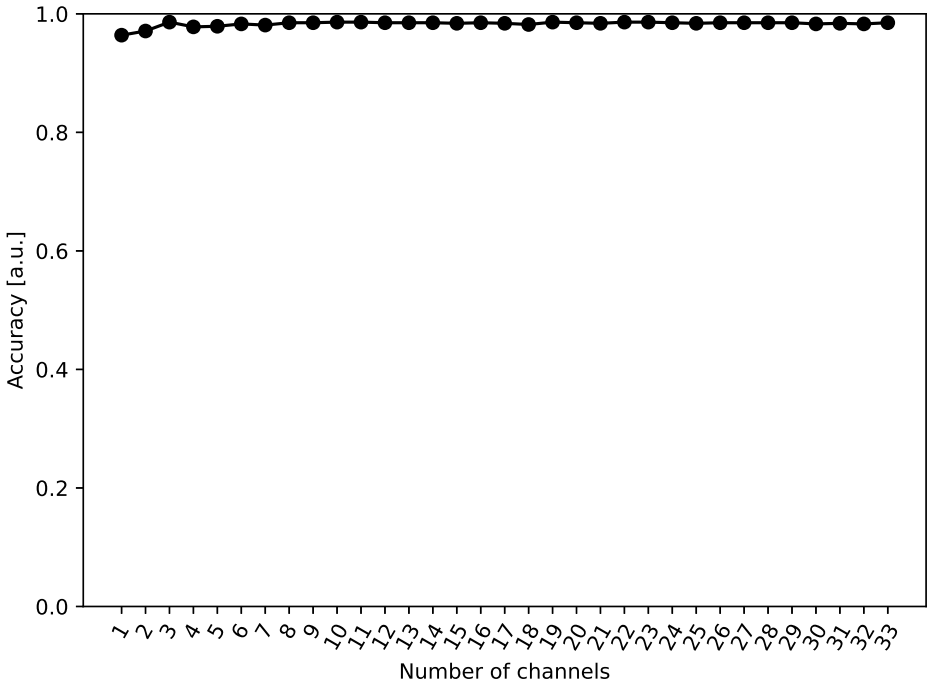


(a) Accuracy gradient boosting.

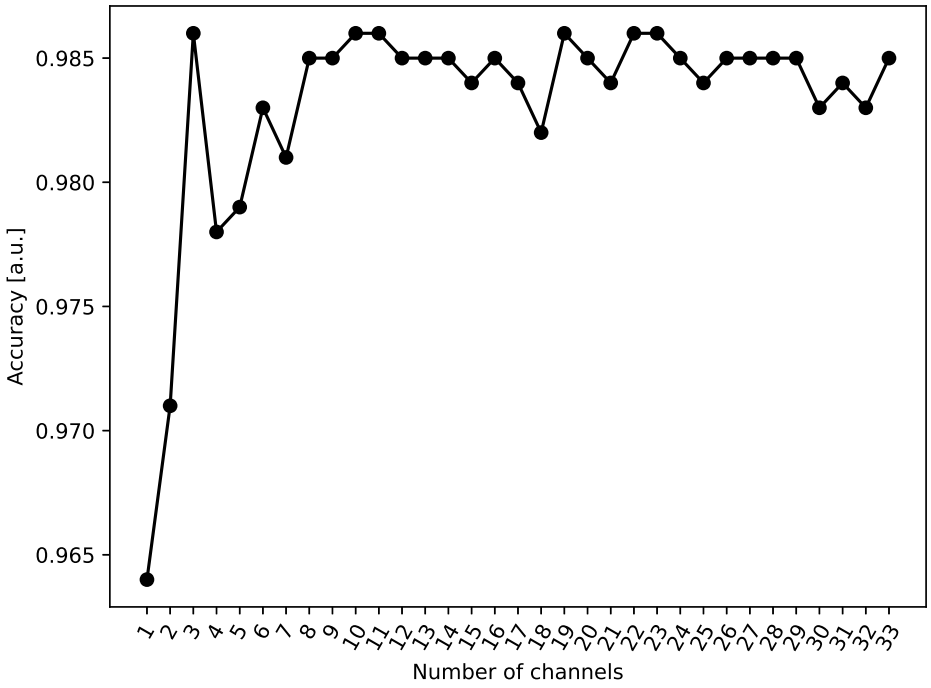


(b) Accuracy gradient boosting, scaled.

Figure 4.8: Accuracy for gradient boosting with an increasing number of channels.



(a) Accuracy SVM.



(b) Accuracy SVM, scaled.

Figure 4.9: Accuracy for SVM with an increasing number of channels.

