Classification of epileptic seizures using accelerometers

Chalmers university of technology

Anders Hildeman hildeman@student.chalmers.se

Abstract

Epilepsy is a neurological disorder that affects a large group of people around the world. The research institute IMEGO has developed a hardware setup for measuring and logging acceleration of limbs on patients suffering from epilepsy. Signal analysis methods are applied to the acceleration data to register when a seizure occurs. In this thesis, the existing seizure classification methods are enhanced by studying the time evolution of the acceleration data over time spans of several minutes around each discrete sample point in time. This is achieved by clustering data and examining the cluster transitions between data points that are close in time.

In this work several clustering, classification and dimension reduction techniques are presented. Moreover, a pilot study that assesses the performance of the proposed classification method under a wide variety of parameters, input data and data transformations, is implemented. The old classification method that was used prior to this thesis is also evaluated and compared against the new method developed in this thesis.

The pilot study shows that the proposed classification method performs better than the method implemented prior to this work. The proposed method performs well on tonic-clonic epileptic seizures even with a sporty and active movement background during the measurements. For strictly tonic seizures, however, the pilot study gives no promising results. It seems like the classification methods are not able to distinguish a tonic seizure from normal activity.

The optimal parameters of the proposed method seems to differ depending on the kind of seizure and background data that is studied. In a real application, tuning the parameter model for the appropriate patient behavior seems to be necessary for adequate classification performance.

During the measurements, acceleration on three limbs were analyzed. The pilot study reveals that in many cases not all three acceleration points are necessary. Which ones are, depends on the patients seizure behavior.

Contents

Chapter 1

Introduction

In this chapter an introduction to the subjects, ideas and prior research of this master thesis is presented.

1.1 Background

IMEGO is a research institute that performs research and develop solutions in the fields of micro electronics, biotechnology and sensor systems. Since 2008, they have been coordinating a project aimed at recognizing epileptic seizures in patients wearing wireless accelerometers. IMEGO has developed both the electronics and software for this purpose. As part of this project and in collaboration with Chalmers Mathematical Center, the purpose of this thesis is to further develop the signal analysis algorithms as well as to conduct a pilot study that assesses the performance of such algorithm.

1.2 Epilepsy

Epilepsy is an umbrella term used for neurological disorders characterized by seizures. Since the seizures might originate in different locations of the brain, the symptoms can differ greatly between patients all diagnosed with epilepsy. The usual symptoms of a seizure might be: uncontrolled motoric movements or spasms, convulsions, emotional or psychological sensations, loss of consciousness and more. These are usually caused by signal feedback oscillations between neurons.

Epilepsy is the most common neurological disorder and it is estimated that around 50 million people all over the world suffer from recurrent epileptic seizures. About 1% of the Swedish population is affected. Moreover, about 10% of the worlds total population will some time in their life experience a grand epileptic seizure.

Some known causes of epilepsy are genetic abnormalities, strokes, brain tumors, brain defects, head trauma and infectious diseases. However, for about half of all the epileptic

seizures, their causes are still not completely understood. These are known as idiopathic seizures. [1] [2]

1.2.1 Seizures by origin

Epileptic seizures are divided into two major classes depending on the initiation. The partial-onset and generalized-onset seizures.

Partial seizures

Partial-onset seizures start in a specific area of the brain. The symptoms depend on the function of that specific area. If the seizure does not alter consciousness it is known as a *simple* partial seizure. Partial seizures that cloud consciousness and cause abnormal repetitious movements are known as *complex* partial seizures. Partial seizures are usually associated with some kind of damage to a part of the brain.

Generalized seizures

Generalized seizures are those seizures where the whole or large parts of the brain experience abnormal electrical activity. This kind of seizure is usually much more dramatic given that many of the brain functions are affected at the same time. In many of these types of seizures, the subject will not have any recollection of the seizure afterwards. Generalized seizures occur without evident damage to the brain but they might sometimes be initiated by a partial-onset seizure. If they are caused by a partial-onset seizure they are known as secondary generalized seizures.

1.2.2 Seizures by motoric symptoms

Seizures that affect motoric functions are usually classified by its symptoms. Often a seizure will exhibit more than one of these symptoms and the seizure can then be divided into symptomatic episodes.

Tonic seizures

A tonic episode of a seizure is the phase of constant muscle contraction. If a tonic seizure affects the heart (which is quite common in generalized seizures) the heart will stop beating due to the contraction of the heart muscle. Contractions are caused by oscillations of neuro-electrical signals with frequencies so high that the muscle fibers will not have time to relax in between each pulse. This causes the muscles to experience a constant contraction. Since the oscillations of the neuro-electrical signals during a seizure usually start at high frequencies that gradually decrease, the tonic episode is seen in the beginning of a seizure.

One typical position that in many cases is associated with tonic seizures is the so called ”fencing position”. Here, the upper body simulates a fencer with one arm stretched out to the side and the other one tucked in while the head is rotated to look in the direction of the stretched arm.

Atonic seizures

This is the opposite of tonic seizures. During an atonic phase the affected muscles relax instead of contract.

Clonic seizures

A clonic seizure, or phase of a seizure, is the phase where a part of the body is shaking.

Absences

These are sudden moments of absence with vacant stare and loss of attention. This kind of seizure does not usually occur after the age of 20.

1.2.3 Treatment

A large number of epileptic disorders can be controlled with medication. These pharmaceuticals usually decrease neural activity in the brain. Some types of epilepsy can be treated with brain surgery. There are also special diets, behavioral therapy and nerve stimulation techniques that inhibit certain kinds of epileptic seizures.

1.2.4 Diagnosis and evaluation

To be able to perform a reliable diagnosis and assessment of treatments the medical society needs constant quantitative and reliable data of patients experiencing seizures. The most reliable instrument today is the EEG analysis which measure electrical activity in the brain.

In an EEG analysis, electrodes are either attached to the scalp or surgically attached to the walls of the brain. These electrodes can sense when many neurons close to them fire at the same time. In other words, it is a macroscopic view of the neural activity at the brains surface. Neurons dedicated to similar tasks are located close to each other in the brain, i.e. neurons dedicated to muscular activity of a certain body part are located close to each other in the brain. This makes it possible to see when a certain physical, psychological or sensory function is very active by studying the electrical activity in that part of the brain. The areas in the brain that are close to the surface are generally associated with muscular activity. This makes seizures that affects muscular activity much more suitable for EEG

diagnosis. By EEG it is often possible to distinguish between a true epileptic seizure as compared to non seizure activity or a seizure caused by psychological effects.

However, the electrodes placement on the scalp makes EEG analysis hard to implement in every day life situations. Therefore, currently the most diagnosis and medical evaluation is based on a manual log where patients fill in and describe their seizures. A problem with this system is that patients are not always aware of their seizures given that these are sometimes associated with memory loss or clouded consciousness. The patient might also mistake another sensation for a seizure, or the seizure might occur in a sleeping or semi-awake state of mind. Furthermore, the description of the seizures are qualitative and highly subjective in nature.

Because of this, there is a strong and justified demand for a measuring device that can be used by the patients around the clock in normal situations and that give quantitative and reliable information about seizures.

For further reading see [1] [2].

1.3 Prior work

Some prior work has been done on the subject of classifying seizures from background data using accelerometer data and the current hardware setup (see sec.2.1). Johan Stigwall at IMEGO has developed a signal processing framework for extracting and inspecting accelerometer data. He has also done some analysis of the measurements taken from the first patients observed at Sahlgrenska (see section 2.2). Also, a prior master thesis has been written by Johan Wipenmyr [3] 2010 where he developed 51 summary statistics to use for discrimination (see section 2.3). Those statistics has been adopted in this thesis and act as the feature space of data for the proposed method (see chapter 4) to work on. The prior work has treated each sample in feature space as an independent sample. For patients with very strong seizures this method has been adequate but for the more subtle seizures it has not yielded satisfactory results.

Below follow a brief introduction to the classification method as it were prior to the initiation of this thesis. For more information about the prior work see [4].

1.3.1 Prior classifier

The classifier used prior to this thesis (For a more complete explanation see [3]) are in general terms defined as:

1. Extract feature values from the raw accelerometer data.
2. Cluster all observations in the training set marked as seizures using a k-means cluster analysis with three clusters (see section 3.1.1).
3. Fit a Gaussian distribution to the samples of each of these three clusters as well as a Gaussian for all non seizure samples in the training set.

4. Use quadratic discriminant analysis to classify samples in the test set as belonging to one of the estimated Gaussian. Weight with the "prior" probability of each Gaussian estimated from the training set.
5. Classify the samples of the test set as being a seizure if it is classified as a cluster which belongs to a seizure.
6. Post process by removing all samples in the test set that are classified as belonging to a seizure but do not have large number of samples in their neighborhood that are also classified as seizures.
7. Fill in the "holes" in between those remaining samples that are too close to each other and classified as seizures. In this way samples close to each other are registered as part of the same seizure instead of many small seizures.

1.4 Purpose

For this thesis, the aim from the start has been to make use of the time evolution of seizure acceleration data. The idea are that the movements characterizing a seizure will cause acceleration of the limbs in a certain order. Looking solely at a specific time instance to draw a conclusion on whether a seizure is occurring or not might not yield great classification. Instead, analyzing the acceleration at a certain time instance in the context of several time instances both before and after will give much more information to base a classification decision on. The prior classifier did only infer time evolutionary acceleration behavior from a very near future (the range of one second). This thesis does connect accelerations at time instances as far away from each other as several minutes when making a classification decision.

The main purposes of this thesis are:

- Include time evolutionary information into the decision process of the seizure classifier.
- Enhance the performance of the current classifier method.
- Perform a pilot study, assessing the performance of the proposed classification method.
- Investigate possibilities of decreasing the number of sensors and decreasing the computational complexity of the classifier.

Chapter 2

Data acquisition

In this chapter, information regarding the nature of the input to the classifier are presented. This includes the movements recorded (sec. 2.2), the recording equipment (sec. 2.1) and the data transformation (sec. 2.3).

2.1 Hardware

The task of gathering accelerometer data from a human being in every day life demands some special equipment. IMEGO has developed hardware for this purpose. The hardware consists of the following components:

2.1.1 Accelerometer sensor

A small box containing the actual accelerometers in 3 dimensions as well as some other needed electronics, this box is named *IDM20*. Such a sensor box contains:

- one three-dimensional accelerometer
- A small microcontroller for controlling analog to digital conversion, sampling and packeting of data.
- A radio module for sending the packeted data to a logging unit.

These boxes should be attached to the limbs one wants to measure the acceleration on.

2.1.2 Logging unit

This is the unit that receives packeted accelerometer data from all the sensors, sorts it out and stores it on a media. For some of the measures, this unit has been an ordinary laptop computer equipped with an USB radio module. With data being stored on the

Figure 2.1: Equipment used for measuring accelerometer data of epileptic seizures. Three accelerometer sensors, wrist bands for fixation of sensors to wrists and a portable logger unit for reception and storage of accelerometer data.

pc's hard drive. Also, a portable logging unit based on a Primer 2 development kit has been developed and used for some measurements. [5] The portable Primer 2 based unit being better suited for real life measurements since it is no larger than a cellular phone and lightweight with a battery time of at least one day. Here the data are stored on a regular microSD card.

2.1.3 Fixation

For the wrist sensors, special wrist bands similar to those used on arm watches have been developed. The chest sensors has so far been attached by medical tape. In the final product a smart attachment solution using textiles will probably be used. For women specially designed bras might be another good option. Overall, ideas involving designing clothes to accommodate the sensors are being investigated.

2.1.4 Radio communication

The data is transferred between the sensor boxes and the logging unit using a radio protocol called "ANT" from Dynastream Innovations Inc. This radio protocol is designed for low power, short distance and relatively low baud rate communication. It operates in the 2.4 GHz frequency band, which is the ISM band open for unlicensed use in the most countries worldwide. It has an operational range of about 10 meters (personal experience) and is considered a PAN or "personal area network". This means that the logging unit must be close to the patient at all times of data gathering, preferably carried by the patient or located in the same room (if a relatively small room). The baud rate of "ANT"

is at most 19 kb/s if using a certain mode of communication known as burst[6]. However this mode has its disadvantages, with the communication mode currently used, known as broadcast, the maximum baud rate is about 12.8 kb/s. This puts the epilepsy application close to the maximum communication speed of the protocol.

2.1.5 Sampling procedure

The microcontroller in the sensor box sample acceleration in three dimensions about 50 times per second. This accelerometer data as well as battery voltage and sensor temperature are then packeted and sent to the logging unit at appropriate intervals. The logging unit receives data from several sensor boxes at the same time. It merges the data and stores it in to a single data file. This data file is in later stages used for signal processing and classification.

2.2 Data

For evaluating the method, data have been gathered in two ways:

2.2.1 Sahlgrenska data

At Sahlgrenska university hospital, patients are held for observation and EEG analysis for a couple of consecutive days with the purpose of properly diagnosing their state and evaluating treatment. Under the supervision of a neurologist, IMEGO has had the opportunity to carry out accelerometer measurements on epilepsy patients which have given their consent. Patients are supervised by nurses around the clock and the EEG-data is analyzed by neurologists after the observations. In such an controlled environment, it is possible to record exactly when a patient is having a seizure. The advantage of this procedure is that the seizure accelerometer data is recorded and then confirmed by experts and patients themselves. The disadvantage is that the measurements are done in an unnatural setting. Because the patients movements are restricted by the length of the cords for the electrodes connected to their scalps (the most they are able to do is eat, sleep, watch tv, etc) day-to-day activities such as walking, cleaning, performing manual labor, etc, are never recorded into these data sets. Furthermore, certain procedures performed on a patient after a seizure might interfere with the actual seizure behavior or become discriminatory themselves. For example, the nurse needs to confirm the seizure and awakening from the seizure by communicating and touching the patient each time. The classifier might accidentally be trained to recognize the movements associated with such procedures. If a classifier is trained for this behavior it will not be able to find the seizures in real life when no nurse is there to perform the procedure during each seizure. Finally, the fact that the patient is constrained to a bed might also remove seizures that occur at standing posture. This could be problematic if the classifier is trained on a patient who lies on a bed but is used on the same patient in every day life.

2.2.2 Simulated seizures

This data are gathered by performing measures on individuals who are not experiencing true epileptic seizures. However, the individual performs a certain choreographed movement pattern (which is believed to resemble an epileptic seizure) at specific times. The advantage of this way of measuring is that the patient can be measured during normal life activities as well as during some more active tasks. Consequently, there is no risk at all of missing a seizure reference point because of uncertainties in EEG data analysis or other problems that might occur in the data recorded on real patients. The disadvantage is though, that the movement pattern might not resemble a true epileptic seizure or that there might be small differences in each ”seizure” movements due to the fact that these ”seizures” are actually cognitively controlled actions.

2.3 Features

In the prior master thesis [3] a number of summary statistics were derived from the accelerometer data. These statistics (henceforth referred to as ”features”) serve the purpose of presenting the accelerometer data in ways believed to hold greater discriminatory power between seizures and non-seizures, compared to the raw acceleration values. In this thesis, such features serve as the data space on which the proposed method performs the classification.

The features are calculated each second and are comprised of accelerometer values gathered at 50 Hz from one or several sensors and dimensions during a time span of one or four seconds (depending on feature). The wider four second time intervals are used for the ”DC” values (see sec. 2.3.1). Since the features are extracted each second the wider time interval will make the DC components from neighboring time instances dependent. In most cases the features from different time instances might be regarded as independent samples but one should be aware of this DC dependence of neighboring time instances.

For each second the information is gathered in a neighborhood using a hanning window. The reason for this is to avoid ”edge effects” and reduce frequency aliasing. See figure (2.2) for further understanding.

2.3.1 DC-values

The DC-values are calculated for each accelerometer, i.e. each sensor and dimension separately. It is a measure of the average acceleration during the time interval defined by the window function (A four second wide window). Since most of the movements have a shorter time span than 4 seconds, these measures will mostly capture the gravitational acceleration which makes them good indicators of orientation for each of the sensors.

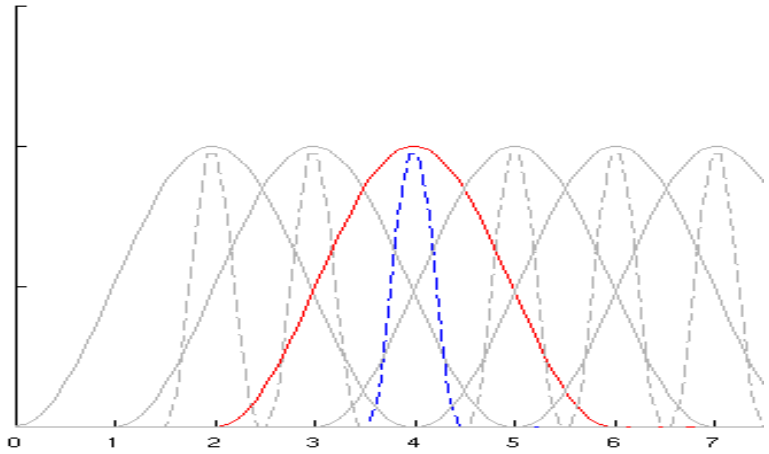

Figure 2.2: Feature extraction windows for neighboring time instances. The overlap of the wide windows (solid lines) as well as the mutual exclusiveness of the thinner windows (dashed lines) are clearly seen.

2.3.2 Signal magnitude area

Signal magnitude area (SMA) is a measure of the overall activity of a sensor during the windowing interval [3][7]. It includes acceleration in all directions to distinguish those time instances where there is a lot of movement, i.e.

$$\text{SMA} = \frac{\sum_i \left(\frac{|x_i|+|y_i|+|z_i|}{3} \cdot w_i \right)}{\sum_i w_i}, \tag{2.1}$$

where w_i is the weight for time instance i. and defined by the windowing function used. x,y and z are directions in the three dimensional acceleration space. SMA features are calculated for each sensor separately. Therefore the SMA measure will have larger values for acceleration in several basis directions compared to acceleration in only one dimension.

2.3.3 Vector magnitude

Vector magnitude (VM) is also a measure of the overall activity of a sensor. It is basically the same as SMA but uses the vector magnitude instead of the sum of the absolute values [3], i.e.

$$\text{VM} = \frac{\sum_i \left(\sqrt{x_i^2 + y_i^2 + z_i^2} \cdot w_i \right)}{\sum_i w_i}. \tag{2.2}$$

Due to the presence of gravitational acceleration which will affect a still accelerometer, acceleration in directions perpendicular to g will be underestimated by the VM value.

It is not obvious which measure will discriminate a seizure from a non-seizure best, since placement of the basis directions in a smart way might cause SMA to be more or less

discriminatory than the slightly less biased measure of VM. Therefore both are included even though it is believed that one of them can be excluded later on. VM features are calculated for each sensor separately.

2.3.4 Mean absolute magnitude difference

Mean absolute magnitude difference is the third of the overall activity measures. It is defined as

$$\text{MAMD} = \frac{\sum_i \left(\left| \sqrt{x_i^2 + y_i^2 + z_i^2} - 1 \right| \cdot w_i \right)}{\sum_i w_i}. \tag{2.3}$$

Mean absolute magnitude difference works like vector magnitude but subtracts the gravitational acceleration from the measure [3]. Here the acceleration are normalized to 1G, i.e. an accelerometer in a still environment on earth will register only the gravitational acceleration of magnitude 1. Therefore the MAMD measurement yield a value of 0 when the accelerometer are still (contrary to VM which will register a value of 1). MAMD features are calculated for each sensor separately.

2.3.5 Periodicity

The periodicity values (PER) measure how periodic the movements are around a time instance. These features are based on the FFT transform (fast fourier transform) and compare the magnitude of the largest frequency component to the average frequency magnitude, i.e. when one particular frequency is prevalent it is an indication of a periodic movement.

$$\text{PER} = \sum_{s=x,y,z} \frac{\max_{\omega \in \Omega}(|F_s(\omega)|)}{\text{mean}_{\omega \in \Omega}(|F_s(\omega)|)}, \tag{2.4}$$

where F_s are the fourier transformation of the acceleration in the s direction during a time window ω.

Periodicity are calculated for each sensor separately.

2.3.6 Frequency bands

The frequency band features are also based on the FFT transform and are calculated for some mutually exclusive frequency intervals. The idea here is to identify movements of a certain frequency that might be typical of a seizure.

The spectrum is divided among the frequency bands 0.75-2.25, 2.25-3.75, 3.75-5.25, 5.25-8.25, 8.25-13.25 and 13.25-25 Hz

Since the accelerometer data is gathered at 50 Hz the Nyquist frequency is 25 Hz (Shannon sampling theorem: No frequency component larger than half the sampling frequency can be identified [8]) and a frequency higher than that can not be extracted from the data. In the case where data have been gathered using radio transmission, there is a possibility of

data loss. In such cases the missing samples are estimated by interpolation. This might make the highest frequency band less reliable. Frequency band features are calculated as a sum for each sensor module.

2.3.7 Correlation

While all the other features have been calculated separately for each sensor. The correlation features measure the correlation between different sensors. There are one definition for the linear correlation (see eq.2.5) and one for the circular correlation (see eq.2.6).

The correlation measures how similar two sensors move compared to each other. If both the left and right arm perform the same movement this will give a high response in the correlation features that compares the arm sensors with each other.

The linear correlation,

$$\text{lincorr}(f,g) = \sum_{k=\kappa} \sum_{i=\{x_f,y_f,z_f\}} \sum_{j=\{x_g,y_g,z_g\}} \frac{\mathbf{a_i}[k]\mathbf{a_j}[k]}{|\mathbf{a_i}[k]|\,|\mathbf{a_j}[k]|}, \tag{2.5}$$

measures similarity if two sensors move in phase, i.e. the acceleration is more or less identical on both sensors such as linear spasms. The circular correlation,

$$\text{circorr}(f,g) = \sum_{k=\kappa} \sum_{i=\{x_f,y_f,z_f\}} \sum_{j=\{x_g,y_g,z_g\}} \frac{\mathbf{a_i}[k]\alpha_\mathbf{j}[k]}{|\mathbf{a_i}[k]|\,|\alpha_\mathbf{j}[k]|}, \tag{2.6}$$

measures the similarity if two sensors move with a 90° phase shift such as in a circular movement. Here f and g are sensors, $\mathbf{a_i}[k]$ are the acceleration at the $\mathbf{i}$ direction and sensor at time instance k and $\alpha_\mathbf{i}$ are the same as $\mathbf{a_i}$ but with a 90° phase shift. κ are the set of sample points inside the window of the current time instance.

Chapter 3

Theory

In this chapter, general methods and concepts used in this thesis are explained. The actual implementation of the concepts introduced in this chapter are later presented in chapter 4.

3.1 Clustering

Clustering or cluster analysis is the task of dividing a data set into subsets where members of a subset share some common characteristics. Clustering is a form of unsupervised learning and the subsets ("clusters") retrieved are usually not known a priori. Compared to the related field of classification, where the subsets are known in advance and the task is to find discriminatory properties, clustering is used more as an exploratory analysis tool.

Examples of usage

- In high dimensional data (higher than 3 dimensions) it is difficult to visualize the data distribution. Cluster analysis is a way of extracting information from a data set that is not possible to evaluate by eye.
- Cluster analysis can be used as a form of dimension reduction when the number of clusters are fewer than the dimension of the data. The membership values are then the new dimensions. This application of cluster analysis is probably more useful when using a "fuzzy" cluster algorithm, i.e. the membership values are continuous instead of discrete or dichotomous.
- Taxonomy, using cluster analysis as a tool for defining different classes of data.

Overview

A cluster algorithm strives to minimize a cost function, using some kind of similarity measure. Choice of algorithm and cost function depends on the application. The subject of cluster analysis and measures are vast and in this section only theory relevant for the thesis and overall understanding will be presented.

There are basically three types of cluster algorithms[9]:

- The *combinatorial algorithms* cluster data without any assumption of an underlying probability model.
- The *mixture modelling* algorithms assume that the data are distributed according to a parameterized probability model consisting of component density functions, where a component density function can be any parameterized probability density function. These algorithms try to fit the parameters of the presumed probability model to some training data. This is usually done by maximum likelihood or corresponding Bayesian methods or estimates thereof.
- The *mode seeking* algorithms also assume an underlying probability model but use a non parametric approach.

A human being is very well adept at recognizing clusters in 1-,2- or 3-dimensional data. Take a look at the data set in figure 3.1. Intuitively a human will find 2 separate clusters in this image. A smaller circular shaped cluster on the left and a larger more vertically outspread cluster on the right. However to accomplish this clustering with computer algorithms might not be so obvious. First of all, the algorithm can not presume a certain shape, since the two clusters have different shapes. Secondly, the algorithm can not presume a specific number of clusters since it could have been a different number of clusters in the figure but a human would still recognize them.

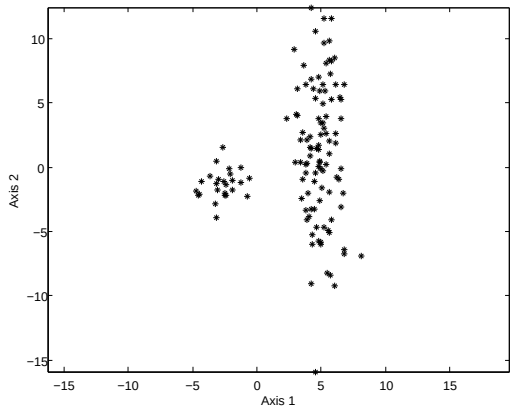

Figure 3.1: A 2 dimensional data set generated by two different Gaussian.

Data are usually vast, therefore, clustering algorithms need to work with huge amounts of data and the computational complexity will rapidly increase. It is usually important to use some heuristics to implement a computational simple algorithm that in general would

not have worked so well, but because of information known a priori about the problem might be possible to tune for better performance.

3.1.1 K-means

This is one of the simplest algorithms and a standard method for clustering. It is a combinatorial and iterative algorithm with a fixed number of clusters specified in advance. Each cluster is characterized only by a centroid. Sample points are members of only one cluster which is the cluster with the closest centroid to their location.

The k-means algorithm is generally defined as follows[9][10]:

1. Perform some normalization on the data such that all dimensions of the data have the same scaling. Usually the data is normalized with the z-score.

$$\hat{X}_i = \frac{X_i - \mu}{\sigma}, \tag{3.1}$$

   where X is the original observation and $\hat{X}_i$ is the z-score for the same observation.

2. Choose initial values for the k centroids. The choice of initial points might be random, based on prior knowledge or both.

$$C_i = \varpi_i, i = 1,...,k, \tag{3.2}$$

   where ϖ_i is a, possibly random, process for choosing the starting location of the i:th cluster.

3. Assign all observations to their respective closest centroid.

$$m_j = \underset{i}{\operatorname{argmin}} \|X_j - C_i\|_{l^2}^2, j = 1,...,N, \tag{3.3}$$

   where N is the number of observations.

4. Calculate the mean of all observations belonging to each cluster. Assign the centroids the new values of these means.

$$C_i = \frac{1}{J_i} \sum_{i=1}^{k} X_j \omega\{m_j = i\}, \tag{3.4}$$

   where J_i is the number of observations that are currently members of the i:th cluster, k are the number of clusters, ω is 1 if the statement inside the curly brackets are true and 0 if false, m_j is the membership function of the j:th observation (assuming a value in between 1 and k).

5. Steps 3 and 4 are iterated until m_j values do not change anymore.

As can be seen from the algorithm this clustering method might not actually find the global minimum and can be sensitive to the initial values Φ_i. However, it is a relatively

fast algorithm that does not assume any probability distribution. Therefore it is widely popular and usually regarded as the standard method.

It uses the Euclidean distance measure from the cluster centers. This creates linear borders that are not adapting to a special shape of a cluster. This can clearly be seen in figure (3.2) which is a k-means clustering of the data set from figure (3.1).

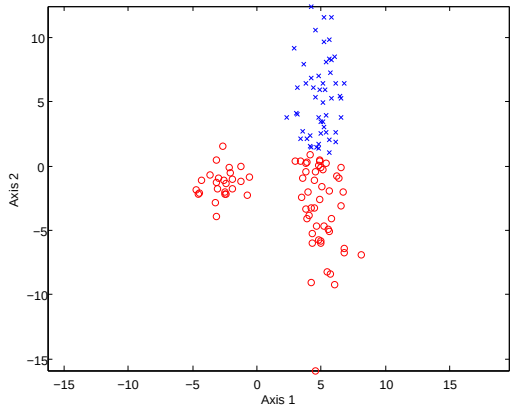

Figure 3.2: K-means clustering with $k = 2$ on the data set of fig.3.1.

3.1.2 Gaussian mixture

Gaussian mixture clustering is an example of a *mixture modelling* method. As the name suggests it assumes an underlying probability model of k Gaussian distributions.

$$f_{gm}(x) = \sum_{i=1}^{k} \pi_i f_i(x) \tag{3.5}$$

where f_i is the probability density function of a Gaussian distribution and f_{gm} is the probability density function of the Gaussian mixture distribution. π_i is the mixture coefficient. The mixture coefficients are all positive and sum up to unity.

Just as the k-means algorithm the general implementation of Gaussian mixture clustering has the drawback of performing a fit of exactly k clusters even though the data might be better suited for another value of k. However, the shape of the clusters are adapted to Gaussian shapes, which in many applications is a satisfactory shape adaption.

A Gaussian mixture clustering of the data in data set figure (3.1) can be seen in figure (3.3). The figure portray a perfect fit. This is a little misleading since the original data set was actually derived from two Gaussian distributions, the figure does nonetheless show a great advantage of Gaussian mixture clustering over k-means clustering.

Maximum likelihood estimate

Usually the fit of Gaussian distributions to the data given is performed by maximum likelihood estimation. The maximum likelihood estimate ($\hat{\theta}_{MLE}$) of a parameter (θ) is

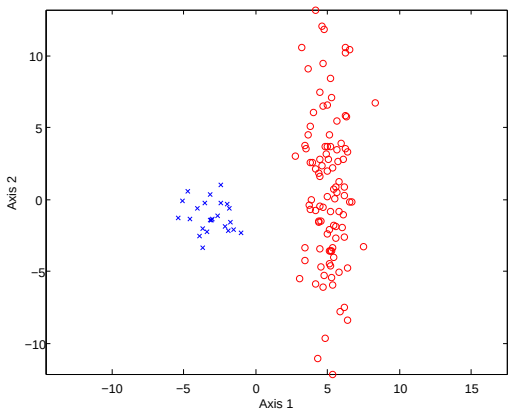

Figure 3.3: Gaussian mixture clustering with $k = 2$ on the data set of fig.3.1.

defined as the parameter value for which the likelihood function

$$L(\theta, \mathbf{x}) = \prod_{i=1}^{N} f(x_i|\theta) \tag{3.6}$$

of the given data set ($\mathbf{x}$) reaches its maximum. Where $f(x|\theta)$ is the density or mass probability function of a distribution with the presumed probability model and θ is the parameter vector.

The likelihood of a parameter set (θ) from the study of a sample ($\mathbf{x}$) is defined by the likelihood function (eq. 3.6)[9]. It is a measure of how likely it is that $\mathbf{x}$ was the outcome of samples of a random variable with an assumed probability model with parameters θ. The likelihood function is thought of as a function of θ with the data $\mathbf{x}$ being fixed. The maximum likelihood estimator for Θ is then

$$\hat{\theta}_{MLE} = \underset{\theta}{\operatorname{argmax}}\, L(\theta, \mathbf{x}) \tag{3.7}$$

Assuming that the likelihood function is the best measure to assess the fit of parameters to a sample, the $\hat{\theta}_{MLE}$ is the best fit of any parameters to the given sample.

Often it is easier to calculate the log-likelihood function due to the sum instead of products, i.e.

$$l(\theta, \mathbf{x}) = \log L(\theta, \mathbf{x}) = \sum_{i=1}^{N} \log f(x_i|\theta) \tag{3.8}$$

Since the likelihood function is positive and argmax is the same for both functions it is satisfactory to maximize the log likelihood instead.

In the case of a Gaussian mixture model the log-likelihood function would be defined as eq. 3.9.

$$l_{GM}(\theta, \mathbf{x}) = \sum_{i=1}^{N} \log \left(\sum_{j=1}^{k} \pi_j f_{\theta_j}(x_i) \right) \tag{3.9}$$

Where $\theta = (\pi_1, \theta_1, ..., \pi_k, \theta_k)$ and f_{θ_j} is the probability density function of a Gaussian with parameters $\theta_j = (\mu_j, \sigma_j)$.

To find the MLE (maximum likelihood estimate, $\operatorname{argmax} l(\theta, \mathbf{x})$) the usual analytical way would be to differentiate the log likelihood function. Set derivatives to zero as well as some additional constraints to prevent local maximas or minimas and saddle points to be detected. Solve the system of equations. For mixture models these equations usually turn out to be highly non linear and impossible to solve analytically. In reality numerical methods are used to find MLE estimates.

EM algorithm

A popular algorithm for estimating $\hat{\theta}_{MLE}$ is the EM algorithm. This is an iterative algorithm which is proven to converge towards a global maximum likelihood. It also has the advantages of low cost per iteration, low storage demand and relative ease of programming. It is known however to converge very slowly in some situations[11].

The algorithm is in a more generalized setting a method for acquiring the MLE from missing data. In the case of a mixture probability model the missing data are which observations that are derived from which component probability distribution. Here each observation from the mixture distribution is seen as being derived from one and only one particular component distribution. Let $\Delta_i^{(j)}$ denote the source of the observation (where $\Delta_i^{(j)} = 1$, if the j:th component distribution is the source of the i:th sample and $\Delta_i^{(j)} = 0$, otherwise). Then the dichotomous Δ-matrix is the missing data and equation eq.(3.9) can be written as:

$$l_{GM}(\theta, \mathbf{x}) = \sum_{i=1}^{N} \log \left(\sum_{j=1}^{k} \Delta_i^{(j)} \pi_j f_{\theta_j}(x_i) \right) = \sum_{i=1}^{N} \sum_{j=1}^{k} \Delta_i^{(j)} \left(\log \pi_j + \log f_{\theta_j}(x_i) \right), \quad (3.10)$$

where the last equality comes from the fact that $\Delta_i^{(j)}$ can only be 1 for one j for each i value. If the Δ-matrix was known the parameters would be quite easy to estimate. π_j would simply be the number of i values for which $\Delta_i^{(j)} = 1$ divided by N. The μ_j would be the sample mean of all x_i for which $\Delta_i^{(j)} = 1$ and σ_j would be the corresponding standard deviation. Therefore the problem would be solved if the Δ-matrix was known. The EM algorithm searches for Δ.

EM stands for Expectation and Maximization. The algorithm consists of an expectation step and a maximization step which is successively performed in each iteration step. To start, some initial values are chosen for the parameters (θ). If no prior information is known the μ_j-values can be set to x_i for some i:s chosen at random. The σ_j values can all be set to the standard deviation of the entire sample set. The component coefficients π_j can all be set to $1/k$.

In the algorithm a "soft" membership matrix values with nonzero values for all component distributions are used (ψ) instead of the "hard" membership matrix (Δ) of the ideal scenario. The algorithm is defined as[9]:

1. Choose initial values of the parameters ($\tilde{\theta}$).

2. Expectation step: Compute the membership matrix

$$\psi_i^{(j)} = \frac{\tilde{\pi}_j f_{\tilde{\theta}_j}(x_i)}{\sum_{l=1}^{k} \tilde{\pi}_l f_{\tilde{\theta}_l}(x_i)} \tag{3.11}$$

3. Maximization step: Maximize likelihood

$$\tilde{\mu}_j = \frac{\sum_{i=1}^{k} \psi_i^{(j)} x_i}{\sum_{i=1}^{k} \psi_i^{(j)}}, \quad \tilde{\sigma}_j^2 = \frac{\sum_{i=1}^{k} \psi_i^{(j)} (x_i - \tilde{\mu}_j)^2}{\sum_{i=1}^{k} \psi_i^{(j)}}, \quad \tilde{\pi}_j = \frac{1}{N} \sum_{i=1}^{N} \psi_i^{(j)} \tag{3.12}$$

4. Iterate step 2 and 3 until convergence.

3.2 Classification

There is a demand for robust classifiers in a large and growing number of applications: Recognizing cancer tumors from moles, radar echoes from electronic noise, sad faces from happy faces, etc. In this thesis classification is the main core of the problem, classifying accelerometer values as belonging to seizure movements or not.

For the subject of the thesis let us make a distinction between cluster analysis and classification. Even though cluster analysis (see section 3.1) can be used for classifying data and therefore can be viewed as a method of classification it does not infer some prior knowledge in to the classification. In this thesis "classification" refers to methods of "supervised learning". In this concept some prior information, usually a data set with known class memberships, is used to train the classifier to classify new data (contrary to cluster analysis which is often referred to as "unsupervised learning"). The classifier will then draw conclusions from the training data and use this to classify new data.

Usually a classifier needs to make a choice if an observation belongs to a class or not. The null hypothesis being that the observation does not belong to the class. Such problems occur when having only two classes or when having several classes but with a strong background class so that the other classes do not generally compete with each other. In such cases, classifying a sample with respect to a particular class can yield four outcomes.

- A True positive. A correct classification of the observation as belonging to the class of interest. Correctly rejecting a false null hypothesis.
- A False positive. An incorrect classification of the observation as belonging to the class of interest when it do not. This is also known as a type I error and occurs when incorrectly reject a true null hypothesis.
- A True negative. A correct classification of an observation as not belonging to the class of interest or a correctly accepting the null hypothesis.
- A False negative. An incorrect classification of an observation as not belonging to the class of interest when it actually does. This is usually known as a type II error and occurs when incorrectly accepting the null hypothesis.

Usually a classifier will not perform a completely accurate classification on a given problem. Depending on the application there is usually an asymmetric severeness of the type I error to the type II error. Therefore it is common in many implementations to weigh the classifier parameters to decrease the risk of one type of error on the cost of an increase of the other kind of error.

3.2.1 Quadratic discriminant analysis

Quadratic discriminant analysis is a parametric classification method based on normal theory. It approximates all classes with Gaussian distributions. This does not say that the samples from the classes need to be Gaussian distributed. The method can work well

on non Gaussian data as well but is not guaranteed to do so. Quadratic discriminant analysis (QDA) is based on Bayes theorem.

Theorem 3.1 *Bayes theorem*
If $H_1, ..., H_n$ are exclusive and exhaustive events on a sample space Ω and K are some event in the same sample space. Then

$$\mathbb{P}(H_i|K) = \frac{\mathbb{P}(H_i)\mathbb{P}(K|H_i)}{\mathbb{P}(K)} \tag{3.13}$$

[12] Since each observation in the data is considered sampled from one and only one class, the classes are considered exclusive events. It is further assumed that the amount of classes in existence are known so the classes are also exhaustive (an observation must be a member of one of the known classes). The event that an observation belongs to the i:th class can then be denoted by H_i, using the notation from theorem 3.1.

Having a sample $\mathbf{x}$, the best classification would be the class with the maximum conditional probability. The class which under the condition of sample $\mathbf{x}$ has the largest probability of all classes. This is known as the maximum a posteriori probability (MAP) decision [13], where the "a posteriori" comes from viewing the probability after the sample was observed, i.e.

$$\underset{i}{\operatorname{argmax}}\, \mathbb{P}(H_i|\mathbf{x}) = \underset{i}{\operatorname{argmax}}\, \frac{\mathbb{P}(H_i)\mathbb{P}(\mathbf{x}|H_i)}{\mathbb{P}(\mathbf{x})} \tag{3.14}$$

Since the denominator is constant for every choice of i, it can be omitted. Since the nominator is always positive and the maximum of the logarithm has the same argmax as eq.(3.14) the maximization can be expressed as:

$$\underset{i}{\operatorname{argmax}}\, [\mathbb{P}(H_i)\mathbb{P}(\mathbf{x}|H_i)] = \underset{i}{\operatorname{argmax}}\, [\log(\mathbb{P}(H_i)) + \log(\mathbb{P}(\mathbf{x}|H_i))] \tag{3.15}$$

Since the classes are assumed to have Gaussian distributions the rightmost logarithm can be written as:

$$\log \mathbb{P}(\mathbf{x}|H_i) = \log f_{\theta_i}(\mathbf{x}) = -\frac{1}{2}\log\left[(2\pi)^d|\Sigma_i|\right] - \frac{(\mathbf{x}-\mu_i)^T\Sigma_i^{-1}(\mathbf{x}-\mu_i)}{2}, \tag{3.16}$$

where $\theta_i = (\mu_i, \sigma_i)$ are the parameters of the i:th class Gaussian distribution. $\mathbb{P}(H_i)$ is the prior probability of an observation belonging to the class i, let us denote it by π_i.

The quadratic discriminant is defined as

$$g_i(\mathbf{x}) = \log(\pi_i) - \frac{1}{2}\log\left[(2\pi)^d|\Sigma_i|\right] - \frac{(\mathbf{x}-\mu_i)^T\Sigma_i^{-1}(\mathbf{x}-\mu_i)}{2} \tag{3.17}$$

and the maximum discriminant corresponds to the same class as implied by the Bayes classification (eq. 3.14).

The reason why it is called quadratic is because of the quadratic term as opposed to linear discriminant (see section 3.2.1). Because of the quadratic terms the decision boundaries of a QDA classifier can be described by quadratic functions.

For training a QDA classifier the training data are used to estimate the parameters $\theta = (\pi_1, \mu_1, \Sigma_1, ...)$. The π_i is estimated by the number of observations belonging to the i:th class divided by the total number of samples in the training data. However the π_i-values are often simply used as weights for weighing false negatives to false positives. The μ_i and Σ_i are estimated using only the observations belonging to the i:th class in the training data. It is important to have enough observations in every class in the training data. If there are too few observations in a class the covariance matrix Σ might be ill posed.

When classifying a new sample ($\mathbf{x}$), the quadratic discriminants $g_i(\mathbf{x})$ are calculated and the i corresponding to the largest discriminant value is chosen as the class of origin.

Compared to more advanced parametric classifiers the QDA usually behaves surprisingly well. According to [9] this can be explained by the more reliable parameter estimates. While the quadratic constraint gives a bias to the model the comparably stable parameter estimates make up for this bias. Especially when using training sets which are small or unreliable.

Linear discriminant analysis

Linear discriminant analysis uses basically the same theoretical construction as QDA but assumes a common covariance matrix for all the classes. When comparing two discriminants with each other, using the same covariance matrix we obtain:

$$\begin{aligned} g_i(\mathbf{x}) > g_j(\mathbf{x}) \Leftrightarrow \log f_{\theta_i}(\mathbf{x}) + \log \pi_i > \log f_{\theta_j}(\mathbf{x}) + \log \pi_j \Leftrightarrow \\ \Leftrightarrow \mathbf{x}^T \Sigma^{-1}(\mu_i - \mu_j) > \frac{(\mu_i + \mu_j)^T \Sigma^{-1}(\mu_i - \mu_j)}{2} - \log(\frac{\pi_i}{\pi_j}) \end{aligned} \tag{3.18}$$

As can be seen, there is a linear decision boundary between the two classes. Therefore the name ”linear discriminant analysis”, the decision boundaries are hyperplanes. The LDA classifier, even though it seems like an extreme simplification, is known to generally perform well on a large set of diverse classification problems [9]. This is believed to derive from the stable estimation of the parameters. The linearity constraint introduces a bias but the low variance of the parameters are known to make up for the bias compared to more advanced classifiers [9]. Just as LDA, QDA is considered a very simple classifier and QDA and LDA share the same pros and cons. The greatest reason for choosing LDA instead of QDA is that the amount of training data compared to the number of dimensions makes the covariance matrix estimation unreliable. Pooling all the data from all classes together can then increase the stability of the covariance matrix considerably.

An interesting note is the fact that the LDA arises automatically in a two class least square regression without any assumption of Gaussian distributions [9].

3.2.2 K-nearest neighbors

The k-nearest neighbor (KNN) classifier is a non parametric classifier that can accommodate arbitrary decision boundaries. It is a simple and relatively intuitive algorithm

that makes very mild assumptions on the structure of the classes. The downsides are that the classification can, for certain data sets, exhibit unstable classification and usually the classifier requires comparable large storage space and computational complexity compared to parametric classifiers [9].

In KNN a new data sample is compared with the training data. The k-number of training data samples that are closest to the new sample (a distance measure, often Euclidean distance) is used to classify it. Usually a simple majority vote among those k training samples decide which class the new sample is assigned to (see eq. 3.19). It is also possible to have fuzzy membership functions, meaning that more than one membership function might be nonzero and the observations are to some degree part of several classes at the same time. In such a case, the amount of the k neighbors that belongs to a certain class can be used as the membership value for that class and observation.

$$\begin{aligned} l &= \underset{j}{\operatorname{argmax}} \sum_{i \in \{i; \mathbf{y}_i \in N_k(\mathbf{x})\}} m_j(\mathbf{y}_i), \\ m_j(\mathbf{x}) &= \begin{cases} 1, & j = l \\ 0, & \text{otherwise} \end{cases}, \end{aligned} \tag{3.19}$$

where $\mathbf{x}$ is a new observation of unknown origin, $\mathbf{y}_i$ is the i:th observation with known origin, $m_j(\mathbf{x})$ is the membership function of class j for observation $\mathbf{x}$, $m_j(\mathbf{y}_i)$ is the membership function of class j for observation $\mathbf{y}_i$ and $N_k(\mathbf{x})$ is the neighborhood hypersphere of $\mathbf{x}$ large enough to include exactly k of the nearest neighbors in the training set.

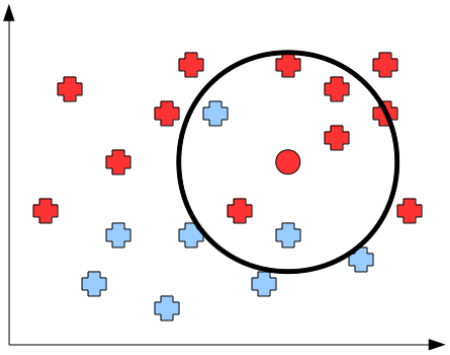

Figure 3.4: Test sample (circle) classified using k-nearest neighbor classification with majority decision and $k = 7$. Training set (cross)

Curse of dimensionality

An effect that occurs when increasing the dimensionality of a classification problem while maintaining the same number of observations is the so called ”curse of dimensionality”. This effect causes a reduction in classification performance under such circumstances. Since adding dimensions to a data set will add additional information, it might seem strange that the classification performance would decrease.