Chapter 5

Summary, Discussion, and Further Work

5.1 Summary and Discussion

This work presents a patient-dependent and a patient-independent method for epileptic seizure detection. Three different machine-learning algorithms were used to classify seizure and seizure-free periods. Feature and channel importance were found for the patient-independent approach. The most important feature(s) was selected manually for each algorithm across all patients and all channels. Using this, channel importance was found. The method detects epileptic seizures with high performance for all three machine learning algorithms using these features and channels, especially the random forest classifier.

The patient-dependent approach was used as a starting point for the patient-independent approach and to verify the high performance. Verification included balancing and unbalancing the dataset, changing the size of the training set, using different performance measures, and running the algorithm on a different dataset. More patients from this dataset should be tested for an even better comparison. The two patients presented indicate some generalisability. Since the patient-dependent and patient-independent methods are based on the same architecture, the experiments conducted on the patient-dependent method also apply to the patient-independent method. Thus, the experiments prove the robustness of both approaches.

The random forest and SVM performance went up when using fewer features, as shown in Table 4.4 and Table 4.5. This can be because too many features might result in the method overfitting. Another explanation is that features that may cause noise in the classification are removed, as only the most important ones are chosen. The performance of the gradient boosting went down when using fewer features. This can be because too few features were used, and the classifier needed more information in order to classify the signal correctly.

Compared to the literature where SVM was the most popular choice for classification [28], random forest is the machine learning method that performs the best in this work. Even when using only one channel and one feature for the patient-independent method, the accuracy reaches 97.6% when using random forest, shown in Figure 4.7b, while only reaching 88.4% for gradient boosting, shown in Figure 4.8b, and 96.4% for SVM, shown in Figure 4.9b. When making the dataset very imbalanced,

either towards more seizures or towards more seizure-free periods, and when making the training set minimal (using only 3% for training), the performance was still high for the random forest and gradient boosting and slightly lower for SVM.

Some explanation of the high performance measures can come from visual inspection of Figure 2.1. The signal containing a seizure can easily be distinguished from the seizure-free signal. However, this does not explain the difference in performance between the classifiers.

Gradient boosting has a lower performance than the other machine learning methods. Specificity is significantly lower. As shown in Figure 4.2b, there is a high rate of false positives, which explains the specificity score. This can be due to noisy data [29]. Because the signal is not filtered, and the epileptic seizures are noisy, this might be an explanation for the lower performance. It could also be due to the value of the learning rate, and that changing this will give a better performance.

There can be multiple reasons why random forest performs better than SVM. The difference in performance is most visible for an imbalanced dataset, a small training set, and when a reduced number of channels is used. SVM is good at capturing the complexity of the data if it has a high dimension. However, if the dimension of the data is low, SVM might not be a good choice. High dimension data can be defined as data where the number of observations is equal to or smaller than the number of features, and low-dimension data as data where the number of observations is much larger than the number of features [30]. So the reason why SVM does not perform as well might be that the number of features (between 1 and 16 in this work) is lower than the number of observations, which is the number of intervals on all channels for all recordings of all patients.

Even though the performance of the method is high, some properties could have been explored, or explored in a more systematic manner. These are listed below.

- **Noise reduction**
  Noise reduction is a widely used prepossessing step to reduce the impact of noise, such as blinking and other movements from the patient or noise from the electrical grid. The latter is usually a significant problem when dealing with EEG signals [31]. This can, for example, be done the same way as presented in [32].

- **Learning rate**
  The learning rate for gradient boosting should be found using an optimisation algorithm such as grid search.

- **Kernel function**
  The kernel for the SVM classifier was set to a linear kernel. Experimenting with different kernels might give an even better performance.

- **Systematic feature selection**
  One drawback of the presented method is that the features are chosen manually. Doing this more systematically might lead to even better performance measures when using a few features and channels.

- **Other decomposition methods**
  Other methods for decomposition, such as EMD and MWD, are frequently used in epilepsy classification problems [8] and are therefore relevant to explore.

The main drawback of the method presented, and the reason that a clear conclusion is not given is that performance of the classifiers, as well as the feature importance and channel importance, changes for each run. This makes it difficult to conclude which feature and channel to use for each machine-learning method. The reason for using MDI instead of MDA when finding channel importance was because MDA gave an importance of zero for all channels. Non-zero values were obtained when using MDI. For gradient boosting, many of the channels still have an importance of zero. Some channels are irrelevant for the classification, as expected, but 30 channels out of 33 are worth further investigation. The reason for this, combined with why different runs yield different importance, is left for further work. To get a more generalisable conclusion, running the algorithm multiple times and summing the importance could be an option. This may give better results but does not solve the problem or explain why it happens. The results from this work are still useful, as the method presented results in high performance measures, even when using only one feature and one channel.