One way of explaining this in the KNN sense is by imagining a scenario in one dimension with x number of training observations inside a line of length unity. The Euclidean distance between a new observation of unknown origin and its nearest neighbour in the training set, let us call the distance y, which is simply the absolute difference in between their one dimensional coordinates. If we would keep the same x number of training observations but have two dimensions, the observations forced to be inside a box with borders of unit length. If we would insert a new observation of unknown origin into this box somewhere, than on average the Euclidean distance to the nearest training data would be further (assuming the new observation was sampled from a uniform distribution). This is realized by studying the Euclidean distance measure between two points in a d-dimensional space.

$$l(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=1}^{d}(x_i - y_i)^2} \tag{3.20}$$

Increasing d will always increase the measure by extra terms inside the square root.

This shows that with higher dimensions the average space between new points and training points increases as well as average space between the training points themselves. With further distance between the new sample and the training data the uncertainty increases. With those extra dimensions the actual classification boundary might be more complex and might require more training data to account for. This is called ”curse of dimensionality”.

This explanation gives an understanding of why it is attractive to work in low dimensions and vast training data with KNN. The curse of dimensionality is, however, present for all kinds of classification and regression techniques. For a parametric classifier the effect can be understood by the following explanation: The parameters of a multidimensional parametric classification model are dependent on the number of dimensions. By increasing the dimensionality, the number of parameters of the model increases as well. Keeping the training data size constant while increasing the parameters will decrease the quality of the estimated parameters. A lower estimation quality of the parameters will reduce the classification performance. [14]

Computer complexity

As mentioned above the KNN algorithm has a comparably higher memory usage and is computationally more complex than most of the parametric classifiers. The straightforward way presented above requires calculation of the distance from each new sample point to all sample points in the training set. The computational complexity of this is of the order $O(MNd)$, i.e.

$$\sum_{j=1}^{M}\sum_{k=1}^{N}\sum_{i=1}^{d}(x_{ki} - y_{ji})^2 \Rightarrow O(M \cdot N \cdot d), \tag{3.21}$$

where y_{ji} are the i:th dimension of the j:th observation in the training set, which has M number of observations. x_{ki} are the i:th dimension of the k:th observation in the test set, which has N number of observations.

Since M is usually needed to be quite large to give a satisfying classification performance, the algorithm will be quite computationally complex. Also M will be dependent on d because of the curse of dimensionality so a larger d will increase the complexity more than is obvious from the equation.

Moreover the M training observations need to be stored in memory and accessed quite frequently during the classification. Since M might be very large and/or the classifier might be realized on a small embedded system this is a big drawback of the straightforward KNN algorithm. Some reduction techniques exist but the KNN algorithm is still quite computational and memory demanding compared with parametric classifiers.

Computational reduction can often be achieved due to redundancy in training data information. Often the decision boundaries are defined by only a few samples around the borders while the bulk of the samples are present deep inside, far from the decision boundaries. Removal of such inner points would not affect the classification performance while reducing the computational complexity and storage demand heavily. Another method of reducing the computational complexity is to divide the sample space in to mutually exclusive domains. When testing a new sample, the k nearest neighbors would only need to be evaluated among the training data inside the same domain and neighboring domains of the test observation.

3.3 Transition matrix

For this thesis the aim is to find epileptic seizures using accelerometers. For the more subtle seizures, the feature value at a certain time instance in itself might not be discriminatory of a seizure. Also, some time instances in what is marked as a seizure is normal behavior prior or after the actual seizure. Here, the key idea is that a seizure is made up of certain episodes in a specific order. Each episode might be recognizable through the feature values for those time instance. The discriminatory information then lies in the order and duration of such episodes (see fig. 3.5).

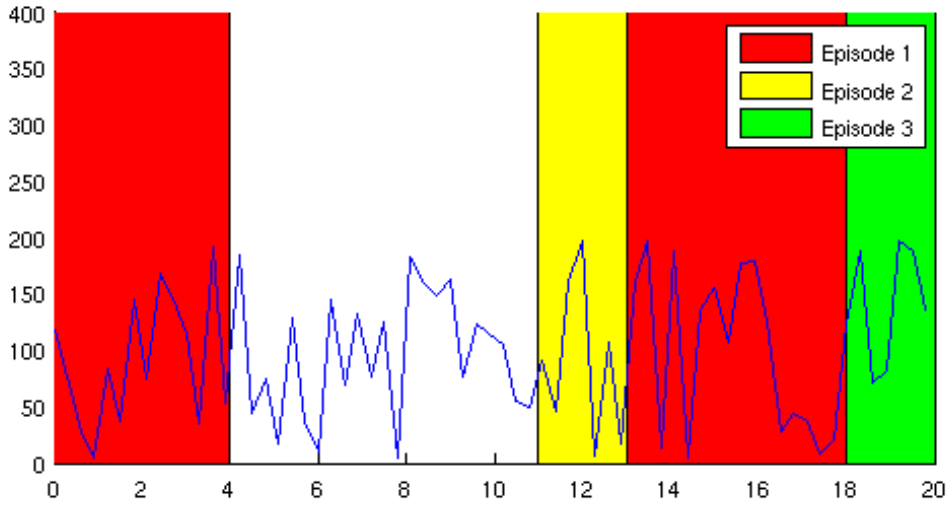

Figure 3.5: Example of how a seizure might be characterize by a certain order and duration of its episodes.

In this thesis, we have developed a "transition matrix" extraction technique. These "transition matrices" has been developed to represent the duration and order of episodes during a fixed time interval. Each transition matrix represent the order and duration of episodes during the interval between two points in time (t_1 and t_2). There is a fixed time length, k, between t_1 and t_2. The time interval is then moved one discrete time step for every extracted transition matrix. In this way there is a transition matrix representing the order and duration of episodes during any time interval of length k amongs the data.

Prior to the extraction of the transition matrices, the episode memberships of each time instance are assumed to be known. The diagonal of the transition matrix simply holds the sum of all membership values for a certain episode during the specified time interval. The diagonal of the first row is the sum of the episode enumerated as '1' etc. The summation is carried out through a convolution with the sum kernel K_s,

$$K_s = [1, ..., 1, 0, ..., 0]. \tag{3.22}$$

The amount of '1': in the K_s kernel is equal to the k-value, i.e. it sums the k number of time instances membership value forward in time from the base of the transition matrix.

The amount of '0': is one less than number of '1':s. The diagonal values are calculated using

$$T_{ii}[n] = \{f_i * K_s\}[n] = \sum_{m=n}^{N-1} f_i[m]K_s[n-m], \tag{3.23}$$

where $f_i[m]$ is the membership value of the i:th cluster and m:th time instance, $T_{ii}[n]$ is the value of the i:th diagonal element of the n:th transition matrix. The Kernel index $K_s[x]$ is from the middle of the kernel, i.e. index values $x > 0$ are zero valued and index values $x \leq 0$ have the value of one.

The non diagonal elements of the transition matrix hold measures of the transition from one episode to another. These measures are extracted by iterating through all time instances in the time interval k similarly to how the diagonal was acquired. However, instead of just summing the membership values of each cluster it sums up transition probabilities between two episodes.

The probability values are represented by the $G_j[n]$-values. These values are the difference between the number of j-episodes that exist in a small neighbourhood, l, after the n:th time instance as compared with the number of j-episodes that exist in the neighbourhood with length l prior to the n:th time instance. If the value is negative, i.e. there are more j-episodes present prior than after n, the value will be thresholded to a zero value.

The G values are then multiplied with the membership functions $f_i[n]$ to register a probability of a transition from episode i to episode j. One might think that it would be easier to present a transition as $f_i[n]$ multiplied with $fj[n+1] - fj[n]$ but due to the presence of noise, this would not have been a robust measure of transitions. The G values works by averaging in some sense.

$G_j[n]$ is defined as

$$G_j[n] = \max\left(\{f_j * K_t\}[n], 0\right), \tag{3.24}$$

$$\tag{3.25}$$

where K_t is the smaller convolution kernel, referred to as the "transition kernel" (in comparison with the "summation kernel" K_s):

$$K_t = [1, ..., 1, -1, ..., -1], \tag{3.26}$$

with the length of two times l number of time instances. The length of l should be longer than 1 second but not much larger. A value around 10 seconds should give relatively stable transition estimates.

The actual values on the non diagonal of the n:th transition matrix, T_{ij}, for the i:th row and j:th column is the transitions between episode i to episode j summed up over the same time interval k as was used in the diagonal. This is simply defined as the sum of all $G_j[n]f_i[n]$ values for all $n \in [t_1, t_2]$. The summation is performed by a convolution with the K_s kernel just as in the case with the diagonal.

$$T_{ij}[n] = \left\{(G_j f_i) * K_s\right\}[n], \quad i \neq j \tag{3.27}$$

See figure 3.6 for a simple sketch of the fields in the matrix, also see figure 3.7 for further understanding on the summation and transition kernels.

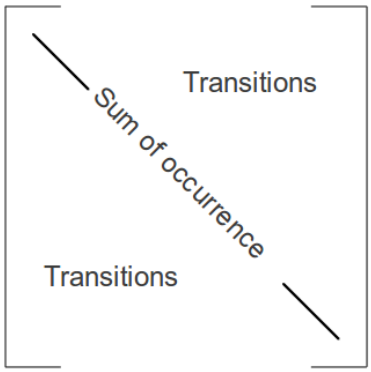

Figure 3.6: Explanatory sketch of a transition matrix.

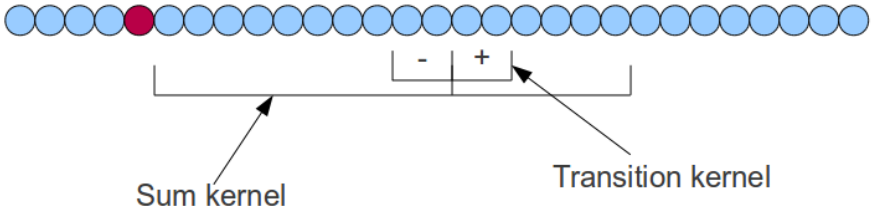

Figure 3.7: Explanatory sketch of the kernels summing up values at different time instances. The colored time instance show the base of the current transition matrix. The time interval of the sum kernel is marked. Also the time interval for one n-value of G_j is marked.

3.3.1 Extreme value characterization

Logging extreme values in a specific time interval is an alternative to the regular transition matrix. It is mainly implemented for working on feature space data directly without a prior clustering analysis. The idea is that some seizures might have high extreme values in different dimensions inside them. By gathering the extreme values for each dimension during a specified time interval a typical seizure behavior might be found. This idea was developed to mimic the way I (the author) was searching for seizures in the data using graphs of the feature space.

There is no matrix notation associated with this use. Instead there are the original dimensions of the features. Where each time instance will have the value of the chosen extreme value for each dimensions. The extreme values are found in a neighborhood of k number of time instances forward in time from the base instance.

3.4 Dimension reduction

Since classifiers perform worse if the dimensionality of a system increases while the number of samples are constant (see sec 3.2.2), reducing the dimensionality while still keeping the discriminatory information is useful. This section recalls two techniques that can be used for reducing dimensionality while minimizing loss of important information.

Beside these two techniques there are many other ways of performing dimension reduction. The cluster analysis techniques described in section 3.1 can for instance be used as ways of reducing dimensionality, as long as the number of clusters are lower than the number of dimensions. One can also look at the way that cluster analysis is used in this thesis as a way of performing dimension reduction as well (see sec 4), even though the cluster analysis in this thesis also serves other purposes. For a more exhaustive introduction to dimension reduction in general the interested reader should have a look at [13] and the other related articles in the same issue of the journal.

3.4.1 Principal component analysis

Principal component analysis (PCA) is a popular dimension reduction technique that does not infer any prior knowledge of what kind of information that is important. It is in this aspect a blind technique that only reduces dimensions based on the variation of data itself.

The principal component analysis will find an orthogonal basis of the data with basis vectors sorted by the overall variance present in their direction. By this technique the linear subspace that accounts for a certain amount of the overall variance can be extracted. PCA can be very effective when some of the sample variables are linearly dependent of each other which usually occurs in real life applications with many dimensions. Typically for an implementation a variance threshold or a maximum number of dimensions are stated. The linear subspace which accounts for the variance threshold are then extracted using PCA.

Assuming that the data has been normalized in advance such that each dimension is of the same order (z-score), the subspace of lower variance will represent the information which are not changing as much as the others inside the data set. Information that do not change might not be as important in an implementation. Beside from the fact that this is probably an indication of that less important data are present in this subspace, this subspace will be more affected by noise since the actual data themselves do not differ so much. Therefore, in applications it is customary to regard the low variance subspace as noise or at least disturbed by noise. In fact the reduction of this ”noise” might actually enhance the classification due to greater disparity at low variances [13]. Except from reducing dimensions and suppressing noise, PCA, by removing linear dependence among the data dimensions in itself might be an attractive feature since some analysis methods are based on inversion of covariance matrices. With linear dependence among the dimensions of data, this will cause ill conditioned matrices and thereby decrease stability and usefulness of such methods.

How to determine the threshold between useful data and noise is up to the implementation. Usually some information are known or believed in advance that might help to set a proper threshold.

In a problem such as in this thesis, where the amount of data belonging to seizures and non-seizures are asymmetric and the possibility for non-seizure data to be of very different nature, it is not necessarily that discriminatory information is located in a subspace of high variance. The variance between the non seizure data alone might be much higher than that of seizure to non seizure data. Therefore a quite high variance threshold is recommended in such problems to make sure that the discriminatory information is included.

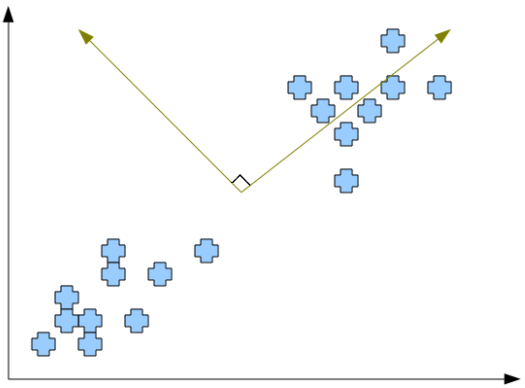

Figure 3.8: Data mainly occupies the diagonal insinuating that the two variables correlate. PCA spans the data space with a new basis with the first basis vector parallel to the diagonal of most variance. In this example the first basis vector is probably sufficient for representing the data in most cases.

Algorithm

Principal component analysis is based on estimation of the covariance matrix. Before estimating the covariance matrix the data need to be normalized so that each dimension have the same magnitude (if same unit/magnitude is not guaranteed implicitly).

The covariance matrix is estimated from the data.

Since a covariance matrix is real valued and symmetric it is possible to diagonalize it by a orthogonal matrix (Spectral theorem [15]). The new representation of the covariance matrix

$$C = VDV^T, \tag{3.28}$$

where C are any covariance matrix, D are a diagonal matrix of eigenvalues and V are an orthogonal matrix with columns being eigenvectors corresponding to the eigenvalues of D, , first maps the data in to the eigenspace with, V^T, and then represents the covariance matrix in the eigenspace where the data is obviously uncorrelated as seen by the diagonal covariance matrix, D. The variances can be seen as the diagonals of D and can be sorted from largest to smallest, make sure to keep track of which eigenvalue corresponds to which eigenvector.

Once D and V are sorted in descending order the total variance of the subspace including eigenvectors $\{v_1, ..., v_n\}$ can be calculated by:

$$T_n = \frac{1}{N} \sum_{i=1}^{n} D_i, \tag{3.29}$$

where N is the dimensionality of the original data space, D_i the i:th eigenvalue after the eigenvalues have been sorted and T_n the total variance of the linear subspace accounting for the largest amount of variance using n dimensions.

The principal component space is then chosen as $\tilde{V}_n$, which is the space spanned by the n first eigenvectors after they have been sorted by the eigenvalues in descending order. n is either chosen by a threshold on the number of dimensions to be used or by a variance threshold fulfilled by T_n.

3.4.2 Partial least squares discriminant analysis

Partial least squares discriminant analysis (PLS-DA) is one way of performing a goal oriented dimension reduction. Here the linear subspace with best discriminatory power between classes known a priori are extracted. It can be compared with PCA (see sec.3.4.1) where the variance is no longer the regular variance among data but instead the variance between data values of different classes. A basis direction with high variance is characterized by large difference in overall values between data samples corresponding to one class compared to the other classes.

By applying PLS-DA to a training set with known class memberships it is possible to reduce the dimensionality while retaining the discriminatory information from the data. Since fewer dimensions increases performance of a classifier the contribution of a PLS-DA based dimension reduction can be substantial.

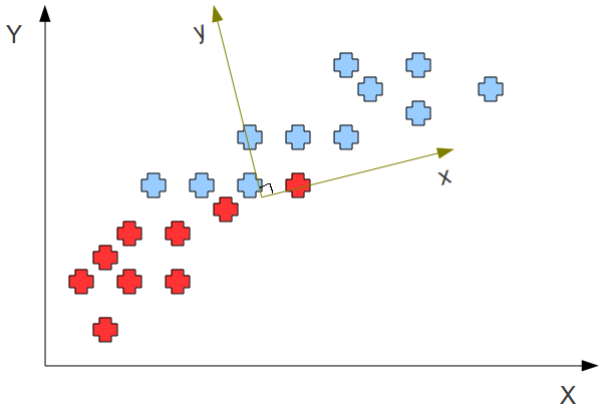

Figure 3.9: A new basis is extracted where the new x-axis have the best discriminatory properties between the red and blue data of all single directions possible.

Partial least square regression

PLS-DA is based on partial least squares regression (PLS-R). PLS-R can be thought of as a mixture of principal component analysis and multiple linear regression. It both transforms the dependent and independent variables to orthogonal subspaces and fits a linear model to the data at the same time. In addition, the algorithms are iterative and the user control the error bound with power to stop the algorithm when the desired precision has been fulfilled.

The PLS model has a PCA style linear subspace dependence of the dependent variables (Y) to their subspace Q and independent variables (X) to their subspace T. These dependences are known as the outer relations,

$$X = TP^T + \tilde{E} \tag{3.30}$$

$$Y = UQ^T + \tilde{F}. \tag{3.31}$$

Moreover a linear relation is assumed between T and Q just as in multiple linear regression,

$$U = T\beta + \epsilon. \tag{3.32}$$

This linear assumption is known as the inner relation. [16]

Here T is the subspace of X, P the basis of the subspace and $\tilde{E}$ the error term introduced by the dimension reduction. Similarly U is the subspace of Y with the basis Q and error term $\tilde{F}$, β the linear models coefficient vector and ϵ the error term from the linearization.

By successively increasing the number of components to use in T and performing the analysis of variance on the least squares regression defined in eq.(3.32) it is possible to compare the current variance measure with a threshold value. In this way the PLS regression can stop when demanded precision has been fulfilled. For each new principal component the problem is deflated by the prior principal components and therefore the algorithms scale well and there is very little overhead in checking variance before continuing with next principal component.

When working with categorical data such as in PLS-DA the same regression model is used but here the dependent variables have discrete values. In the case of PLS-DA one dependent variable vector for each class is needed, i.e. Y has as many columns as classes. If there is only two classes present or if there is one "mother" class outside of the smaller classes, one column can be omitted in Y. Hence PLS-DA is just using partial least squares regression on categorical data.

The name "partial least squares" derives from the fact that a least squares solution is solved in each iteration of the original algorithm.

In the PLS model, the observed variables X and Y are regarded as observations of some underlying true variables referred to as latent variables. Hence neither X or Y are assumed to be independent. In many applications of PLS modelling the aim is to find the number of those latent variables. This model is therefore popular in areas such as chemometrics and econometrics where an exploratory analysis of latent variables might be the real goal (and not necessarily discriminant analysis or regression in their own right). [17]

3.5 Model assessment

Evaluating the model is not a trivial task and depends on the data and the system one wants to model. A very simple model might assume too much about the system and therefore have a strong bias. This will cause the model to perform badly. On the other side of the spectrum there are models that are too complex. These models will learn to classify the training set very well but might discover tendencies that are not general for the system but exclusive for the training set itself. These models therefore have a large variance. This is known as over fitting and will cause the model to perform badly on new data.

There is always a bias-variance tradeoff when modelling a system. In a setting with vast amount of data the ideal procedure is to divide data into a training set, a validation set and a test set. The training set is used to fit the model. The validation set is compared with the fitted model to recognize over fitting, see fig.(3.10). Finally when the model is fitted with optimal parameters to approximate the validation set a test set is classified to assess the performance of the model.

Here the idea is to have the test set completely separated from the two other sets until the complete model with the optimal parameters has been fitted. In this way the test set performs like the model is expected to perform in reality. It is a good idea to perform the assessment several times with the training, validation and test sets randomly and exclusively sampled from the data. In this way a more stable assessment will emerge since which samples belonging to which set will change the prediction result somewhat, [9].

In the setting of this thesis the data are not that vast. An accelerometer measurement usually have less than 20 seizures and always fewer than 40. Since the time dependence is included transition matrices covering the same seizure will be closely related and can not be considered as independent observations. Therefore the data are too sparse to be able to do a proper training/validation/test division for assessment. Instead, here the validation set is skipped altogether and the assessment uses some sporadic enumerations of parameters together with the sparse data technique ”cross-validation” (see sec.3.5.3) for assessing the models.

When fitting a model and assessing it against a test set (or validation set) there are different ways of dividing up these sets. Each method has its advantages and disadvantages. Below, we give a brief introduction to the four most common and useful methods.

3.5.1 Data separation

The most intuitive way of dividing the data sets for assessment is the data separation method. This involves separating the data from the start and using one training set, one test set and one validation set. If performing the analysis several times new data are used in every run.

This method is best suited for data rich environments since it uses much more data

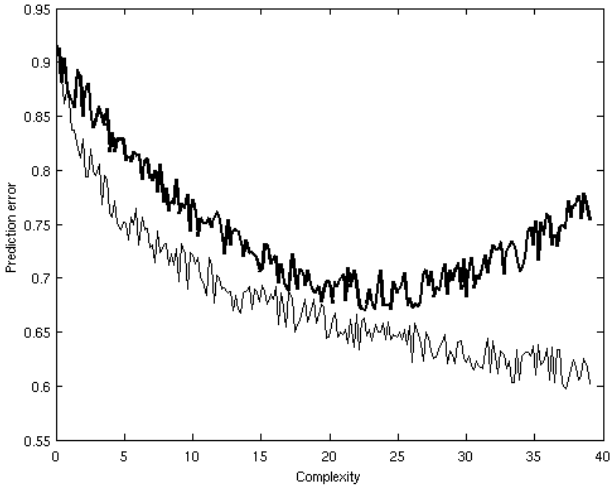

Figure 3.10: Prediction error as a function of model complexity. The training set (thin line) monotonically decreasing since it is used for fitting the model. The validation set (thick line) having a global minimum at a point where the bias-variance tradeoff is optimal.

than the other methods. It is advantageous in the aspect that it is the most realistic assessment since each assessment analysis will use data which were not present in the other ones.