5.2 Further Work

There is still a way to go before a portable device for predicting epileptic seizures is a reality. The first step can be to predict the seizures, as detection will not have any value for the user.

Another step is to adjust the method to make real-time predictions, as a portable device will receive data in real-time.

An additional topic worth investigating is using other classifiers, such as Local Outlier Factor (LOF), to detect epilepsy in an early stage. Patients who have just been diagnosed will have little data to train the machine learning algorithms on. Then the epileptic seizures can be detected as outliers to regular brain activity.

Some work has already been done on the connection between sleep and epilepsy, but more research is needed [33, 34]. Different stages of sleep are relevant for our understanding and treatment of epilepsy, as the seizure occurs more often during sleep, and the type of seizure is different during sleep than when the patient is awake.

References

[1] Ali Hossam Shoeb. *Application of machine learning to epileptic seizure onset detection and treatment*. PhD thesis, Massachusetts Institute of Technology, 2009.

[2] Paolo Detti, Giampaolo Vatti, and Garazi Zabalo Manrique de Lara. Eeg synchronization analysis for seizure prediction: A study on data of noninvasive recordings. *Processes*, 8(7), 2020.

[3] Amaral L. Glass L. Hausdorff J. Ivanov P.C. Mark R. Mietus J.E. Moody G.B. Peng C.K. Goldberger, A. and H.E. Stanley. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation*, 101(23):e215–e220, 2000.

[4] Paolo Detti. Siena scalp eeg database (version 1.0.0). *PhysioNet*, 2020.

[5] Magnus Själander, Magnus Jahre, Gunnar Tufte, and Nico Reissmann. Epic: An energy-efficient, high-performance gpgpu computing research infrastructure. *arXiv preprint arXiv:1912.05848*, 2019.

[6] Ali Shahidi Zandi, Reza Tafreshi, Manouchehr Javidan, and Guy A. Dumont. Predicting epileptic seizures in scalp eeg based on a variational bayesian gaussian mixture model of zero-crossing intervals. *IEEE Transactions on Biomedical Engineering*, 60(5):1401–1413, 2013.

[7] WHO. Epilepsy. https://www.who.int/news-room/fact-sheets/detail/epilepsy, (n.d.). Last accessed 07.12.2022.

[8] J. Prasanna, M. S. P. Subathra, Mazin Abed Mohammed, Robertas Damaševičius, Nanjappan Jothiraj Sairamya, and S. Thomas George. Automated epileptic seizure detection in pediatric subjects of chb-mit eeg database—a survey. *Journal of Personalized Medicine*, 11(10):1028, Oct 2021.

[9] Abhijit Bhattacharyya and Ram Bilas Pachori. A multivariate approach for patient-specific eeg seizure detection using empirical wavelet transform. *IEEE Transactions on Biomedical Engineering*, 64(9):2003–2015, 2017.

[10] Suparerk Janjarasjitt. Epileptic seizure classifications of single-channel scalp eeg data using wavelet-based features and svm. *Medical & biological engineering & computing*, 55(10):1743–1761, 2017.

[11] Luis Alfredo Moctezuma and Marta Molinas. Eeg channel-selection method for epileptic-seizure classification based on multi-objective optimization. *Frontiers in neuroscience*, 14:593, 2020.

[12] Sylvia Bugeja, Lalit Garg, and Eliazar E Audu. A novel method of eeg data acquisition, feature extraction and feature space creation for early detection of epileptic seizures. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 837–840. IEEE, 2016.

[13] Amara Graps. An introduction to wavelets. *IEEE computational science and engineering*, 2(2):50–61, 1995.

[14] Arthur Petrosian. Kolmogorov complexity of finite sequences and recognition of different preictal eeg patterns. In *Proceedings eighth IEEE symposium on computer-based medical systems*, pages 212–217. IEEE, 1995.

[15] Michael J Katz. Fractals and the analysis of waveforms. *Computers in biology and medicine*, 18(3):145–156, 1988.

[16] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt. A comparison of waveform fractal dimension algorithms. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 48(2):177–183, 2001.

[17] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[18] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[19] Clément Bénard, Sébastien Da Veiga, and Erwan Scornet. Mda for random forests: inconsistency, and a practical solution via the sobol-mda. *arXiv preprint arXiv:2102.13347*, 2021.