3.5.2 Resubstitution

This method uses the same data for both training and testing. Since it reuses the information the assessment is positively biased and will always give an optimistic prediction. How optimistic the prediction is depends on the general system, model, parameters and sample size. It might be hard to estimate the bias but one should always regard the assessment some at least somewhat optimistic. The advantage of this method is that it can work with sparse data and is fast to compute.

3.5.3 Cross validation

Cross validation reuses data to be able to perform unbiased assessments without having vast data. The idea here is to divide the data set into several mutually exclusive and exhaustive subgroups known as ”folds”. An analysis is then performed where all but one fold is used for training and the leftover fold is used for evaluation (see figure 3.11). This analysis is performed with a new fold acting as the evaluation fold until every single one of the folds have been used for evaluation. In this way all of the data has been used as a test set without introduction of a positive bias and the training set is as large as all but one fold. The tradeoff is that training and classification is performed several times. The size of the folds will control the computational complexity-variance tradeoff. When using few folds, the variance will be high since the prediction depends on in which fold

the samples were assigned. When using a large number of folds the variance will be lower since more of the data are used for training. On the other hand there will be much more computations performed since the analysis needs to be rerun for each fold.

The disadvantage of this method is that it is computationally very demanding when using a large number of folds and might have a high variance when using a small number of folds.

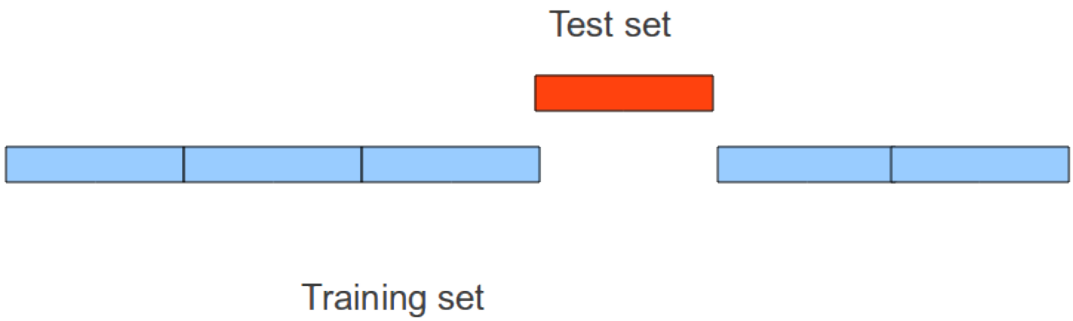

Figure 3.11: Sketch of how a data set can be divided into ”folds”. The red fold are left out of the training set and used for evaluation instead. For each iteration of the crossvalidation loop, a different fold is left out.

Chapter 4

Proposed method

In this chapter, the proposed classification method developed in this thesis work is presented. The chapter provides instructions on how to apply the methods described in chapter 3 to the data described in chapter 2.

The chapter is divided in to several steps, which should be performed in consecutive order. Most of these steps are represented for both the training and test sets. The training set is used to fit the parameters of the steps and the test set is later subjected to the algorithm in each step.

These steps follow the flow chart in figure (4.1) which are as following.

1. **Preprocessing** The preprocessing stage loads the selected features and uses the training set to perform a principal component analysis in order to reduce the dimensionality. The test set are later reduced using the principal components acquired by the training set.

2. **Clustering** During the clustering stage the seizures in the training set are clustered to find typical seizure behaviors. The test set is later categorized into those clusters.

3. **Transition matrix** In this stage transition matrices are extracted to account for the presence and transitions between different clusters during a fixed time length. This stage is the same for the training and test sets.

4. **PLS-DA** The dimensions of the transition matrices of the training set are reduced by partial least squares discriminant analysis. In this way a goal oriented dimension reduction is performed. The test sets transition matrices can later be reduced into the same linear subspace as was discovered in the training set.

5. **Classification** The classification are performed after the PLS-DA dimension reduction on the transition matrices. The classifiers parameters are fitted by the training data and the classification is later performed on the test data.

6. **Post-processing** The post-processing step is only for the test sets. It involves melting together points close in time which are classified as seizures. After this, the points classified as seizures but located too far away in time from other seizure classifications are removed.

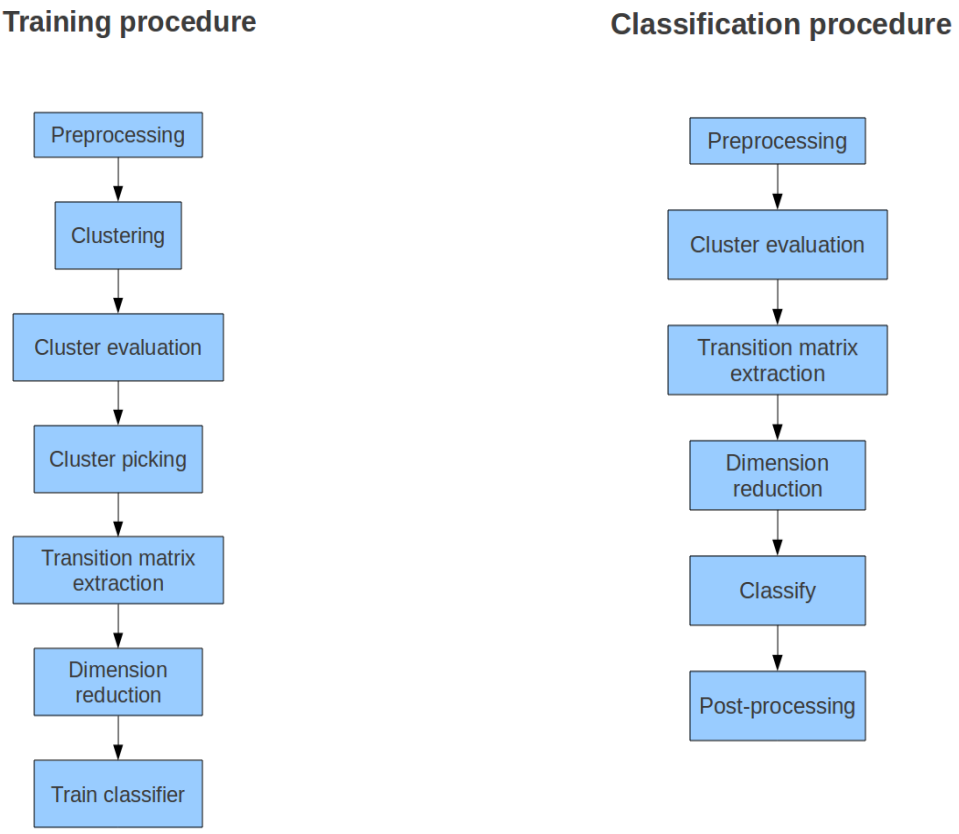

Figure 4.1: Block diagram depicting the different stages of training and classification of seizures.

Before these stages can commence the actual data have been sampled and stored by a suitable hardware configuration. Depending on how and where the signal processing is performed the implementation of the method might vary. For instance if the signal processing is performed on all gathered data at a powerful workstation offline, there might be different requirements than if it is performed in real time on an embedded system. Even though the implementation might be quite different between those diverse systems, the instructions in this chapter are written in general terms so it will apply to them both.

4.1 Preprocessing

4.1.1 Data transformation

The accelerometer data are originally delivered as 16 bit scalar numbers, one for each axis and sensor, sampled 50 times a second. This raw data is then to be transformed in to

51 summary statistics for each second (51 if using 3 sensors). These summary statistics are defined and explained in section 2.3. Let us refer to the new data space from this transformation as the ”feature space” compared to the original raw data space that we might refer to as simply the ”data space”. The reason for this data transformation is to extract a feature space that should respond more clearly to differences in motorical movements from seizures and none seizures. Also, this transformation introduces some concept of time into the feature space even though the time dependence is in the range of a second.

As explained in section 2.3, there are some dependence of observations from neighboring time instances in the feature space. More than this dependence the biomechanical movements that the accelerometer data are derived from introduces inherent dependencies between samples close to each other. This last dependency can be explained by a simple example: For an accelerometer attached to a wrist to register downward acceleration compared to an accelerometer at the chest, the arm needs to be raised somewhat above the lowest position prior to the downward acceleration to be able to perform a downward motion. Here, the constraint originates from the fact that the wrist is connected to the torso by an arm.

These dependencies show that the samples in the feature space are not fully independent of each other, therefore one has to be careful in applying any method that assumes independent observations.

The features that should be extracted are chosen based on some prior knowledge or all might be included. The classifier might work better on some specific feature sets compared to using all features together (see section 3.2.2.

4.1.2 Principal component analysis

After the data have been transformed into feature space the dimensions are normalized. This is done by computing the sample mean and variance for all dimensions separately (51 dimensions if using all sensors and features) from a training set. The features are then first translated by the sample mean and then divided by the sample standard deviation. These normalizations are performed since different features will have different magnitudes and units. When analyzing the feature space later on the features with larger values will usually be weighted heavier even though their discriminatory power is not necessarily better than other features. By forcing the dimensions to be of the same order of magnitude, these problems are avoided. It is however important that the training data are representable of later data so that the normalization works well on new data as well (the normalization parameters mean and standard variations will be estimated from the training data).

It is expected that some of the features might be correlated and that their is some noise introduced in the measures. Therefore a principal component analysis is performed (see section 3.4.1). This is a first step to reduce the dimensionality while keeping the most of the relevant data. A relevant threshold is chosen such that the subspace used accounts for a certain percentage of the total variance. The complement to this subspace might

be regarded as noise since it accounts for low variance. Another positive side effect of removing low variance subspaces is that the risk of ill conditioned covariance matrices is greatly reduced in the clustering stage in the case of a covariance based clustering method. Using a high value of the variance parameter reduces the risk of removing the discriminatory important subspaces while still eliminating the obviously linear dependence among features and removing some noise and those features that hardly react at all. In this thesis, the variance threshold has been 90% which has worked well and usually reduced the dimensionality to about half.

4.2 Clustering

When having pre-processed all data the next task is to find typical "episodes" of the seizures. Here several options can be chosen.

First of all the most suiting clustering algorithm might depend on the type of seizure and the type of background data. For some data there might exist a clustering setting which identifies a cluster which in itself is completely discriminatory between a seizure and none seizure (this is the ideal situation). In other cases there might be nothing out of the ordinary found in the seizures and the clusters will be present everywhere in the data. The most data however will have certain behaviors that are typical but not exclusive for seizures. Here the discriminatory information might lie in the order and duration of such behavior. The cluster analysis is performed to find such behavior that in a later stage can be analyzed for discriminatory classification.

The clustering techniques that are implemented and tested are listed below. Each one have its advantages and disadvantages.

4.2.1 Fuzzy Gaussian mixture

The fuzzy Gaussian mixture option uses Gaussian mixture clustering (see sec 3.1.2). The word "fuzzy" refers to the membership function allowing for partial membership in several clusters at the same time. Therefore, samples located in between two clusters might be equal members of both clusters.

Using this option k number of Gaussian distributions will be estimated from the seizure data exclusively (k being a parameter). The reason for using the seizure data exclusively is to find clusters typical to seizures. The Gaussian mixture model is acquired by using training data to estimate the mean and covariance matrices of the Gaussian distributions using the EM algorithm (see sec 3.1.2). Choosing a too large k compared to the number of dimensions and training seizure data will yield unstable estimates. Therefore one can choose to use only diagonal covariance matrices to increase the stability of the estimates on the expense of the adaptivity of the model. Choosing a small enough k the most attractive attribute of this clustering algorithm is the shape adaptivity of the clusters. The clusters will adapt their shape to the actual data.

4.2.2 Hard Gaussian mixture

The hard Gaussian mixture is basically identical to the fuzzy one but the samples get classified only to one cluster each, i.e. the membership functions are dichotomous. The algorithm works exactly like the fuzzy one but the last stage is introduced where membership are awarded solely to the cluster with highest membership value for each sample.

The advantage of this hard membership will be that the next stage in the proposed method (transition matrix stage) might yield more robust transition matrices due to the dichotomous membership values. The drawbacks are that information will be lost on observations that are in the middle close to several clusters at the same time.

4.2.3 Hard K-means

The K-means option uses the regular k-means algorithm (see sec 3.1.1). A fast combinatorial algorithm that do not assume any probability model. Here k number of clusters are chosen in advance and searched for in the seizure data. K-means uses hard dichotomous membership values for each sample and a hyperspherical shape.

4.2.4 Seizure separated K-means

The seizure separated K-means option performs clustering on each seizure separately. These clusters are then compared with non seizure data and the other seizures. Here there are k number of clusters to search for in each seizure. Of those k, the n number of best clusters are chosen. Since it is hard to compare clusters from different clusterings. The chosen clusters are all derived from the one seizure that gave the best seizure clusterings.

The clusters that are present in as many seizures as possible and not so common outside the seizures are chosen as the clusters. With this option there is a possibility to really find good behavior that is typical for seizures and uncommon otherwise. However, this is computationally more complex since there are so many k-means clusterings to perform and evaluation of them as well. Also seizure separation might not work with Gaussian mixture since there might be too few data observations to acquire stable estimates.

4.2.5 No clustering

There is also the possibility of choosing no clustering and work with the feature space in the next step instead.

4.3 Transition matrix extraction

In the clustering stage all samples were assigned a membership value to each of the extracted clusters. These membership values are used to compute transition matrices for all time instance in the data (see sec 3.3). The idea is that these transition matrices will have discriminatory properties since they hold information of the time evolution of cluster memberships during a time interval.

When computing transition matrices there are basically two options. One option is the regular transition matrices defined in section 3.3, these are the true transition matrices. An alternative is to search for extreme values inside the time intervals. This choice is suited for handling feature data without clustering but it might give some valuable information when working with fuzzy membership functions as well.

4.3.1 Extreme value characterization

Logging extreme values in a specific time interval is an alternative to the regular transition matrix (see sec.3.3.1).

The extreme values to search for might be minimum, maximum or maximum absolute value. The dimensionality of the extreme value "transition matrix" will be equal to the dimensionality of the feature space (4.2).

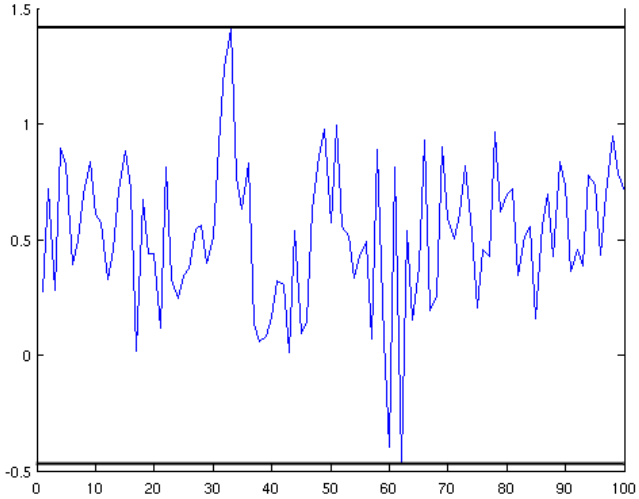

Figure 4.2: Example of how the maximum value and minimum value are identified with the black horizontal lines.

4.4 Classification of transition matrices

This is the stage where actual supervised learning will occur. In the training phase a PLS-DA dimension reduction will first reduce the dimensionality.

4.4.1 Dimension reduction

A linear dimension reduction is performed using partial least squares discriminant analysis (PLS-DA, see sec 3.4.2). This reduction is similar to the principal component analysis but instead of extracting a subspace according to the variance of the data itself, the partial least square algorithm extracts the subspace which accounts for variances of the data compared to the result (result in this case being either part of seizure or not part of seizure). In this way a subspace which holds the important discriminating information is extracted. Just as with the principal component analysis the variance threshold is an adjustable parameter with optimal values being different depending on the data.

When the dimension reduction has been trained on the training data the linear transformation matrix is stored and used to transform new data.

4.4.2 Classification

The reduced data are now ready for classification. In this thesis two different kinds of classifiers have been implemented. The first one being the quadratic discriminant analysis (see sec 3.2.1) and the other being K-nearest neighbors (see sec 3.2.2).

QDA

Choosing the QDA classifier, multivariate Gaussian distributions are estimated to the seizure and non seizure transition matrices of the training data. The new transition matrix data are then classified to belong to the distribution with highest a priori probability (see sec 3.2.1).

The parameters for this classifier are the prior probabilities of seizure and non-seizure. Originally the method uses the occurrence of seizures and non seizures in the training data to estimate the prior probabilities. However if one does not believe that this is a correct estimation or one wants to decrease one type of misclassification error on the cost of the other one (type I or II) it is possible to adjust these values.

KNN

In order to choose the KNN classifier, the training data are used to outline the decision boundaries. A new sample is classified to belong to a seizure if k of its nearest neighbors fulfill some logical statement (see sec 3.2.2). Which logical statement is a parameter in

itself. the choices are either a majority vote among the k nearest neighbors, the presence of at least one seizure in those k neighbors or if all should be.

Here the parameters are the k-value, being the number of nearest neighbors to examine and which logical statement that should be fulfilled (in the evaluation, only majority decision has been used).

4.5 Post-processing

When new unknown data have gone through all the stages up and including the classifications stage, some post-processing needs to be performed to remove single classifications far from others. Since the seizure behavior should (theoretically) be recognized in all transition matrices that include the seizure in their time span, some single classification points alone in their neighborhood should not occur. If such classifications have occurred they have to be misclassifications and should therefore be removed. This is performed by using morphological operators.

First a morphological close operation is performed on the data to close gaps in between points close to each other that are all classified as seizures. This needs to be done since noise might yield that a few points in the middle of a seizure segment are not classified as a seizure. After this a morphological open operation is performed to remove those stray seizure classifications that after the ”close” operation are not part of a larger seizure segment.

The morphological operations ”close” and ”open” are based on two other morphological operation, ”dilate” and ”erode” [8]. These operations are explained below, see figure (4.3) for further understanding.

4.5.1 Dilate

Morphological dilate in the setting of one dimensional binary data works by adding ones around already existing ones by a defined radius. This causes those ”islands” of ones to grow or dilate by the given radius.

4.5.2 Erode

Erode is the opposite of dilate. Here the zeros will ”dilate” by a radius causing the ”islands” of ones to shrink or erode by given radius.

4.5.3 Open

Morphological open is an operation that removes small stray blocks that are not connected to other blocks. This is done by first performing a ”erode” operation which will remove

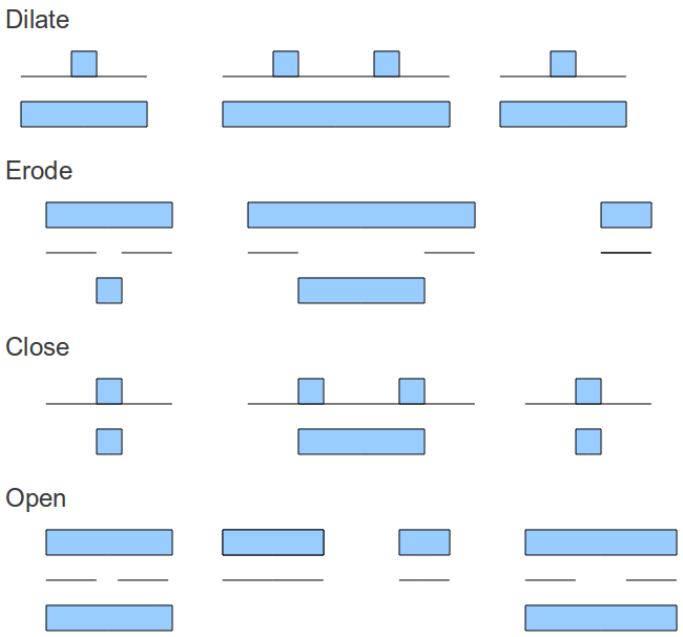

Figure 4.3: Explanatory sketch of binary morphological operations in one dimension. The top blocks in each operation sketch show the distributions of seizures before the operation. The bottom blocks show the distribution of seizures after the operation. The lines in between the top and bottom blocks show the radius of the operation.

too small segments completely. After this a ”dilate” operation restores the shape of the segment that was large enough to survive the erosion.

4.5.4 Close

The morphological close operation is used to close small gaps between ones. This is done by first performing an ”dilate” operation which will increase all segments by a certain radius. An ”erode” operation is thereafter performed to shrink the segments by the same radius. After these two operations small gaps have been filled in or closed.

Chapter 5

Results

5.1 Evaluation method

In this chapter, results from the evaluations of the proposed classification method (see chapter 4) are presented in order to establish assess the classification performance. Since there was an abundance of both, features and parameters to choose from, the aim of the evaluation is to establish:

- Best performance of the proposed method on each of the data sets used in the analysis. These data sets are the acceleration measurements made on real persons (see sec. 2.2).
- Best parameters for each of the data sets used in the analysis
- Which features holds discriminatory information and which do not.
- How well the proposed method perform compared to the already existing method (see sec. 1.3).

The two main measurements used to assess the results in this chapter are "sensitivity" and "selectivity".

Sensitivity being defined as:

$$\text{Sensitivity} = \frac{Y_{tp}}{Y_{ap}}, \tag{5.1}$$

where Y_{tp} are the number of true positives (actual seizures classified as seizures) and Y_{ap} are the number of actual seizures. In words, true positives divided by actual positives or the amount of existing seizures that was found.

Selectivity being defined as:

$$\text{Selectivity} = \frac{Y_{tp}}{Y_{tp} + Y_{fp}}, \tag{5.2}$$

where Y_{fp} are the number of false positives (number of sequences wrongfully classified as seizures). In words, the amount of areas classified as seizures that actually were seizures.

Both measurements assumes values between 0 and 1 and large values characterizes good classification for both measurements.

5.1.1 Data limitations

Since the proposed method incorporate time evolutionary information from the measured data it means that each seizure can only be measured with one sample point (transition matrix), if sample points are supposed to be completely independent. This means that the setting is extremely sparse in data associated with seizures. The available seizure logs from Sahlgrenska does usually contain less than 10 seizures each, with the most numerous being 22 seizures. During the measurements, experts (doctors and nurses) witnessed actual seizures and have provided their opinion after analyzing EEG data scans. Also, video recordings has been available to view the actual movements recorded by the accelerometers.

Inconsistencies are common among the measurements with confirmed seizures, meaning that some seizures show a behavior inconsistent with other seizures. Since data are sparse it is even more important that all seizures are consistent. This means that from the available data, only some seizures can be included while the rest needs to be marked as ”don't care regions”, i.e. regions neither regarded as seizures or non seizures. Thus the data useful for training the classifier is actually less than it were like originally.

From this background, only two data sets from the Sahlgrenska measurements were satisfactory for evaluating the performance of the proposed method, see sec.5.2 and 5.3. Two sets of non authentic, enacted seizures have been included to assess the performance of the proposed method without the restrictions inherent with the authentic Sahlgrenska data, see sec.5.4 and 5.5.

5.1.2 Assessment

All results have been assessed through cross validation (see sec.3.5.3).

Here, the folds of the cross validation are divided such that all seizures are represented mutually exclusive by different folds. There might be some folds with only background data in them as to make all folds comparable in size.

After the extraction of transition matrices, matrices that do not span a time interval including seizures are chosen so that there is no overlap between those matrices chosen for the training data.

The transition matrices accounting for seizure data on the other hand are chosen such that 30 different transition matrices accounts for each seizure. These matrices are aligned such that, except for the seizure itself, they account for slightly different amounts of data before and after (see fig.5.1). Thus the transition matrices accounting for the same seizure are not independent but accounts for different variances of matrices depicting the same seizure. All transition matrices depicting the same seizure are sorted in to the same cross validation fold so that no bias will be introduced.

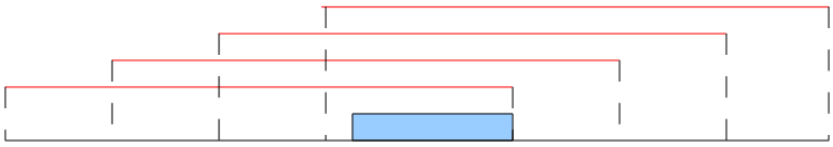

Figure 5.1: A sketch of how several transition matrices are chosen that accounts for the same seizure but including different parts of the background prior and after the actual seizure.

Due to stochastic elements in the evaluation process (choices of folds, clustering analysis, choices of background transition matrices to incorporate), different runs of the algorithm might yield different results. Therefore several runs using the same parameters are done to evaluate the expected performance of that particular model.

The need of repeating evaluations several times together with cross-validation, clustering analysis and transition matrix extraction of large amounts of data, make model assessment a computationally demanding task. This eliminates the possibility of enumerating through all combinations of parameter- and feature-sets. Instead, some parameter set is chosen at start that works quite well (with all features present). From this base, small variation in parameters are performed. Such variations might be to change the clustering algorithm, change the clustering parameters slightly, change amount of variances accounted for in PLS-DA dimension reduction, change classification method or change parameters to the classification method slightly. In this way, a hint on what yields better results are acquired.

For each parameter set, 5 evaluations are performed to assess the stability of the result. From the assessment of the parameters, the best parameter set is later chosen to evaluate the feature sets. Here the idea is to divide the features in to different subgroups. These groups are then excluded one at a time to assess the discriminatory information extracted from that particular group. The aim of this is to draw conclusions of which features that are unnecessary.

As mentioned earlier (see sec. 3.2.2), reducing the dimensionality is in itself attractive, the dimension reduction methods in the proposed method do, however, somewhat take care of this problem. The main reason for reducing the features used are that acquiring them comes with a cost. For an embedded battery powered system, more computations mean a shorter battery operation time. Also, excluding all features from a sensor means that the sensor itself can be excluded, which is cheaper and more comfortable for the user.

Furthermore, the method used in the prior master thesis (see sec. 1.3) has been evaluated using the same evaluation procedure as the method proposed in chapter 4. This is done in order to make direct comparison between the new and older classification method as well as to evaluate the older method with an unbiased evaluation procedure (re-substitution was used in the prior evaluation).

For the old method, the choice of feature and parameter sets are based on the analysis used in the prior thesis, see sec. B.

Since the parameter- and feature-sets evaluated in this thesis are not exhaustive, there is no guarantee that the found results are optimal. Moreover, data with more number of seizures would improve classification further. These results should therefore be seen as a pilot study showing tendencies more than assessing the optimal performance and parameter-/feature-sets. However, if a model yields satisfactory results in this evaluation they are very likely to perform at least as well on sets with more seizures (as long as the seizures are consistent).

5.1.3 Feature groups

The feature groups (see sec.2.3 for detailed explanation of the features) that are excluded one at a time for each run are as follows:

- **VM and MAMD** All VM and MAMD values are excluded to assess if SMA alone contain the same information. This three measures are all ways of measuring general activity. Therefore they are believed to hold similar information. For further understanding see the definitions at sec.2.3.
- **VM and SMA** Similarly but to assess if MAMD alone contains the same information.
- **SMA and MAMD** Similarly but to assess if VM alone contains the same information.
- **DC** To assess if DC values are needed for discrimination. If orientation is an important feature, this means that different models need to be used depending on whether the subject is standing/sitting or lying down, and maybe also in which direction they are lying down.
- **CORR** All correlation values (both linear and circular) to assess if these are important features.
- **PER** All periodicity values to assess if these are important.
- **FREQ** All frequency values to assess if these are important.
- **Highest FREQ** The frequency band of $13.5 - 25$ Hz is excluded. If some sample points are lost in the radio transmission, the highest frequency band will be most heavily influenced by the loss.
- **Sensor 1** All features derived from sensor 1, which is the sensor attached to the chest.
- **Sensor 2** All features derived from sensor 2, which is the sensor attached to the left arm.
- **Sensor 3** All features derived from sensor 3, which is the sensor attached to the right arm.

- **Differentials** This is not actually a separate feature group but a transformation of all features. Instead of using the actual feature values, the difference between feature values neighboring each other are chosen. It is in this aspect an approximation of the feature space partial derivatives.

  One idea was that these differentials might be more obviously discriminatory than the actual feature values. This transformation is used to test this idea.

5.2 Patient 7

Patient 7 refers to the data measured at Sahlgrenska from a patient suffering from epilepsy with tonic-clonic seizures (see sec. 2.2). This patient experienced powerful tonic-clonic seizures 11 times during the 44 hours when the measurement was conducted. The seizures seem (according to the video) to be consistent. There was one more seizure confirmed by the experts from EEG data. This seizure did not, however, create movements that was recognizable as anything out of the ordinary when studied on the video. Therefore it was marked as a ”don't care”-region, i.e., a region that is excluded from training data since it belongs neither to the seizure or non seizure category.

5.2.1 Parameter evaluation results

Figure (5.2) show the results from the parameter set evaluations (to see the complete results see appendix sec. A.1).

Since there is often a trade off between sensitivity and selectivity, it is hard to objectively judge which parameter set is the most accurate.

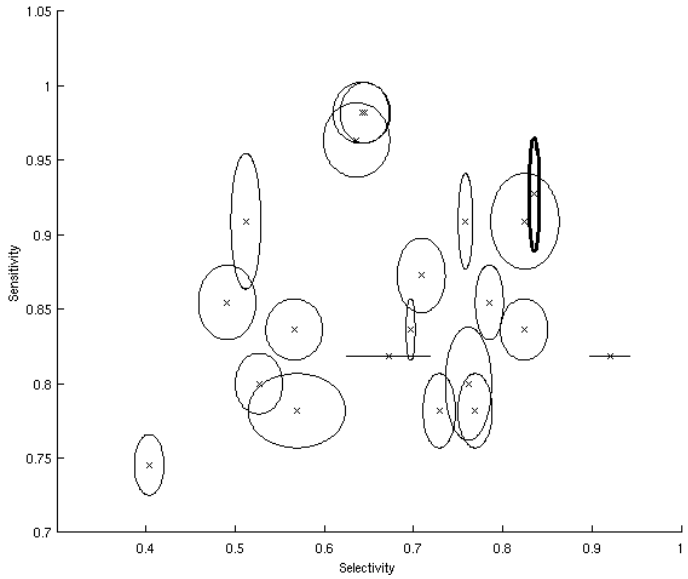

Figure 5.2: Plot of the results using all evaluated parameter sets. The x marking the means of each parameter set and the ellipses marking the standard deviation in sensitivity and selectivity directions. The standard deviation ring of the parameter set that was interpreted as the best result is shown with thick lines.

5.2.2 Best performance

How well do the proposed classification method perform on patient 7 as its best? There are two effects that need to be included in the concept of performance. As many of the

Features used:	All features				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	85%	11	11	2
	100%	85%	11	11	2
	91%	83%	11	10	2
	91%	83%	11	10	2
	82%	82%	11	9	2
Mean	92.8%	83.6%	11	10.2	2.0
Median	91.0%	83.0%	11	10.0	2.0
STD	7.5%	1.3%	0	0.8	0.0

Table 5.1: Results from the parameter set that showed overall the best results and was later used to evaluate feature sets.

seizures as possible should be found and as few non seizures should be classified as seizures. It is not obvious how to choose the best performance since some models perform better in sensitivity but worse than others in selectivity. For a real application, both of them need to be relatively good. What "relatively good" is, depends on the application.

Here, a high sensitivity are attractive since the patient might experience clouded consciousness during the seizure or be insecure about if it actually was a seizure or not. It is therefore important that as many seizures as possible are found. Having a functionality that makes it possible for the patient to mark when he/she thinks he/she have seizures might be a good complement.

A high selectivity is important for an accurate diagnosis and/or treatment evaluation. With a system that alarms, in real-time, each time it discovers a seizure, it would be possible for the patient to mark when he/she thinks that he/she experienced a seizure. Even though a patient might have a hard time knowing when he/she actually experienced a seizure he/she might be more convinced on when he/she does not have one (for example if they do some activity that causes seizure classifications). So, as long as the selectivity is quite high it might be possible to reject those false alarms on the basis of the patient's judgement. With this reasoning, a high sensitivity is more important than a high selectivity. How much more important it is, is still subjective and depends on the application.

Several parameter sets are candidates as the best parameter set depending on how the sensitivity/selectivity tradeoff is weighted. The set that was considered optimal is presented in table (5.1). This parameter set had a reasonably high and consistent selectivity while having a high sensitivity. This parameter set was later used in the evaluation of the feature sets.

Since the measurements were performed on Sahlgrenska during the EEG observation, the patient was restricted to activities close to the bed and with some restrictions on movements due to the electrodes connected to the scalp and their cords. The patient suffered from very powerful clonic-tonic seizures. The places with false alarms were probably places were the patient stood up from the bed quite fast e.g. This is not so similar to

Clustering method	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
Kmeans, 4 clusters	78.2%	5.0%	72.9%	3.6%
Kmeans, 5 clusters	78.2%	5.0%	76.9%	3.8%
Kmeans, 3 clusters	80.0%	7.6%	76.1%	5.2%
Gm hard, 4 clusters	81.8%	0.0%	67.2%	9.4%
Gm hard, 3 clusters	90.9%	9.1%	51.2%	3.3%
Gm fuzzy, 3 clusters	85.5%	5.0%	49.1%	6.4%
Gm fuzzy, 4 clusters	78.2%	5.0%	56.9%	10.9%
Kmeans seizure separated, 3 of 7 clusters	90.9%	6.4%	82.5%	7.6%
Kmeans seizure separated, 3 of 7 clusters	83.6%	4.1%	82.4%	5.3%
Kmeans seizure separated, 3 of 7 clusters	92.7%	7.6%	83.5%	1.2%
Kmeans seizure separated, 3 of 7 clusters	81.8%	0.0%	92.0%	4.5%
Kmeans seizure separated, 3 of 7 clusters	98.2%	4.1%	64.2%	6.5%
Kmeans seizure separated, 3 of 7 clusters	96.4%	5.0%	63.6%	7.4%
Kmeans seizure separated, 3 of 7 clusters	98.2%	4.1%	64.5%	5.5%
Kmeans seizure separated, 4 of 7 clusters	85.5%	5.0%	78.5%	3.1%
Max values	90.9%	6.4%	75.8%	1.6%
Max values	83.6%	4.1%	69.7%	1.0%
Min values	74.5%	4.1%	40.4%	3.3%
Min values	80.0%	4.1%	52.7%	5.3%
Max absolute values	83.6%	4.1%	56.6%	6.4%
Max absolute values	87.3%	5.0%	70.9%	5.4%

Table 5.2: Performance compared to clustering method for patient 7. For some of the methods one sees different results for the same clustering method. This is due to other changes such as the classifier or PLS-DA variance.

a seizure in actual movements but since the background data contain so little powerful movements, it might have been more similar to a seizure than to the rest of the data in the measurement.

5.2.3 Best parameter set

Which parameters played an important role in increasing the performance of the classifier? The parameter set that yielded the best performance can be seen in table 5.1. When comparing the complete evaluation data (see appendix A.1) several conclusions can be drawn. First of all the seizure separated K-means clustering methods seems to perform better than the other ones (see table 5.2).

The regular K-means clustering works quite well but has low selectivity. The non clustering extreme-value-searching methods work surprisingly well, showing a 90% sensitivity and 75% selectivity for max-value-searching with pls variance 70% and KNN with k = 3. A possible reason why the extreme-value-searching methods perform well might be due to the tranquil background data in the measurement. A real life measurement on a normally active patient during their every day life would probably show worse results for those methods.

The Gaussian mixture based clustering methods had quite good sensitivity but were not able to generate a good selectivity.

Clustering: Kmeans seiz. sep. , 3 of 7 Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
70%	91.0%	6.4%	82.6%	7.7%
80%	83.8%	4.0%	82.4%	5.3%
Clustering: Max values Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
70%	91.0%	6.4%	75.8%	1.8%
80%	83.8%	4.0%	69.4%	0.9%
Clustering: Min values Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
70%	74.8%	4.0%	40.2%	3.3%
80%	80.2%	4.0%	52.6%	5.3%
Clustering: Max absolute values Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
70%	83.8%	4.0%	56.6%	6.3%
80%	87.4%	4.9%	70.8%	5.4%

Table 5.3: Comparison between accounting for 70% or 80% in the PLS-DA dimension reduction. There are slightly smaller variances with 80% but the sensitivity is larger when accounting for only 70% of the variance.

Generally, almost all of the clustering algorithms generated good sensitivity values. The selectivity performance on the other hand were quite bad on almost all parameter setups except the seizure separated K-means with KNN classifier.

When assessing how much variance to account for in the PLS-DA dimension reduction, depends on the clustering method used. In the result evaluation for patient 7 the results can be compared in Table 5.3. These results do not show a clear advantage or disadvantage of using 70% over 80%. In the case of using seizure separated K-means it might be slightly better to account for only 70% but it is hard to draw a general conclusion, the variance seems slightly larger as well when using 70% variance.

When it comes to classification, QDA seems to yield very high sensitivity but with a tradeoff on selectivity. In cases where selectivity is not so crucial, QDA might be a good alternative. The probability weight did not seem to change the results of QDA classification, a sign of large separation between the seizure and non seizure Gaussian.

If selectivity is quite important, KNN would be regarded as a better classifier. It seemed like using 5 nearest neighbors gave better sensitivity while choosing 1 nearest neighbor gave better selectivity. Usually in literature, choosing 1 nearest neighbor will yield higher variance in the result, in this analysis however it gave more stable results than both 3 and 5 nearest neighbors. 3 nearest neighbor does not seem to have a specific advantage but the results do not show such bad results that it is possible to conclude that choosing 3 nearest

Clustering: Kmeans seiz. sep. , 3 of 7 PLS-DA variance: 80%				
Classifier	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
QDA, Prior prob. weights [10 1]	98.2	4.0	64.6	5.4
QDA, Prior prob. weights [1 10]	96.4	4.9	63.8	7.2
QDA, Prior prob. weights [1 1]	98.2	4.0	64.2	6.4
KNN, k = 1	82.0	0.0	92.0	4.5
KNN, k = 5	92.8	7.5	83.6	1.3
KNN, k = 3	83.8	4.0	82.4	5.3

Table 5.4: Comparison between classifiers on patient 7 data set.

neighbors is actually worse. All three choices give quite good results (both sensitivity and selectivity higher than 80%).

5.2.4 Feature evaluation results

The statistics of the feature group exclusions can be seen in table 5.5. This is to be compared with the results from table 5.1. There might be variations in the results due to the stochastic elements of the classification method (see chapter 4), therefore only large changes in performance can be attributed to the feature reduction.

For further information of the feature set evaluation results, see appendix sec. A.1.

Feature set	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
Without VM and MAMD	91.0 %	6.4 %	84.8 %	3.6 %
Without SMA and MAMD	96.4 %	4.9 %	83.0 %	2.4 %
Without SMA and VM	85.6 %	10.3 %	82.4 %	1.8 %
Without DC	89.2 %	7.5 %	81.8 %	1.6 %
Without CORR	89.2 %	9.9 %	86.4 %	3.5 %
Without PER	91.0 %	6.4 %	86.4 %	4.3 %
Without FREQ	91.0 %	6.4 %	72.4 %	5.9 %
Without highest FREQ	94.6 %	4.9 %	84.2 %	4.4 %
Without sensor 1	89.2 %	4.0 %	84.4 %	3.7 %
Without sensor 2	91.0 %	0.0 %	83.0 %	0.0 %
Without sensor 3	71.2 %	4.0 %	50.6 %	22.4 %
Without sensor 1 and 2	89.2 %	4.0 %	72.2 %	16.0 %
Without sensor 2 and 3	74.8 %	4.0 %	65.4 %	2.3 %
Without sensor 1 and 3	78.4 %	8.0 %	48.6 %	6.9 %
Using feature differentials	92.8 %	7.5 %	59.2 %	6.1 %

Table 5.5: Results from feature set evaluation of patient 7 using the parameter set in table 5.1.

5.2.5 Best feature set

Which feature groups were important and which were not? When looking at Table 5.5 one can clearly draw some conclusions.

The first three rows assess if the three measurements for overall activity are exchangeable. Exclusion of vector magnitude or signal magnitude area did not seem to affect the result. The measure of mean absolute magnitude difference on the other hand decreased the sensitivity significantly. Both the variance and the average of the sensitivity values were lower without MAMD, a sign that mean absolute magnitude difference actually hold discriminatory information not included in either vector magnitude or signal magnitude area.

The exclusion of DC values did not significantly affect the result. The orientation of the patient during seizures might not be as important when the seizures are powerful jerkings.

Neither correlation, periodicity of the highest frequency band (13.25-25Hz) seemed to affect the results.

Excluding all frequency bands decreased the selectivity from around 80% to 70%. A surprisingly small effect considering the seizures for patient 7 were strong tonic-clonic jerkings.

Since the evaluation was performed by removing one feature group at a time and the result did not show any dramatic decrease in performance, this implies that there are redundancy in the total feature set. However, bear in mind that the patient was mostly still on the bed and some strong movements can easily be registered in several features in such case.

Analyzing the sensors, it seems like excluding sensor 1 or 2 (chest and left arm) did not affect the results significantly. Excluding sensor 3 (right arm) on the other hand did decrease the performance considerably. By this the conclusion can be drawn that the right hand (sensor 3 mounted at the right wrist) plays an important role in discriminating between seizures.

Removing sensor 1 and 2 did decrease the selectivity to about 70%. This shows that sensor 1 and 2 hold redundant information important for increased selectivity. The redundancy being derived from the fact that removing one of them at a time did not change the result.

Removing sensor 3 together with one of the other two did reduce the performance considerably. However the sensitivity is still above 70% and the selectivity above 40%. Keeping such a good classification can probably be assessed largely to the few occasions where the patient actually performs strong movements at all during the measurements.

Finally, analyzing the feature differentials did not show a reduction in sensitivity but did show a significant reduction in selectivity to about 60%. It is not possible to completely rule out the use of feature differentials but in the current setting it was not better than using the regular feature values.

5.2.6 Old classification method evaluation

We also evaluated the old classification method, used prior to this master thesis (see sec. 1.3). The old method is very sensitive to which features that has been included. The features used for evaluation can be seen in table 5.7, which were analyzed to be the optimal features for patient 7 using the evaluation procedure from [3], see sec. B.1. The old classification method has been evaluated with the same cross validation procedure as the other results in this thesis. There are two evaluated parameter sets, the first being exactly the same parameter set as suggested by the evaluation procedure from the prior thesis [3], the second evaluation having basically the same parameters apart from the weighting of prior probabilities set much higher for non seizures than seizures. This causes a small decay in sensitivity but a large gain in selectivity (see table 5.6).

Classifier:		Old method, Prior prob. weights [1 1]			
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
Mean	82.0%	41.8%	11	9.0	12.6
Median	82.0%	41.0%	11	9.0	13.0
STD	0.0%	1.1%	0	0.0	0.5

Classifier:		Old method, Prior prob. weights [1 10^{15}]			
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
Mean	78.4%	63.2%	11	8.6	5.0
Median	82.0%	64.0%	11	9.0	5.0
STD	4.9%	1.1%	0	0.5	0.0

Table 5.6: Results using the old classification method that existed prior to this thesis.

5.2.7 Performance of old method

The old classification method used prior to this thesis did not incorporate the evolution in time when classifying seizures. For this thesis, the evaluation procedure used on the proposed classification method was also used for evaluating the old classification method. The reason for this was to make a direct comparison between the two methods. The results from the old classification method can be seen in table 5.6.

Using the exact model that yielded optimal results by the prior evaluation procedure, see appendix B.1, the old classification model got a very stable result with sensitivity of 82% and selectivity of about 41%. Since the selectivity was inadequate, the prior probability

Feat #	Feature name
24	Frequency band 3.75 - 5.25 Hz, Sensor 1
3	DC, Z axis, Sensor 1
1	DC, X axis, Sensor 1
5	DC, Y axis, Sensor 2
40	Linear correlation, Sensors 1 & 2

Table 5.7: The optimal feature set for the old method according to sec. B.1

weight was changed to reduce type II errors. The results with weights of [1 10^{15}] increased the selectivity to 63% while reducing the sensitivity to 78%. However, with this change in prior weights the stability was reduced and the variance of the sensitivity was increased from 0 to 4.9%.

5.3 Patient 14

Patient 14 refers to the data measured on a person suffering from epilepsy with tonic seizures. The measurement was performed on Sahlgrenska, in the setting described in sec. 2.2. The patient experiencing subtle tonic seizures characterized by holding a "fencing position" (see sec.1.2) for a couple of seconds and then returning to normal. For a few of the seizures, small vibrations are visible in the videos.

This data set includes 22 seizures during a time span of about 6 days.

5.3.1 Parameter evaluation results

Figure 5.3 show the results of all parameter set evaluations (to see more details see appendix sec. A.2).

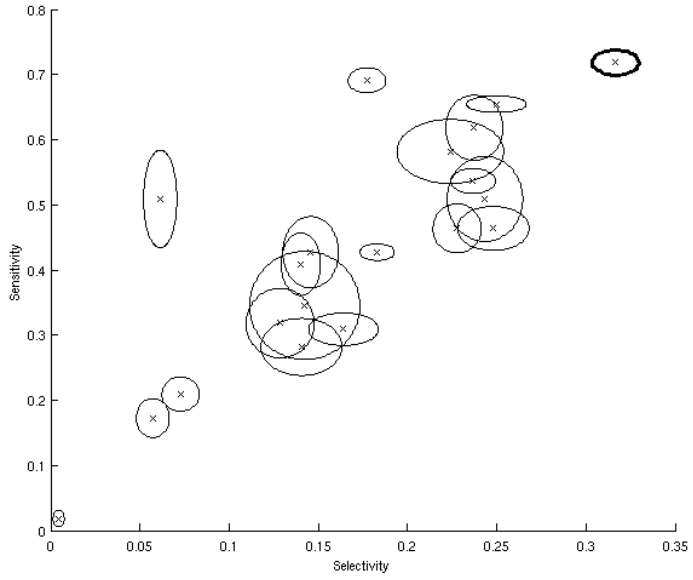

Figure 5.3: Plot of the results for patient 14 using all evaluated parameter sets.