[20] Clément Bénard, Sébastien Da Veiga, and Erwan Scornet. Mda for random forests: inconsistency, and a practical solution via the sobol-mda. *arXiv preprint arXiv:2102.13347*, 2021.

[21] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[22] Bernhard E Boser, Isabelle M Guyon, and Vladimir N Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, 1992.

[23] Shenghuan Zhang, Brendan McCane, Phoebe S-H Neo, Shabah M. Shadli, and Neil McNaughton. Trait depressivity prediction with eeg signals via lsboost. In *2020 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8, 2020.

[24] Luis Alfredo Moctezuma and Marta Molinas. Classification of low-density eeg for epileptic seizures by energy and fractal features based on emd. *Journal of biomedical research*, 34(3):180, 2020.

[25] Luis Alfredo Moctezuma Pascual et al. Distinción de estados de actividad e inactividad lingüística para interfaces cerebro computadora. Master's thesis, Benemérita Universidad Autónoma de Puebla, 2017.

[26] Luis Alfredo Moctezuma and Marta Molinas. Subject identification from low-density eeg-recordings of resting-states: A study of feature extraction and classification. In *Future of Information and Communication Conference*, pages 830–846. Springer, 2019.

[27] Fadi Thabtah, Suhel Hammoud, Firuz Kamalov, and Amanda Gonsalves. Data imbalance in classification: Experimental evaluation. *Information Sciences*, 513:429–441, 2020.

[28] F Lotte, L Bougrain, A Cichocki, M Clerc, M Congedo, A Rakotomamonjy, and F Yger. A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of Neural Engineering*, 15(3):031005, apr 2018.

[29] Alexandros Agapitos, Anthony Brabazon, and Michael O'Neill. Regularised gradient boosting for financial time-series modelling. *Computational Management Science*, 14(3):367–391, 2017.

[30] Jianping Hua, Waibhav D Tembe, and Edward R Dougherty. Performance of feature-selection methods in the classification of high-dimension data. *Pattern Recognition*, 42(3):409–424, 2009.

[31] Agata Nawrocka and Andrzej Kot. Methods for eeg signal analysis. In *2011 12th International Carpathian Control Conference (ICCC)*, pages 266–269. IEEE, 2011.

[32] Luis Alfredo Moctezuma. Towards universal eeg systems with minimum channel count based on machine learning and computational intelligence. 2021.

[33] Bo Jin, Thandar Aung, Yu Geng, and Shuang Wang. Epilepsy and its interaction with sleep and circadian rhythm. *Frontiers in Neurology*, 11:327, 2020.

[34] Håkon Stenwig, Andres Soler, Junya Furuki, Yoko Suzuki, Takashi Abe, and Marta Molinas. Automatic sleep stage classification with optimized selection of eeg channels. *bioRxiv*, 2022.

Appendix A

Results for all Patients Using a Patient-Dependent Approach

Patient 1	RF	GB	SVM
Accuracy	**1.0**	0.998	0.998
F1 score	1.0	0.995	0.995
Sensitivity	1.0	0.997	0.997
Specificity	1.0	1.0	1.0
AUROC	1.0	0.999	0.999
Patient 2			
Accuracy	**1.0**	0.997	0.997
F1 score	1.0	0.985	0.985
Sensitivity	1.0	1.0	1.0
Specificity	1.0	0.944	0.944
AUROC	1.0	0.972	0.972
Patient 3			
Accuracy	**1.0**	**1.0**	**1.0**
F1 score	1.0	1.0	1.0
Sensitivity	1.0	1.0	1.0
Specificity	1.0	1.0	1.0
AUROC	1.0	1.0	1.0
Patient 4			
Accuracy	**1.0**	**1.0**	**1.0**
F1 score	1.0	1.0	1.0
Sensitivity	1.0	1.0	1.0
Specificity	1.0	1.0	1.0
AUROC	1.0	1.0	1.0
Patient 5			
Accuracy	**1.0**	**1.0**	**1.0**
F1 score	1.0	1.0	1.0
Sensitivity	1.0	1.0	1.0
Specificity	1.0	1.0	1.0
AUROC	1.0	1.0	1.0