There is one parameter set that perform clearly better than the others. This one can be seen in table 5.8

5.3.2 Best performance

With this patient there was one parameter set superior in both sensitivity and selectivity to the others (see table 5.8). However, even for that parameter set the result was unsatisfactory, only about 70% of the seizures were found and the selectivity was only around 30%.

Features used:	All features				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	77%	33%	22	17	34
	73%	33%	22	16	33
	73%	28%	22	16	42
	68%	34%	22	15	29
	68%	31%	22	15	34
Mean	71.8%	31.8%	22	15.8	34.4
Median	73.0%	33.0%	22	16.0	34.0
STD	3.8%	2.4%	0	0.8	4.7

Table 5.8: Results from the parameter set that showed overall best results and was later used to evaluate the feature sets.

The performance on this data set is not adequate for a real life application. The patient studied did have a large amount of seizures, which all resulted in the patient assuming a fencing position. It was quite obvious from the videos when the patient experienced a seizure but these kinds of seizures do not have an original acceleration pattern. Assuming the fencing position (which was assumed in a moderate arm speed) is not much different from stretching the arms for any other reason. Once the fencing position was assumed, no further acceleration was performed for any of the sensors. When regarding merely the acceleration at wrists and chest, there are no real difference between sitting still and stretching your arms as compared to having a seizure. The conclusion drawn is that these kinds of seizures are probably not possible to classify correctly by this proposed classification method. If they are possible at all using only accelerometers are left unsaid.

However, any device able to identify the orientation of the limbs would be able to find these kinds of seizures. If using accelerometers, these together with gyroscopes and/or magnetometers might make it possible to derive orientation of the limbs at all times by integration in time. This might however put a considerable demand for high resolutions on the sensors as well as on the numerical integration method. Another solution would be to attach cords to the pivot points of the limbs. These cords could change resistance when bended. This would make it possible to know the orientation of the limbs at all times.

5.3.3 Best parameter set

The parameter set that yielded the best performance can be seen in table 5.8. A comparison between the clustering methods and the results from the evaluation can be seen in table 5.9).

Basically the extreme-value-searching non clustering methods show the worst results. These methods give selectivity results below 10% and sensitivity results at most 20%. The maximum-absolute-value-searching method have a performance comparable to randomly

Clustering method	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
Kmeans, 4 clusters	61.8%	10.0%	23.7%	3.2%
Kmeans, 4 clusters	58.2%	9.9%	22.4%	6.0%
Kmeans, 5 clusters	50.9%	13.0%	24.3%	4.2%
Kmeans, 5 clusters	53.6%	3.8%	23.7%	2.5%
Kmeans, 5 clusters	65.5%	2.5%	25.0%	3.3%
Kmeans, 5 clusters	71.8%	3.8%	31.7%	2.6%
Kmeans, 5 clusters	69.1%	3.8%	17.7%	2.1%
Kmeans, 6 clusters	46.4%	7.5%	22.8%	2.7%
Kmeans, 6 clusters	46.4%	6.7%	24.8%	4.0%
Gm fuzzy, 5 clusters	30.9%	5.0%	16.4%	3.9%
Gm fuzzy, 4 clusters	34.5%	16.6%	14.2%	6.2%
Gm fuzzy, 4 clusters	42.7%	2.5%	18.3%	1.9%
Gm fuzzy, 4 clusters	28.2%	8.7%	14.1%	4.5%
Gm fuzzy, 3 clusters	42.7%	10.9%	14.6%	3.1%
Gm fuzzy, 3 clusters	40.9%	9.6%	14.0%	2.2%
Gm hard, 4 clusters	31.8%	10.7%	12.9%	3.8%
Kmeans seizure separated, 3 of 7 clusters	50.9%	14.9%	6.1%	1.8%
Max value	20.9%	5.2%	7.3%	2.1%
Min value	17.3%	5.9%	5.7%	1.8%
Max absolute value	1.8%	2.5%	0.4%	0.6%

Table 5.9: Performance compared to clustering method for patient 14.

placing the seizure classifications anywhere in the time span.

The seizure separated K-means method shows a surprisingly bad performance as well. One interpretation for this is that there are no discriminatory information caught in the clusters from only one of the seizures that can be generalized to all seizures present.

The K-means and Gaussian mixture based clustering methods seem to yield similar sensitivity values but the K-means algorithm has slightly larger selectivity. One explanation for this is that K-means clustering is able to search for more clusters since the stability of the covariance matrix introduces numerical problems for Gaussian mixture clustering with too few observations compared so number of clusters.

Here the explanation why the K-means method performs slightly better than the other methods is that it might find information similar between a certain subgroup of the seizures but that do not generalize to all of the seizures. Since it can have more clusters than the Gaussian mixture based method it might find more behaviors to use for classification. In this way it will have found clusters that together find all of the seizures. The seizure separated method did not find these since it was based on the assumption that all seizures have something in common.

Assessing the variance threshold for PLS-DA, the same models have been evaluated but with 70% and 80% variance thresholds. The result can be seen in table 5.10. The amount of PLS-DA variance is highly dependent on the clustering method used. As can be seen in the table, using 80% instead of 70% PLS-DA variance for the K-means algorithm does seem to increase the variance of the results, especially for selectivity, where it is doubled in all three cases. However the changes are so small and tested with so few evaluations that it is not an obvious conclusion.

Clustering: Kmeans, 4 clusters Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
70%	61.8	9.9	23.8	3.0
80%	58.2	10.1	22.6	6.0
Clustering: Kmeans, 5 clusters Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
70%	53.8	3.8	23.8	2.7
80%	51.0	12.9	24.2	4.4
Clustering: Kmeans, 6 clusters Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
70%	46.2	7.4	22.8	2.7
80%	46.2	7.0	24.8	4.1
Clustering: GM fuzzy, 3 clusters Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
70%	42.6	11.2	14.6	3.0
80%	41.0	9.7	14.0	2.3
Clustering: GM fuzzy, 4 clusters Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
70%	28.2	8.8	14.0	4.7
80%	42.6	2.2	18.4	1.7

Table 5.10: Comparison between accounting for different amounts of variance in the PLS-DA dimension reduction for patient 14.

Clustering: K-means, 5 clusters PLS-DA variance: 70%				
Classifier	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
KNN, k = 3	53.8	3.8	23.8	2.7
QDA, Prior prob. weight [1 1]	65.6	2.2	25.2	3.5
QDA, Prior prob. weight [1 10]	71.8	3.8	31.8	2.4
QDA, Prior prob. weight [10 1]	69.2	3.8	17.8	2.3
Clustering: GM fuzzy, 4 clusters PLS-DA variance: 70%				
Classifier	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
KNN, k = 3	34.6	16.7	14.4	6.3
KNN, k = 1	28.2	8.8	14.0	4.7

Table 5.11: Comparison between classifiers on patient 14 data set.

Using 80% instead of 70% on the Gaussian mixture clustering yields no significant difference when using only 3 clusters. Using 4 clusters however increases performance dramatically. Both averages increases while the variances decrease.

Studying the performance of different classifier parameters, see table 5.11. Here QDA has a definite advantage over KNN. An explanation for this might be that, since the changes comparing a seizure to a non seizure are so small, the KNN method might be too unstable to be useful. Using a parametric model with relatively few parameters might yield slightly more stable classifications.

5.3.4 Feature evaluation results

The statistics of the evaluations when excluding certain feature groups can be seen in table 5.12. This is to be compared with the results from table 5.8.

For further information on the feature set evaluation result see appendix sec. A.2.

5.3.5 Best feature set

From table 5.12 some remarkable discoveries can be extracted. First of all, removing sensor 3 did increase selectivity from around 30% to 40% but doubled the variance. This could be explained by at least one but possibly two of the following causes. Sensor 3 might be unimportant for classification of patient 14. The curse of dimensionality (see sec. 3.2.2) might influence this model heavily. By reducing the dimensionality a better classification is possible.

If the curse of dimensionality plays an important role, one reason for the bad performance on this patient might actually be that the dimensions are not reduced that well. This in turn would indicate no linear subspace which holds the most of the discriminatory power.

Feature set	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
Without VM and MAMD	64.4 %	8.0 %	27.8 %	2.5 %
Without SMA and MAMD	70.8 %	5.9 %	34.0 %	4.2 %
Without SMA and VM	75.4 %	2.2 %	32.2 %	3.3 %
Without DC	45.4 %	7.4 %	21.0 %	7.9 %
Without CORR	63.8 %	8.6 %	33.6 %	3.5 %
Without PER	64.8 %	5.1 %	25.0 %	3.1 %
Without FREQ	51.0 %	6.7 %	31.2 %	5.4 %
Without highest FREQ	68.4 %	7.8 %	28.0 %	4.8 %
Without sensor 1	52.8 %	5.4 %	18.2 %	4.5 %
Without sensor 2	29.0 %	5.0 %	16.6 %	2.9 %
Without sensor 3	74.6 %	2.2 %	40.4 %	4.5 %
Without sensor 1 and 2	70.0 %	10.3 %	13.2 %	0.8 %
Without sensor 2 and 3	57.4 %	2.2 %	27.2 %	1.1 %
Without sensor 1 and 3	48.2 %	9.6 %	23.6 %	1.3 %
Using feature differentials	96.0 %	2.2 %	38.2 %	7.2 %

Table 5.12: Results from feature set evaluation of patient 14.

Secondly, the feature differentials show a very high sensitivity and with a slight increase in selectivity as well. Here it certainly seems like the differentials between neighbouring observations in time do hold more discriminatory information than the original features, or alternatively that the discriminatory information are compressed into smaller subspaces easier when using feature differentials (therefore reducing the influence of the curse of dimensionality). It is quite extraordinary that the feature differentials perform so much better than all the others but the selectivity is still unsatisfactory, so no real use of the proposed classification method is expected on patients with similar seizures as patient 14.

5.3.6 Old method evaluation

The features used for evaluation of the old method can be seen in table 5.14. This feature set was analyzed to be the optimal feature set for patient 14 using the evaluation procedure from [3], see sec. B.2. There are two evaluated parameter sets, the first being exactly the same as suggested by the prior thesis [3], the second evaluation having the same parameters apart from the weighting of prior probabilities set to twice as high for non seizures compared to seizures and with the morphological open radius increased by 10. This causes a small decay in sensitivity but a large gain in selectivity (see table 5.13).

5.3.7 Performance of old method

The results from the old classification method evaluation for patient 14 can be seen in table 5.13.

The old method have a high sensitivity and a very low selectivity. When increasing the probability weight for non seizures by 2 and increasing the morphological open radius

Classifier:	Old method, Prior prob. weights [1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
Mean	82.0%	4.0%	22	18.0	464.2
Median	82.0%	4.0%	22	18.0	463.0
STD	0.0%	0.0%	0	0.0	12.0

Classifier:	Old method, Prior prob. weights [1 2] open radius 10				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
Mean	74.8%	11.0%	22	16.4	132.8
Median	73.0%	11.0%	22	16.0	133.0
STD	4.0%	0.7%	0	0.9	4.4

Table 5.13: Results using the old classification method that existed prior to this thesis.

Feat #	Feature name
11	SMA, Sensor 2
2	DC, Y axis, Sensor 1
45	Circular correlation, Sensors 2 & 3
9	DC, Z axis, Sensor 3
6	DC, Z axis, Sensor 2
19	Periodicity, Sensor 1
7	DC, X axis, Sensor 3
15	VM, Sensor 3
44	Linear correlation, Sensors 2 & 3

Table 5.14: The optimal feature set for the old method according to sec. B.2

by 10, the selectivity increases substantially. The sensitivity does however decrease from 82% to 75%. On this patient, the old method performs similarly to the new one. Both classifiers show classification results that are good enough to be non random but not so good that they could be of any use in a real application.

5.4 Patient F1

This F1 data set are measured on two subjects not actually experiencing authentic epileptic seizures. Instead they have been instructed to enact a sequence of movements at some occasions and logging the exact times of these ”seizures”. The ”seizures” are very strong movements and should be regarded as more extreme than the average real epileptic seizures. The background data, i.e. the movements not associated with ”seizures”, are also quite extreme. Activities such as running, climbing and walking around in the woods are included as well as normal every day chores. The data set includes 28 seizures over a time span of 50 hours.

This data set is mainly included to assess the performance of the proposed classification method in a quite extreme background environment. It is also interesting since the data set is actually a fusion of four different measurements by two different persons, performed during different times. Therefore, it will give a hint on the possibility of transferring seizure models between patients suffering from similar forms of epilepsy.

Seizure sequence

The seizure sequence consists of indian jumps followed by swinging of arms from side to side and finally rotating the arms around a horizontal axis. The seizures lasts for about 20-30 seconds.

5.4.1 Parameter evaluation results

Figure 5.4 show the result of all parameter set evaluations (to see more details see appendix sec. A.3).

Features used:	All features				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	90%	28	27	3
	96%	90%	28	27	3
	96%	84%	28	27	5
	96%	93%	28	27	2
	96%	93%	28	27	2
Mean	96.0%	90.0%	28	27.0	3.0
Median	96.0%	90.0%	28	27.0	3.0
STD	0.0%	3.7%	0	0.0	1.2

Table 5.15: Results of the parameter set that showed overall best results and was later used to evaluate the feature sets.

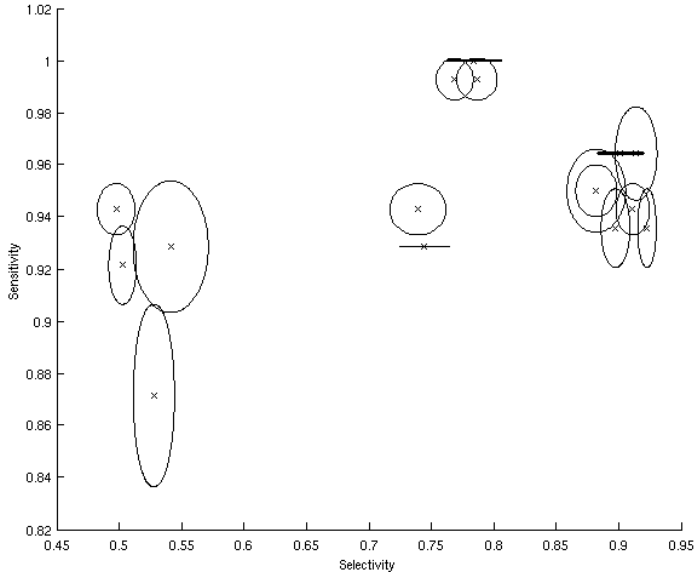

Figure 5.4: Plot of the results using all evaluated parameter sets.

5.4.2 Best performance

For this data set, many parameter sets generated good results. The overall best model can be seen in table 5.15. This is a good and very convincing result that shows two things: First of all, the proposed method is able to identify various reasonably distinct movements against an active background. Secondly; seizure behavior can be generalized over different persons, as long as the movements are similar.

5.4.3 Best parameter set

The parameter set that yielded the best performance can be seen in table 5.15. A comparison between the clustering methods and the results from the evaluation can be seen in table 5.16. It can be seen that the K-means clustering algorithm reaches highest performance among the clustering methods evaluated. The Gaussian mixture based methods show very low selectivity even though they are able to find the real seizures comparably well. It is not yet understood why the Gaussian mixture methods obtain so low selectivity. For some reason they seem unable to filter out the non seizures (remember that the background data are very active in this data set).

The maximum- and minimum-searching, non clustering methods perform very well on this data set as well. They obtain a sensitivity higher than 90% while having selectivity of about 75%. This result is so good that the extreme-value-searching methods might be useful in applications in the future. They do not rely on the clustering of data which have computational advantages as well as a greater transparency for users to assess their behavior. Compared to the clustering methods the extreme-value-searching methods look for obvious patterns and have no random elements which cause less variance in the results.

Clustering method	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
Kmeans, 3 clusters	96.4%	0.0%	90.1%	3.6%
Kmeans, 4 clusters	93.6%	3.0%	92.3%	1.5%
Kmeans, 4 clusters	96.4%	3.6%	91.4%	3.3%
Kmeans, 4 clusters	95.0%	3.2%	88.2%	4.7%
Kmeans, 4 clusters	93.6%	3.0%	89.7%	2.3%
Kmeans, 4 clusters	95.0%	2.0%	88.2%	3.3%
Kmeans, 4 clusters	100.0%	0.0%	78.4%	4.3%
Kmeans, 4 clusters	99.3%	1.6%	76.9%	2.9%
Kmeans, 4 clusters	99.3%	1.6%	78.6%	3.2%
Kmeans, 5 clusters	94.3%	2.0%	91.1%	2.7%
Gm fuzzy, 3 clusters	92.1%	3.0%	50.3%	2.2%
Gm fuzzy, 4 clusters	87.1%	7.0%	52.8%	3.3%
Gm hard, 3 clusters	94.3%	2.0%	49.8%	3.0%
Gm hard, 4 clusters	92.9%	5.1%	54.1%	6.0%
Max value	92.9%	0.0%	74.4%	3.9%
Min value	94.3%	2.0%	73.9%	4.5%

Table 5.16: Performance compared to clustering method for patient F1.

Clustering: K-means, 4 clusters Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
50%	96.4	3.5	91.4	3.1
70%	93.4	2.9	92.4	1.3
80%	94.6	3.1	88.0	4.8

Table 5.17: Comparison between accounting for different amounts of variance in the PLS-DA dimension reduction for patient F1.

Clustering: K-means, 4 clusters PLS-DA variance: 70%				
Classifier	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
KNN, k = 1	93.4	2.9	89.8	2.2
KNN, k = 3	93.4	2.9	92.4	1.3
KNN, k = 5	94.8	1.6	88.2	3.4
QDA, Prior prob. weights [1 1]	100.0	0.0	78.4	4.1
QDA, Prior prob. weights [10 1]	99.2	1.8	77.0	2.8
QDA, Prior prob. weights [1 10]	99.2	1.8	78.6	3.0

Table 5.18: Comparison between classifiers on patient F1 data set.

Only one model was evaluated with different amounts of variance in the PLS-DA step. The result of this evaluation can be seen in table 5.17. The differences are not so significant that one can say that one of them works better than the others. It does not, however, seem like reducing the dimensions to only account for 50% of the variance decreases performance, which means that the discriminatory information exist in a subspace with only half of the PLS-DA variance of the original space.

A comparison between classifiers is shown in table 5.18. The QDA classifiers generate lower selectivity values than the K-nearest neighbor classifiers but with the tradeoff that they have a slightly higher sensitivity.

No real difference can be seen depending on the prior probability weight for the QDA classifiers. This indicates that the Gaussian are well separated (see sec. 3.2.1).

For the KNN classifiers, no real difference in performance can be seen depending on choosing 1, 3 or 5 nearest neighbors.

5.4.4 Feature evaluation results

The statistics of the feature set combinations can be seen in table 5.19. This is to compare with the results from table 5.15.

For further information of the feature set evaluation result see appendix sec. A.3.

5.4.5 Best feature set

Which features hold discriminatory information? Results can be viewed in table 5.19. The general activity measurements signal magnitude area, vector magnitude and mean absolute magnitude difference seem to hold similar information. No apparent decrease in performance is visible when excluding two of them at a time.

When removing the DC features a slight decrease in selectivity is noticed. If this is due to discriminatory power of the DC values or simply due to randomness is hard to say. In any case, there is no big drawback in performance when removing the DC values. The same

Feature set	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
Without VM and MAMD	96.0 %	0.0 %	91.2 %	2.7 %
Without SMA and MAMD	100.0 %	0.0 %	90.8 %	2.2 %
Without SMA and VM	95.4 %	1.3 %	90.6 %	1.3 %
Without DC	96.8 %	1.8 %	87.8 %	2.5 %
Without CORR	96.0 %	0.0 %	92.4 %	1.3 %
Without PER	96.0 %	0.0 %	91.8 %	3.4 %
Without FREQ	96.8 %	1.8 %	86.0 %	1.4 %
Without highest FREQ	96.0 %	0.0 %	93.0 %	0.0 %
Without sensor 1	96.2 %	2.5 %	88.8 %	2.7 %
Without sensor 2	94.2 %	1.6 %	91.8 %	1.6 %
Without sensor 3	100.0 %	0.0 %	93.0 %	0.0 %
Without sensor 1 and 2	96.0 %	0.0 %	84.2 %	1.8 %
Without sensor 2 and 3	93.6 %	1.3 %	89.4 %	1.3 %
Without sensor 1 and 3	96.2 %	2.5 %	88.8 %	1.6 %
Using feature differentials	89.2 %	4.1 %	68.0 %	4.2 %

Table 5.19: Results from feature set evaluation of patient F1.

is true when removing the frequency components. Since the frequency bands and DC values are the groups with most components and two groups that seems very typical for seizures, this result is quite interesting. However, these groups might contain information present in some other feature groups as well (at least for the seizure behavior present in data set F1).

Excluding correlation, periodicity and the highest frequency band (13.25-25Hz), one at a time, does not affect the classification performance.

Removing one of the sensors does not render a significant decrease in classification performance. Removing sensor 1 and 2 simultaneously does, however, decrease the selectivity a little bit to about 84%.

Using feature differentials decreases selectivity to only 68% while decreasing the sensitivity to about 90%. This is the only significant decrease in performance noticed on all of the feature sets analyzed (and feature differentials are not really a group of features but a transformation of the original features).

It seems like all groups tested have redundant information with the complement feature set. Therefore, no real conclusion can be drawn from this analysis except that feature differentials do not perform well in this setting and that probably a large part of the features can be removed. For this kind of "seizures", it would work to only use one sensor, even with diverse background activity.

5.4.6 Old method evaluation

The features used for evaluation of the old method can be seen in table 5.21. These features were analyzed to be the optimal feature set set for patient F1 using the evaluation procedure from [3], see sec. B.3. There are two evaluated parameter sets, the first being exactly the same parameter set as suggested by the prior thesis [3], the second evaluated

with the same parameters apart from the weighting of prior probabilities set much higher for non seizures compared to seizures. This causes a small decay in sensitivity but a large gain in selectivity (see table 5.20).

Classifier:	Old method with prior prob. weight [1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
Mean	93.0%	42.4%	28	26.0	35.4
Median	93.0%	43.0%	28	26.0	35.0
STD	0.0%	1.5%	0	0.0	2.3

Classifier:	Old method with prior prob. weight [1 10^{20}]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
Mean	86.6%	71.0%	28	24.2	10.0
Median	86.0%	71.0%	28	24.0	10.0
STD	1.3%	0.0%	0	0.4	0.0

Table 5.20: Results using the old classification method that existed prior to this thesis.

5.4.7 Performance of old method

The results from the old method evaluation for patient F1 can be seen in table 5.20.

With the original model used prior to this thesis, the average sensitivity is 93% with a selectivity of 42.4%. Increasing the prior probability weight from [1 1] to [1 10^{20}], the sensitivity decreases to 86.6% but with a drastic increase in selectivity up to 71%. These values are almost good enough to be used in an application (at least if the selectivity is not that important, such as if the patient are able to reject false alarms when using the classifier hardware). The huge amount of change in prior probability weights that was needed to change the result show that the Gaussian are well separated.