Patient 6	RF	GB	SVM
Accuracy	**1.0**	**1.0**	0.997
F1 score	1.0	1.0	0.996
Sensitivity	1.0	1.0	0.995
Specificity	1.0	1.0	1.0
AUROC	1.0	1.0	0.998
Patient 7			
Accuracy	**1.0**	0.997	0.997
F1 score	1.0	0.985	0.985
Sensitivity	1.0	1.0	1.0
Specificity	1.0	0.944	0.944
AUROC	1.0	0.972	0.972
Patient 8			
Accuracy	**1.0**	0.996	0.996
F1 score	1.0	0.992	0.992
Sensitivity	1.0	1.0	1.0
Specificity	1.0	0.975	0.975
AUROC	1.0	0.988	0.988
Patient 9			
Accuracy	**1.0**	0.995	0.995
F1 score	1.0	0.992	0.992
Sensitivity	1.0	0.994	1.0
Specificity	1.0	1.0	0.972
AUROC	1.0	0.997	0.986
Patient 10			
Accuracy	**1.0**	0.993	0.997
F1 score	1.0	0.987	0.994
Sensitivity	1.0	0.992	1.0
Specificity	1.0	1.0	0.978
AUROC	1.0	0.996	0.989

Table A.1: Performance measurements using all channels and all features for a patient dependent approach. Patient 1-10.

Patient 11	RF	GB	SVM
Accuracy	**1.0**	0.997	0.978
F1 score	1.0	0.987	0.913
Sensitivity	1.0	0.997	0.983
Specificity	1.0	1.0	0.9
AUROC	1.0	0.998	0.941
Patient 12			
Accuracy	**0.999**	0.988	0.997
F1 score	0.999	0.987	0.985
Sensitivity	1.0	0.984	1.0
Specificity	0.997	0.994	0.944
AUROC	0.998	0.989	0.972
Patient 13			
Accuracy	**1.0**	0.997	0.995
F1 score	1.0	0.996	0.993
Sensitivity	1.0	1.0	0.997
Specificity	1.0	0.989	0.989
AUROC	1.0	0.995	0.993
Patient 14			
Accuracy	**1.0**	**1.0**	0.997
F1 score	1.0	1.0	0.995
Sensitivity	1.0	1.0	1.0
Specificity	1.0	1.0	0.985
AUROC	1.0	1.0	0.993
Patient 15			
Accuracy	**0.992**	0.989	0.963
F1 score	0.989	0.986	0.952
Sensitivity	0.995	0.990	0.967
Specificity	0.981	0.986	0.953
AUROC	0.988	0.988	0.96

Patient 16	RF	GB	SVM
Accuracy	**1.0**	**1.0**	0.983
F1 score	1.0	1.0	0.977
Sensitivity	1.0	1.0	0.995
Specificity	1.0	1.0	0.945
AUROC	1.0	1.0	0.970
Patient 17			
Accuracy	**1.0**	0.99	0.99
F1 score	1.0	0.971	0.972
Sensitivity	1.0	1.0	0.995
Specificity	1.0	0.9	0.95
AUROC	1.0	0.95	0.972
Patient 18			
Accuracy	0.994	**0.997**	0.981
F1 score	0.986	0.993	0.951
Sensitivity	1.0	1.0	0.994
Specificity	0.952	0.976	0.881
AUROC	0.976	0.988	0.937
Patient 19			
Accuracy	**1.0**	**1.0**	**1.0**
F1 score	1.0	1.0	1.0
Sensitivity	1.0	1.0	1.0
Specificity	1.0	1.0	1.0
AUROC	1.0	1.0	1.0
Patient 20			
Accuracy	**1.0**	**1.0**	0.991
F1 score	1.0	1.0	0.987
Sensitivity	1.0	1.0	0.993
Specificity	1.0	1.0	0.986
AUROC	1.0	1.0	0.989

Table A.2: Performance measurements using all channels and all features for a patient dependent approach. Patient 11-20.

Patient 21	RF	GB	SVM
Accuracy	**1.0**	0.997	0.99
F1 score	1.0	0.992	0.976
Sensitivity	1.0	0.996	0.993
Specificity	1.0	1.0	0.971
AUROC	1.0	0.998	0.982
Patient 22			
Accuracy	**1.0**	**1.0**	0.996
F1 score	1.0	1.0	0.985
Sensitivity	1.0	1.0	1.0
Specificity	1.0	1.0	0.944
AUROC	1.0	1.0	0.972
Patient 23			
Accuracy	**1.0**	0.983	**1.0**
F1 score	1.0	0.979	1.0
Sensitivity	1.0	0.984	1.0
Specificity	1.0	0.98	1.0
AUROC	1.0	0.982	1.0
Patient 24			
Accuracy	0.99	**1.0**	0.959
F1 score	0.988	1.0	0.949
Sensitivity	0.997	1.0	0.966
Specificity	0.974	1.0	0.94
AUROC	0.985	1.0	0.953

Table A.3: Performance measurements using all channels and all features for a patient dependent approach. Patient 21-24.