Feat #	Feature name
29	Frequency band 2.25 - 3.75 Hz, Sensor 2
47	Circular correlation, Sensor 1 with itself
46	Linear correlation, Sensors 1 with itself & 3
5	DC, Y axis, Sensor 2
3	DC, Z axis, Sensor 1
8	DC, Y axis, Sensor 3
20	Periodicity, Sensor 3
40	Linear correlation, Sensor 1 % 2
6	DC, Z axis, Sensor 2
48	Linear correlation, Sensor 2 with itself
49	Circular correlation, Sensor 2 with itself
43	Circular correlation, Sensors 1 & 3
50	Linear correlation, Sensor 3 with itself
45	Circular correlation, Sensors 2 & 3
9	DC, Z axis, Sensor 3
42	Linear correlation, Sensors 1 & 3

Table 5.21: The optimal feature set for the old method according to sec. B.3

5.5 Patient F2

This data referred to as patient F2 was measured on a person not actually suffering from epilepsy. Instead the subject was instructed to carry out a rehearsed seizure sequence. In this set the ”seizures” are supposed to look like real epileptic seizures of quite subtle nature (described below). This data can therefore be thought of as substituting an authentic measurement, but with consistent and reasonably numerous seizures and authentic every day background activity. The data set includes 17 ”seizures” over a time span of about 35 hours.

Seizure sequence

The seizure sequence starts out with the actor assuming a fencing position with right hand stretched and left hand tucked behind the head. The head directed towards the right hand. From this fencing position, the ”seizure” evolves by introducing small jerkings with decreasing frequency and increasing amplitude in the right hand. The whole seizure last for about 15-20 seconds.

5.5.1 Parameter evaluation results

Figure 5.5 show the results from all parameter set evaluations (to see more details see appendix sec. A.4).

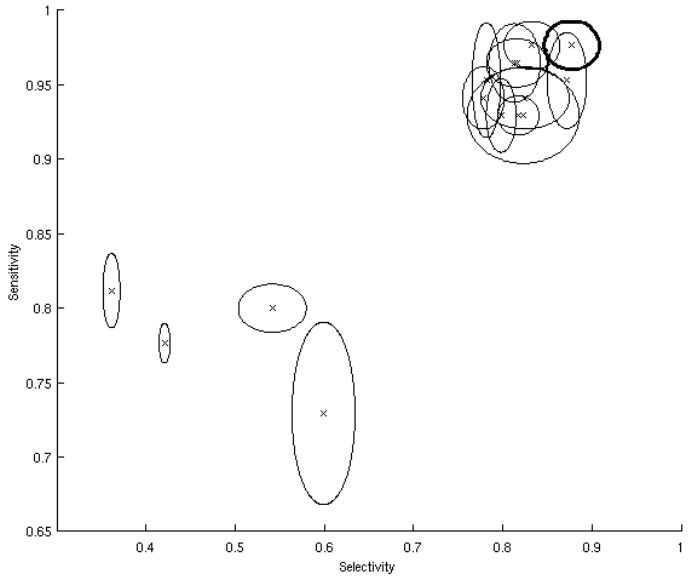

Figure 5.5: Plot of the results using all evaluated parameter sets.

Features used:	All features				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	94%	17	17	1
	100%	85%	17	17	3
	100%	81%	17	17	4
	94%	94%	17	16	1
	94%	84%	17	16	3
Mean	97.6%	87.6%	17	16.6	2.4
Median	100.0%	85.0%	17	17.0	3.0
STD	3.3%	6.0%	0	0.5	1.3

Table 5.22: Results of the parameter set that showed overall best results and was later used to evaluate the feature sets.

Clustering method	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
Gm fuzzy, 3 clusters	97.6%	3.2%	87.7%	6.2%
Gm fuzzy, 4 clusters	95.3%	7.7%	78.2%	3.2%
Gm fuzzy, 4 clusters	94.1%	4.2%	82.5%	9.9%
Gm fuzzy, 4 clusters	92.9%	6.4%	82.3%	12.5%
Gm fuzzy, 4 clusters	92.9%	4.9%	79.8%	3.3%
Gm fuzzy, 4 clusters	92.9%	2.6%	81.7%	4.6%
Gm fuzzy, 4 clusters	96.5%	3.2%	81.6%	6.8%
Gm fuzzy, 4 clusters	94.1%	4.2%	77.8%	4.7%
Gm fuzzy, 4 clusters	95.3%	6.4%	87.2%	4.4%
Gm hard, 3 clusters	97.6%	3.2%	83.3%	6.3%
Gm hard, 4 clusters	96.5%	5.3%	81.3%	5.8%
K means, 3 clusters	80.0%	3.2%	54.2%	7.6%
K means, 5 clusters	72.9%	12.2%	59.9%	7.0%
Max values	77.6%	2.6%	42.1%	1.2%
Min values	81.2%	4.9%	36.1%	1.9%

Table 5.23: Performance compared to clustering method for patient F2.

5.5.2 Best performance

How well do the proposed method perform on patient F2 at its best? The figure 5.5 answer this question. For this data set, several models generated good results. The overall best model can be seen in table 5.22. This result shows that the task of classifying quite subtle tonic-clonic seizures with adequate precision in a realistic background setting is possible.

5.5.3 Best parameter set

The parameter set that yielded the best performance can be seen in table 5.22. A comparison between the clustering methods and the results from the evaluation can be seen in table 5.23.

Clustering: GM fuzzy, 4 clusters Classification: KNN, k = 3				
Variance	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
50%	92.8	6.6	82.2	12.2
70%	95.2	7.8	78.0	3.1
80%	94.0	4.2	82.4	10.0

Table 5.24: Comparison between accounting for different amounts of variance in the PLS-DA dimension reduction for patient F2.

Clustering: GM fuzzy, 4 clusters PLS-DA variance: 70%				
Classifier	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
KNN, k = 1	92.8	2.7	81.6	4.6
KNN, k = 3	95.2	7.8	78.0	3.1
KNN, k = 5	92.8	5.0	79.8	3.3
QDA, Prior prob. weights [1 1]	96.4	3.3	81.6	6.8
QDA, Prior prob. weights [10 1]	94.0	4.2	77.8	4.5
QDA, Prior prob. weights [1 10]	95.2	6.6	87.0	4.3

Table 5.25: Comparison between classifiers on patient F2 data set.

Table 5.23 shows that only the Gaussian mixture based clustering methods are able to yield adequate selectivity. Probably, the shape adaptivity of the Gaussian mixture methods is needed to fit the data satisfactorily in order to acquire the discriminatory information.

As seen in table 5.24, there is no big difference in performance when accounting for 50%, 70% or 80% of the variance in the PLS-DA dimension reduction. However, both 50% and 80% PLS-DA variance causes significantly larger variance among the selectivity values, as compared to using 70% variance. Therefore, using 70% of the PLS-DA variance would be recommended for stability reasons.

The K-nearest neighbor method seems to perform equally well when choosing k = 1, 3 or 5. There is no significant difference between them, K = 3 having a slightly larger sensitivity value than the others but with greater variance for the sensitivity as well. No real difference can be concluded.

The QDA classifiers seem to work at least as well as the KNN and weighting the prior probability for non seizures to be 10 times more likely than a seizure increases the selectivity from about 80% to 87% without changing the sensitivity significantly.

5.5.4 Feature evaluation results

The statistics of the feature set combinations can be seen in table 5.26. This is to compare with the results from table 5.22.

For further information on the feature set evaluation result see appendix sec. A.4.

Feature set	Sensitivity		Selectivity	
	Mean	STD	Mean	STD
Without VM and MAMD	90.4 %	6.8 %	80.2 %	9.4 %
Without SMA and MAMD	97.6 %	3.3 %	78.4 %	3.2 %
Without SMA and VM	96.4 %	5.4 %	83.6 %	3.6 %
Without DC	89.2 %	6.6 %	60.8 %	5.4 %
Without CORR	94.0 %	4.2 %	83.4 %	4.7 %
Without PER	94.0 %	4.2 %	80.0 %	5.3 %
Without FREQ	89.2 %	6.6 %	64.8 %	2.5 %
Without highest FREQ	92.8 %	5.0 %	88.0 %	6.5 %
Without sensor 1	95.2 %	5.0 %	71.4 %	9.2 %
Without sensor 2	85.6 %	3.3 %	63.2 %	5.0 %
Without sensor 3	88.0 %	7.3 %	52.2 %	4.6 %
Without sensor 1 and 2	91.6 %	5.4 %	51.2 %	4.1 %
Without sensor 2 and 3	88.0 %	7.3 %	40.8 %	3.3 %
Without sensor 1 and 3	90.4 %	5.4 %	46.4 %	4.6 %
Using feature differentials	88.0 %	4.2 %	60.4 %	4.8 %

Table 5.26: Results from feature set evaluation of patient F2.

5.5.5 Best feature set

The results when excluding certain feature groups can be viewed in table 5.26. For this patient, the general activity measurements signal magnitude area, vector magnitude and mean absolute magnitude difference do seem to hold their own contribution to the discriminatory power of the feature set. Excluding two of them will decrease the classification result somewhat. The mean absolute magnitude difference seems to have the most of the information since only using this measure will decrease the result the least.

Removing the DC features will affect the sensitivity somewhat and the selectivity deeply. The selectivity decreases to only 60% which renders the model unusable for most applications. Similarly removing the frequency components seems to affect the discriminatory power equally.

Removing correlation, periodicity or the highest frequency band (13.25-25Hz), one at a time, will reduce performance somewhat but will still yield satisfactory results.

Removing sensor 1 will reduce selectivity to 70% while keeping the sensitivity at around 95%. Removing sensor 2 or sensor 3 will decrease the performance much more severely.

Removing two sensors at a time will render results of about 90% sensitivity but only 45% selectivity.

Using feature differentials yield a sensitivity in the order of about 90% but with a selectivity of 60%, below what could be a satisfactory performance in the most implementations.

5.5.6 Old method evaluation

The features used for evaluation of the old method can be seen in table 5.28. These features were analyzed to be the optimal feature set for patient F2 using the evaluation procedure from [3], see sec. B.4. There are two evaluated parameter sets. The first being exactly the same parameter set as suggested by the prior thesis [3], the second evaluation having the same parameters apart from the weighting of prior probabilities set much higher for non seizures than seizures. This causes a increase in both sensitivity and selectivity (see table 5.27).

Classifier:	Old method, Prior prob. weight [1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
Mean	76.2%	12.0%	17	13.0	94.2
Median	76.0%	12.0%	17	13.0	93.0
STD	3.9%	0.7%	0	0.7	3.0

Classifier:	Old method, Prior prob. weight [1 10^{20}]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
Mean	84.4%	37.8%	17	14.4	23.6
Median	82.0%	38.0%	17	14.0	23.0
STD	3.3%	1.1%	0	0.5	0.9

Table 5.27: Results using the old classification method that existed prior to this thesis.

Feat #	Feature name
19	Periodicity, Sensor 1
36	Frequency band 3.75 - 5.25 Hz, Sensor 3
30	Frequency band 3.75 - 5.25 Hz, Sensor 2
45	Circular correlation, Sensors 2 & 3
37	Frequency band 5.25 - 8.25 Hz, Sensor 3
14	VM, Sensor 2
40	Linear correlation, Sensors 1 & 2
35	Frequency band 2.25 - 3.75 Hz, Sensor 3
38	Frequency band 8.25 - 13.25 Hz, Sensor 3
5	DC, Y axis, Sensor 2
46	Linear correlation, Sensor 1 with itself
7	DC, X axis, Sensor 3
49	Circular correlation, Sensors 2 with itself
29	Frequency band 2.25 - 3.75 Hz, Sensor 2
47	Circular correlation, Sensor 1 with itself

Table 5.28: The optimal feature set for the old method according to sec. B.3

5.5.7 Performance of old method

The results from the old classification method evaluation for patient F2 can be seen in table 5.20.

As can be seen in the table there are sensitivity values that are quite satisfactory but the selectivity is very low. When increasing the prior probability weights for the non seizures, it is possible to increase the performance of the method but the selectivity does never go above 40%, to low to be usable in a real implementation.

Chapter 6

Discussion

In this chapter, interpretation and conclusions are drawn from the results in chapter 5. The chapter first make a general interpretation of the results from the four data sets studied in chapter 5. Later, a section addressing future work and possible alternatives are presented. Finally, a section that draws general conclusions and compares them to the questions and tasks of the thesis (see sec. 1.4) are presented.

6.1 General interpretation

The proposed method has been evaluated on four different data sets consisting of measurements with four different seizure behaviors as well as some differences in background activity. All four data sets do however have something in common: They all share the properties that the seizures are consistent and numerous (at least 10).

Exactly how dependent the relation between number of seizures and classification performance is, is hard to subjectively assess. It is however believed that more than 10 seizures with the same behavior pattern should be present to fit a reasonable classification model.

6.1.1 Best performance

For patient 14 the proposed classification method did not seem to perform satisfactorily. This shows that those types of tonic seizures are not expected to be classified correctly by the proposed classification method. This reduces the possible market for such an automatic seizure logger somewhat. However, epilepsy is a common disorder and there is a large group of people suffering from seizures which this classification method shows promising results for.

The 3 other data sets have shown satisfactory results, both sensitivity and selectivity above 80% for all of them. From this the conclusions is that: the kinds of seizures exhibited by patient 7 and patient F2 are possible to classify with satisfactory results. Also, patient

F1 showed that powerful movements can be classified even with a background of high activity and that movements can be generalized between different persons. This last conclusion is quite important, since it makes it possible to increase a seizure data base by fusing seizures from several patients suffering from similar types of epilepsy.

6.1.2 Best parameter set

Clustering method

The data sets showed optimal performance with different clustering algorithms. It seems that no single clustering method is adapted for all kinds of seizure movements. The K-means clustering did work quite well with all patients except patient F2. I would draw the conclusion that, if the seizures are subtle (but still specific) a Gaussian mixture based clustering methods is needed. When the seizures are a bit more energetic, K-means based clustering methods is better since they can fit a larger number of clusters. Also, K-means method is less computationally intensive which might be important in embedded implementations.

PLS variance

It is evident that the amount of variance chosen as a threshold in the PLS-DA dimension reduction must be based on both the clustering method and the patient. No general conclusion can be drawn here.

Classifier

Generally K-nearest neighbors seem to yield the best results when both sensitivity and selectivity are important. QDA might be useful when sensitivity are far more important than selectivity.

6.1.3 Best feature sets

Activity measures

For some patients, there was no differences in the results when choosing between the three measures (SMA, VM and MAMD) while for some others there were a difference. It does, however, seem like there is some redundant discriminatory power among the three activity measures. Not all three are expected to be necessary but in some cases two of them may be.

DC values

DC values seem to play an important role for some kinds of seizures. Since they are basically a measure of inclination, they should be used for seizures associated with some orientation. One might suspect that for more subtle seizures, the DC values become more important to rule out non seizure activities based on the orientation.

Highest frequency

The highest frequency band (13.25-25Hz), does not seem to hold any discriminatory information. This is the frequency band that is affected the most if some radio packets are lost. Furthermore, the body is not able to oscillate at these frequencies. The reason why these frequencies might exist in the measurements can be explained by overtones caused by mechanics inside the sensor boxes or the fixation of the sensor boxes to the limbs. These kinds of oscillations might therefore be more typical of certain hardware under special circumstances. It might be hazardous to use such properties since they might not exist in all measurement setups.

Another origin of so high frequency components might be fast acceleration changes in a movement. Such changes will cause the fourier transform for that time interval to contain high frequency components of reasonable amplitude. However, drastic acceleration activity is expected to show in other features (such as the general activity measures). Thus, it is advisable to remove this feature from the algorithm.

Frequency components

The frequency components seem important even though they are not crucial for every patient. At least all seizures containing clonic episodes should be classified with aid of the frequency components. These features should remain in the feature sets for all kinds of epilepsy.

Periodicity and correlation

The periodicity and correlation features did not seem to hold exclusive discriminatory information. Because of the physical interpretation of these features, it is not expected that they are correlated. Therefore, it may be possible to remove them both although that was never tested during the evaluation.

Sensors

It seems that one or two sensors was adequate for all patients except patient F2 in this study. How many sensors or where to put them might however be up to the characteristics of the patients seizures. When analyzing a patient, it would be recommended to tune

the classifier and hardware setup for that particular patient for optimal classification performance.

Feature differentials

The feature differential transformation did cause severe decrease in performance for all data sets except for patient 14, where the performance was instead increased. For patient 14, even though the performance was increased, the results acquired was not good enough to be useful in an actual implementation.

Although feature differentials were not useful on their own, the fact that they improved the results for patient 14 indicates that they could be useful as a complement to the normal ones.

6.1.4 Comparison between old and new method

The new classification method has shown superior to the old method for all data sets, except patient 14 where the two methods performed similarly inadequate. No advantage in performance has been recognizable for the old method. However, the old method is less computationally complex and might therefore yield advantages in implementations on systems restricted by battery or computational power.

6.2 Future work

There is plenty of possibilities for future work. First of all, obtaining more data sets that fulfill the requirements of consistent seizure behavior and many seizures of the kind that affect muscular movement. Once more data sets are available, a larger and more quantative evaluation study should be performed in order to statistically assess the performance of the classifier.

Hardware specific implementation of the proposed method from this thesis can be an interesting task. Here, numerical stability and computational complexity on a small embedded system introduces new challenges. The hardware can also be implemented to interact with the user (in this case the patient). In this way, the hardware can communicate with the user when a classification takes place. The users might be able to give their input on whether the alarm was a probable seizure or a false alarm. This human-machine interaction might be developed further with the introduction of on-line learning, where the classifier is trained with more data in real time as soon as it is gathered. By doing so, the patient might give input which can be used to confirm data and consequently expand the training set.

There is also the possibility to perform statistical analysis of the accelerometer- or feature-data - and study the possibility of predicting seizures in advance.

Furthermore, there are other kinds of classification methods that might perform better than the proposed method from this thesis. Here, a new classification method or an add-on to the existing one can be developed. Methods such as artificial neural networks or hidden Markov chains could for example be used. Also, more complex clustering and classification techniques that might perform better than the existing ones could be implemented.

The classification performance could be enhanced by developing new and more discriminatory measurements (features) derived from the accelerometer data. To be able to find such measurements, it would be necessary to have a good understanding of epilepsy and its complications. A person that studies or have studied medicine would be more appropriate for such a task.

More areas of application for the product could be identified. It could be applied to study similar disorders such as Parkinson, to monitor the status of the elderly, as diagnosing or training tools for athletes, among others. Classification of movements for non-humans such as animals or machines might also be a market for such a device.

Finally, modifying the hardware can also bring new possibilities. As mentioned in the discussion chapter, finding tonic seizures would be much easier if the position of the limbs were known at all times instead of only their acceleration. By introducing hardware capable of finding these positions, new doors would open for seizure classification.

6.3 Conclusion

Let us assess how well the purpose of this thesis (see sec.1.4) has been fulfilled:

- **Include time evolutionary acceleration information into the decision process of the seizure classifier.**

  The time evolutionary acceleration information has been included on a larger time scale throughout the transition matrix extraction (see sec.3.3). Since these transition matrices are used in the final classification decision, this purpose is considered to be clearly fulfilled.

- **Enhance the performance of the current classifier method.**

  The pilot study, which evaluated the performance of the new classification method, shows only advantages compared to the old classification method. The pilot study did only evaluate the two classification methods on four different data sets. It is therefore impossible to say, with complete certainty, that the performance of these four data sets is representative for the majority of possible epilepsy measurements. As far as we know, it seems likely that future measurements will behave similarly to the tested data sets and, if they do, the new classification method will perform much better than the old one.

- **Perform a pilot study, assessing the performance of the proposed classification method.**

The pilot study was performed and some conclusions has been drawn (as far as conclusions can be drawn from the limited number of data sets studied).

First of all, the clustering algorithm that gives the best result is entirely dependent on the kind of behavior studied. Further on, the parameters of the different stages of the classification method are all quite dependant on the data studied. Therefore, it seems like the parameters and clustering techniques to use should be ”tuned” for the particular patient behavior that is being studied. For example, a senior citizen with a restricted movement pattern will have a less active background data compared to an athlete. Therefore, these two patients might demand differently tuned classification methods.

The KNN classifier (see sec.3.2.2) seems to perform better or equal to the QDA classifier (see sec.3.2.1) for all tested data sets. The QDA classifier does however have some computational advantages. If KNN is too demanding, it might be necessary to use QDA anyway. Some optimization of the KNN classifier might solve this problem.

Neither the new or the old classification method seem to be able to find tonic seizures. This is believed to derive from the non-original acceleration pattern of tonic seizures.

Generalizing seizure behavior from one patient model to another seem to be possible. This means that, if two patients suffer from the same kind of seizures, it should be possible to use the classification model from one of the patients to find seizures at the other patient. This could also be used to increase the training set while training the classifier.

Finding very original seizure behavior seems to be possible even when the background data is very active and extreme. This opens up the possibility to have very strong background data and find normal seizures as well.

- **Investigate possibilities of decreasing the number of sensors and decreasing the computational complexity of the classifier.**

  This task was performed by removing some feature groups at a time and evaluate if and how the classification performance changed.

  It seemed like some, but not all, of the general activity measurements should remain in the feature set. The highest frequency band (13.25 - 25 Hz) as well as the periodicity and correlation features can be removed without any decrease in classification performance. The DC and the rest of the frequency bands should remain in the features set.

  The number and placements of sensors needed are completely dependent on the patients seizure and background movement behavior. For the most patients it seems possible to only include 1 or 2 sensors.

  The time derivatives of the features did not show any discriminatory advantage compared to the original feature values, quite the contrary.

Bibliography

[1] Charlotte Sachs, Gösta Blennow, and Jörgen Malmquist. Epilepsi - nationalencyklopedin, 2011. [Online; accessed 29-April-2011].

[2] Encyclopedia Britannica. Epilepsy - encyclopedia britannica, 2011. [Online; accessed 29-April-2011].

[3] Johan Wipenmyr. Utveckling av metoder för diskriminering av epileptiska anfall. Master's thesis, Luleå university of technology, 2011.

[4] J Stigwall, J Wipenmyr, T Petterson, B Rydenhag, and K Malmgren. Automatic registration and classifcation of epileptic seizures using wireless motion sensors. *proceedings of the 7th international conference on wearable micro- and nano-technologies for personalized health, pHealth*, May 2010.

[5] Raisonance. *STM32-Primer2 user manual*, 2009. www.raisonance.com.

[6] Dynastream innovations inc, 228 River Avenue, Cochrane, Alberta, Canada T4C 2C1. *ANT Message Protocol and Usage*, 4.2 edition. thisisant.com.

[7] M. J. Mathie, A. C. F. Coster, N. H. Lovell, and B. G. Celler. Detection of daily physical activities using a triaxial accelerometer.

[8] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image Processing, Analysis and Machine Vision*. Thomson learning, Toronto, third edition, 2008.

[9] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of statistical learning*. Springer, Stanford, California, second edition, 2008.

[10] A.K Jain, M.N Murty, and P.J Flynn. Data clustering: A review. *ACM Computing Surveys*, 31(3):316–323, 1999.

[11] Richard A Redner and Homer F Walker. Mixture densities, maximum likelihood and the em algorithm. *SIAM Review*, 26(2):195–239, 1984.

[12] David Williams. *Weighting the odds*. Cambridge university press, The Pitt building, Trumpington street, Cambridge, United kingdom, 2004.

[13] Jiang Xudong. Linear subspace learning-based dimensionality reduction. *IEEE Signal processing magazine*, 28(2):16–26, 2011.

[14] Anil K. Jain, Robert P.W. Duin, and Jianchang Mao. Statistical pattern recognition: A review. *IEEE Transactions on pattern analysis and machine intelligence*, 22(1):4–37, 2000.

[15] C Lay, David. *Linear Algebra and its application.* Pearson, third edition, 2003.

[16] P Geladi and B.R. Kowalski. Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, 185(C):1–17, 1986.

[17] Svante Wold, Mikael Sjöström, and Lennart Eriksson. Pls-regression: a basic tool for chemometrics. *Chemometrics and intelligent laboratory systems*, 58:109–130, 2001.

Appendix A

Complete results

A.1 Patient 7

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	75%	11	9	3
	82%	75%	11	9	3
	82%	75%	11	9	3
	73%	73%	11	8	3
	73%	67%	11	8	4
Mean	78.4%	73.0%	11	8.6	3.2
Median	82.0%	75.0%	11	9.0	3.0
STD	4.9%	3.5%	0	0.5	0.4

Features used:	All features				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	82%	11	9	2
	82%	75%	11	9	3
	82%	75%	11	9	3
	73%	80%	11	8	2
	73%	73%	11	8	3
Mean	78.4%	77.0%	11	8.6	2.6
Median	82.0%	75.0%	11	9.0	3.0
STD	4.9%	3.8%	0	0.5	0.5

Features used:	All features				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	91%	77%	11	10	3
	82%	82%	11	9	2
	82%	69%	11	9	4
	73%	80%	11	8	2
	73%	73%	11	8	3
Mean	80.2%	76.2%	11	8.8	2.8
Median	82.0%	77.0%	11	9.0	3.0
STD	7.5%	5.3%	0	0.8	0.8

Features used:	All features				
Cluster analysis:	Gm hard, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	82%	11	9	2
	82%	69%	11	9	4
	82%	64%	11	9	5
	82%	64%	11	9	5
	82%	56%	11	9	7
Mean	82.0%	67.0%	11	9.0	4.6
Median	82.0%	64.0%	11	9.0	5.0
STD	0.0%	9.6%	0	0.0	1.8

Features used:	All features				
Cluster analysis:	Gm hard, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	52%	11	11	10
	100%	50%	11	11	11
	91%	50%	11	10	10
	82%	56%	11	9	7
	82%	47%	11	9	10
Mean	91.0%	51.0%	11	10.0	9.6
Median	91.0%	50.0%	11	10.0	10.0
STD	9.0%	3.3%	0	1.0	1.5

Features used:	All features				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	91%	48%	11	10	11
	91%	48%	11	10	11
	82%	60%	11	9	6
	82%	47%	11	9	10
	82%	43%	11	9	12
Mean	85.6%	49.2%	11	9.4	10.0
Median	82.0%	48.0%	11	9.0	11.0
STD	4.9%	6.4%	0	0.5	2.3

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	75%	11	9	3
	82%	56%	11	9	7
	82%	56%	11	9	7
	73%	50%	11	8	8
	73%	47%	11	8	9
Mean	78.4%	56.8%	11	8.6	6.8
Median	82.0%	56.0%	11	9.0	7.0
STD	4.9%	10.9%	0	0.5	2.3

Features used:	All features				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	92%	11	11	1
	91%	77%	11	10	3
	91%	77%	11	10	3
	91%	77%	11	10	3
	82%	90%	11	9	1
Mean	91.0%	82.6%	11	10.0	2.2
Median	91.0%	77.0%	11	10.0	3.0
STD	6.4%	7.7%	0	0.7	1.1

Features used:	All features				
Cluster analysis:	Kmeans seizure separated, 4 of 7 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	91%	77%	11	10	3
	91%	77%	11	10	3
	82%	82%	11	9	2
	82%	82%	11	9	2
	82%	75%	11	9	3
Mean	85.6%	78.6%	11	9.4	2.6
Median	82.0%	77.0%	11	9.0	3.0
STD	4.9%	3.2%	0	0.5	0.5

Features used:	All features				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	91%	83%	11	10	2
	82%	90%	11	9	1
	82%	82%	11	9	2
	82%	82%	11	9	2
	82%	75%	11	9	3
Mean	83.8%	82.4%	11	9.2	2.0
Median	82.0%	82.0%	11	9.0	2.0
STD	4.0%	5.3%	0	0.4	0.7

Features used:	All features				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	85%	11	11	2
	100%	85%	11	11	2
	91%	83%	11	10	2
	91%	83%	11	10	2
	82%	82%	11	9	2
Mean	92.8%	83.6%	11	10.2	2.0
Median	91.0%	83.0%	11	10.0	2.0
STD	7.5%	1.3%	0	0.8	0.0

Features used:	All features				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 1				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	100%	11	9	0
	82%	90%	11	9	1
	82%	90%	11	9	1
	82%	90%	11	9	1
	82%	90%	11	9	1
Mean	82.0%	92.0%	11	9.0	0.8
Median	82.0%	90.0%	11	9.0	1.0
STD	0.0%	4.5%	0	0.0	0.4

Features used:	All features				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	QDA, Prior prob. weights [1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	69%	11	11	5
	100%	65%	11	11	6
	100%	61%	11	11	7
	100%	55%	11	11	9
	91%	71%	11	10	4
Mean	98.2%	64.2%	11	10.8	6.2
Median	100.0%	65.0%	11	11.0	6.0
STD	4.0%	6.4%	0	0.4	1.9

Features used:	All features				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	73%	11	11	4
	100%	69%	11	11	5
	100%	58%	11	11	8
	91%	63%	11	10	6
	91%	56%	11	10	8
Mean	96.4%	63.8%	11	10.6	6.2
Median	100.0%	63.0%	11	11.0	6.0
STD	4.9%	7.2%	0	0.5	1.8

Features used:	All features				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	QDA, Prior prob. weights [10 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	73%	11	11	4
	100%	65%	11	11	6
	100%	65%	11	11	6
	100%	61%	11	11	7
	91%	59%	11	10	7
Mean	98.2%	64.6%	11	10.8	6.0
Median	100.0%	65.0%	11	11.0	6.0
STD	4.0%	5.4%	0	0.4	1.2

Features used:	All features				
Cluster analysis:	Max values				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	73%	11	11	4
	91%	77%	11	10	3
	91%	77%	11	10	3
	91%	77%	11	10	3
	82%	75%	11	9	3
Mean	91.0%	75.8%	11	10.0	3.2
Median	91.0%	77.0%	11	10.0	3.0
STD	6.4%	1.8%	0	0.7	0.4

Features used:	All features				
Cluster analysis:	Min values				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	43%	11	9	12
	73%	44%	11	8	10
	73%	40%	11	8	12
	73%	38%	11	8	13
	73%	36%	11	8	14
Mean	74.8%	40.2%	11	8.2	12.2
Median	73.0%	40.0%	11	8.0	12.0
STD	4.0%	3.3%	0	0.4	1.5

Features used:	All features				
Cluster analysis:	Max absolute values				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	91%	59%	11	10	7
	82%	64%	11	9	5
	82%	60%	11	9	6
	82%	50%	11	9	9
	82%	50%	11	9	9
Mean	83.8%	56.6%	11	9.2	7.2
Median	82.0%	59.0%	11	9.0	7.0
STD	4.0%	6.3%	0	0.4	1.8

Features used:	All features				
Cluster analysis:	Max values				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	91%	71%	11	10	4
	82%	69%	11	9	4
	82%	69%	11	9	4
	82%	69%	11	9	4
	82%	69%	11	9	4
Mean	83.8%	69.4%	11	9.2	4.0
Median	82.0%	69.0%	11	9.0	4.0
STD	4.0%	0.9%	0	0.4	0.0

Features used:	All features				
Cluster analysis:	Min values				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	60%	11	9	6
	82%	56%	11	9	7
	82%	50%	11	9	9
	82%	50%	11	9	9
	73%	47%	11	8	9
Mean	80.2%	52.6%	11	8.8	8.0
Median	82.0%	50.0%	11	9.0	9.0
STD	4.0%	5.3%	0	0.4	1.4

Features used:	All features				
Cluster analysis:	Max absolute values				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	91%	77%	11	10	3
	91%	71%	11	10	4
	91%	67%	11	10	5
	82%	75%	11	9	3
	82%	64%	11	9	5
Mean	87.4%	70.8%	11	9.6	4.0
Median	91.0%	71.0%	11	10.0	4.0
STD	4.9%	5.4%	0	0.5	1.0

Features used:	Without VM and MAMD				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	82%	11	9	2
	91%	83%	11	10	2
	91%	83%	11	10	2
	100%	85%	11	11	2
	91%	91%	11	10	1
Mean	91.0%	84.8%	11	10.0	1.8
Median	91.0%	83.0%	11	10.0	2.0
STD	6.4%	3.6%	0	0.7	0.4

Features used:	Without SMA and MAMD				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	85%	11	11	2
	91%	83%	11	10	2
	100%	79%	11	11	3
	91%	83%	11	10	2
	100%	85%	11	11	2
Mean	96.4%	83.0%	11	10.6	2.2
Median	100.0%	83.0%	11	11.0	2.0
STD	4.9%	2.4%	0	0.5	0.4

Features used:	Without SMA and VM				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	73%	80%	11	8	2
	82%	82%	11	9	2
	82%	82%	11	9	2
	91%	83%	11	10	2
	100%	85%	11	11	2
Mean	85.6%	82.4%	11	9.4	2.0
Median	82.0%	82.0%	11	9.0	2.0
STD	10.3%	1.8%	0	1.1	0.0

Features used:	Without DC				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	82%	11	9	2
	82%	82%	11	9	2
	91%	83%	11	10	2
	91%	83%	11	10	2
	100%	79%	11	11	3
Mean	89.2%	81.8%	11	9.8	2.2
Median	91.0%	82.0%	11	10.0	2.0
STD	7.5%	1.6%	0	0.8	0.4

Features used:	Without CORR				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	85%	11	11	2
	82%	82%	11	9	2
	82%	90%	11	9	1
	82%	90%	11	9	1
	100%	85%	11	11	2
Mean	89.2%	86.4%	11	9.8	1.6
Median	82.0%	85.0%	11	9.0	2.0
STD	9.9%	3.5%	0	1.1	0.5

Features used:	Without PER				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	82%	11	9	2
	91%	83%	11	10	2
	91%	91%	11	10	1
	91%	91%	11	10	1
	100%	85%	11	11	2
Mean	91.0%	86.4%	11	10.0	1.6
Median	91.0%	85.0%	11	10.0	2.0
STD	6.4%	4.3%	0	0.7	0.5

Features used:	Without FREQ				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	64%	11	9	5
	91%	71%	11	10	4
	91%	71%	11	10	4
	91%	77%	11	10	3
	100%	79%	11	11	3
Mean	91.0%	72.4%	11	10.0	3.8
Median	91.0%	71.0%	11	10.0	4.0
STD	6.4%	5.9%	0	0.7	0.8

Features used:	Without highest FREQ				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	79%	11	11	3
	100%	85%	11	11	2
	91%	83%	11	10	2
	91%	83%	11	10	2
	91%	91%	11	10	1
Mean	94.6%	84.2%	11	10.4	2.0
Median	91.0%	83.0%	11	10.0	2.0
STD	4.9%	4.4%	0	0.5	0.7

Features used:	Without sensor 1				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	82%	11	9	2
	91%	83%	11	10	2
	91%	83%	11	10	2
	91%	83%	11	10	2
	91%	91%	11	10	1
Mean	89.2%	84.4%	11	9.8	1.8
Median	91.0%	83.0%	11	10.0	2.0
STD	4.0%	3.7%	0	0.4	0.4

Features used:	Without sensor 2				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	91%	83%	11	10	2
	91%	83%	11	10	2
	91%	83%	11	10	2
	91%	83%	11	10	2
	91%	83%	11	10	2
Mean	91.0%	83.0%	11	10.0	2.0
Median	91.0%	83.0%	11	10.0	2.0
STD	0.0%	0.0%	0	0.0	0.0

Features used:	Without sensor 3				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	73%	42%	11	8	11
	64%	32%	11	7	15
	73%	89%	11	8	1
	73%	40%	11	8	12
	73%	50%	11	8	8
Mean	71.2%	50.6%	11	7.8	9.4
Median	73.0%	42.0%	11	8.0	11.0
STD	4.0%	22.4%	0	0.4	5.3

Features used:	Without sensor 1 and 2				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	75%	11	9	3
	91%	53%	11	10	9
	91%	59%	11	10	7
	91%	83%	11	10	2
	91%	91%	11	10	1
Mean	89.2%	72.2%	11	9.8	4.4
Median	91.0%	75.0%	11	10.0	3.0
STD	4.0%	16.0%	0	0.4	3.4

Features used:	Without sensor 2 and 3				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	73%	67%	11	8	4
	73%	62%	11	8	5
	73%	67%	11	8	4
	82%	64%	11	9	5
	73%	67%	11	8	4
Mean	74.8%	65.4%	11	8.2	4.4
Median	73.0%	67.0%	11	8.0	4.0
STD	4.0%	2.3%	0	0.4	0.5

Features used:	Without sensor 1 and 3				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	73%	42%	11	8	11
	73%	53%	11	8	7
	73%	57%	11	8	6
	82%	41%	11	9	13
	91%	50%	11	10	10
Mean	78.4%	48.6%	11	8.6	9.4
Median	73.0%	50.0%	11	8.0	10.0
STD	8.0%	6.9%	0	0.9	2.9

Features used:	Using feature differentials				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	60%	11	9	6
	91%	50%	11	10	10
	91%	67%	11	10	5
	100%	58%	11	11	8
	100%	61%	11	11	7
Mean	92.8%	59.2%	11	10.2	7.2
Median	91.0%	60.0%	11	10.0	7.0
STD	7.5%	6.1%	0	0.8	1.9

Features used:	Optimal set				
Cluster analysis:	No clustering				
PLS-DA variance:	100%				
Classifier:	Old method, Prior prob. weights [1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	41%	11	9	13
	82%	41%	11	9	13
	82%	41%	11	9	13
	82%	43%	11	9	12
	82%	43%	11	9	12
Mean	82.0%	41.8%	11	9.0	12.6
Median	82.0%	41.0%	11	9.0	13.0
STD	0.0%	1.1%	0	0.0	0.5

Features used:	Optimal set				
Cluster analysis:	No clustering				
PLS-DA variance:	100%				
Classifier:	Old method, Prior prob. weights [1 10^{15}]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	73%	62%	11	8	5
	73%	62%	11	8	5
	82%	64%	11	9	5
	82%	64%	11	9	5
	82%	64%	11	9	5
Mean	78.4%	63.2%	11	8.6	5.0
Median	82.0%	64.0%	11	9.0	5.0
STD	4.9%	1.1%	0	0.5	0.0

A.2 Patient 14

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	77%	25%	22	17	50
	64%	28%	22	14	36
	59%	24%	22	13	42
	59%	22%	22	13	46
	50%	20%	22	11	45
Mean	61.8%	23.8%	22	13.6	43.8
Median	59.0%	24.0%	22	13.0	45.0
STD	9.9%	3.0%	0	2.2	5.2

Features used:	All features				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	59%	28%	22	13	34
	55%	24%	22	12	38
	55%	24%	22	12	39
	50%	22%	22	11	39
	50%	21%	22	11	41
Mean	53.8%	23.8%	22	11.8	38.2
Median	55.0%	24.0%	22	12.0	39.0
STD	3.8%	2.7%	0	0.8	2.6

Features used:	All features				
Cluster analysis:	Kmeans, 6 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	59%	25%	22	13	40
	45%	25%	22	10	30
	45%	21%	22	10	37
	41%	24%	22	9	28
	41%	19%	22	9	39
Mean	46.2%	22.8%	22	10.2	34.8
Median	45.0%	24.0%	22	10.0	37.0
STD	7.4%	2.7%	0	1.6	5.4

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	68%	22%	22	15	54
	64%	32%	22	14	30
	64%	24%	22	14	45
	50%	18%	22	11	51
	45%	17%	22	10	49
Mean	58.2%	22.6%	22	12.8	45.8
Median	64.0%	22.0%	22	14.0	49.0
STD	10.1%	6.0%	0	2.2	9.4

Features used:	All features				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	68%	30%	22	15	35
	55%	27%	22	12	33
	50%	24%	22	11	34
	50%	21%	22	11	41
	32%	19%	22	7	29
Mean	51.0%	24.2%	22	11.2	34.4
Median	50.0%	24.0%	22	11.0	34.0
STD	12.9%	4.4%	0	2.9	4.3

Features used:	All features				
Cluster analysis:	Kmeans, 6 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	55%	31%	22	12	27
	50%	22%	22	11	38
	45%	27%	22	10	27
	45%	23%	22	10	34
	36%	21%	22	8	30
Mean	46.2%	24.8%	22	10.2	31.2
Median	45.0%	23.0%	22	10.0	30.0
STD	7.0%	4.1%	0	1.5	4.8

Features used:	All features				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	68%	31%	22	15	34
	68%	25%	22	15	46
	64%	25%	22	14	43
	64%	23%	22	14	47
	64%	22%	22	14	49
Mean	65.6%	25.2%	22	14.4	43.8
Median	64.0%	25.0%	22	14.0	46.0
STD	2.2%	3.5%	0	0.5	5.9

Features used:	All features				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	77%	33%	22	17	34
	73%	33%	22	16	33
	73%	28%	22	16	42
	68%	34%	22	15	29
	68%	31%	22	15	34
Mean	71.8%	31.8%	22	15.8	34.4
Median	73.0%	33.0%	22	16.0	34.0
STD	3.8%	2.4%	0	0.8	4.7

Features used:	All features				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [10 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	73%	20%	22	16	66
	73%	18%	22	16	71
	68%	18%	22	15	70
	68%	14%	22	15	91
	64%	19%	22	14	60
Mean	69.2%	17.8%	22	15.2	71.6
Median	68.0%	18.0%	22	15.0	70.0
STD	3.8%	2.3%	0	0.8	11.7

Features used:	All features				
Cluster analysis:	Gm fuzzy, 5 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	36%	16%	22	8	41
	32%	22%	22	7	25
	32%	17%	22	7	35
	32%	16%	22	7	36
	23%	11%	22	5	41
Mean	31.0%	16.4%	22	6.8	35.6
Median	32.0%	16.0%	22	7.0	36.0
STD	4.8%	3.9%	0	1.1	6.5

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	55%	20%	22	12	49
	41%	19%	22	9	39
	36%	14%	22	8	49
	32%	15%	22	7	41
	9%	4%	22	2	46
Mean	34.6%	14.4%	22	7.6	44.8
Median	36.0%	15.0%	22	8.0	46.0
STD	16.7%	6.3%	0	3.6	4.6

Features used:	All features				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	55%	17%	22	12	57
	50%	17%	22	11	54
	45%	16%	22	10	53
	36%	13%	22	8	56
	27%	10%	22	6	53
Mean	42.6%	14.6%	22	9.4	54.6
Median	45.0%	16.0%	22	10.0	54.0
STD	11.2%	3.0%	0	2.4	1.8

Features used:	All features				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	55%	18%	22	12	56
	45%	14%	22	10	60
	41%	13%	22	9	62
	32%	13%	22	7	46
	32%	12%	22	7	50
Mean	41.0%	14.0%	22	9.0	54.8
Median	41.0%	13.0%	22	9.0	56.0
STD	9.7%	2.3%	0	2.1	6.7

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	45%	21%	22	10	37
	45%	19%	22	10	44
	41%	18%	22	9	40
	41%	17%	22	9	45
	41%	17%	22	9	45
Mean	42.6%	18.4%	22	9.4	42.2
Median	41.0%	18.0%	22	9.0	44.0
STD	2.2%	1.7%	0	0.5	3.6

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 1				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	41%	19%	22	9	39
	32%	16%	22	7	36
	27%	17%	22	6	30
	23%	10%	22	5	43
	18%	8%	22	4	45
Mean	28.2%	14.0%	22	6.2	38.6
Median	27.0%	16.0%	22	6.0	39.0
STD	8.8%	4.7%	0	1.9	5.9

Features used:	All features				
Cluster analysis:	Gm hard, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	41%	17%	22	9	45
	41%	14%	22	9	54
	36%	16%	22	8	43
	23%	8%	22	5	55
	18%	9%	22	4	39
Mean	31.8%	12.8%	22	7.0	47.2
Median	36.0%	14.0%	22	8.0	45.0
STD	10.7%	4.1%	0	2.3	7.0

Features used:	All features				
Cluster analysis:	Kmeans seizure separated, 3 of 7 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	73%	9%	22	16	164
	59%	7%	22	13	176
	45%	6%	22	10	159
	41%	5%	22	9	186
	36%	4%	22	8	175
Mean	50.8%	6.2%	22	11.2	172.0
Median	45.0%	6.0%	22	10.0	175.0
STD	15.1%	1.9%	0	3.3	10.7

Features used:	All features				
Cluster analysis:	Max value				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	27%	10%	22	6	54
	23%	8%	22	5	54
	23%	7%	22	5	65
	18%	6%	22	4	62
	14%	5%	22	3	61
Mean	21.0%	7.2%	22	4.6	59.2
Median	23.0%	7.0%	22	5.0	61.0
STD	5.0%	1.9%	0	1.1	5.0

Features used:	All features				
Cluster analysis:	Min value				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	23%	7%	22	5	62
	23%	7%	22	5	67
	18%	7%	22	4	57
	14%	4%	22	3	64
	9%	3%	22	2	61
Mean	17.4%	5.6%	22	3.8	62.2
Median	18.0%	7.0%	22	4.0	62.0
STD	6.0%	1.9%	0	1.3	3.7

Features used:	All features				
Cluster analysis:	Max absolute value				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	5%	1%	22	1	80
	5%	1%	22	1	102
	0%	0%	22	0	102
	0%	0%	22	0	93
	0%	0%	22	0	84
Mean	2.0%	0.4%	22	0.4	92.2
Median	0.0%	0.0%	22	0.0	93.0
STD	2.7%	0.5%	0	0.5	10.1

Features used:	Without VM and MAMD				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	59%	24%	22	13	42
	59%	27%	22	13	36
	59%	28%	22	13	33
	68%	30%	22	15	35
	77%	30%	22	17	39
Mean	64.4%	27.8%	22	14.2	37.0
Median	59.0%	28.0%	22	13.0	36.0
STD	8.0%	2.5%	0	1.8	3.5

Features used:	Without SMA and MAMD				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	64%	27%	22	14	37
	68%	34%	22	15	29
	68%	38%	22	15	24
	77%	35%	22	17	32
	77%	36%	22	17	30
Mean	70.8%	34.0%	22	15.6	30.4
Median	68.0%	35.0%	22	15.0	30.0
STD	5.9%	4.2%	0	1.3	4.7

Features used:	Without SMA and VM				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	73%	32%	22	16	34
	73%	32%	22	16	34
	77%	27%	22	17	45
	77%	35%	22	17	32
	77%	35%	22	17	31
Mean	75.4%	32.2%	22	16.6	35.2
Median	77.0%	32.0%	22	17.0	34.0
STD	2.2%	3.3%	0	0.5	5.6

Features used:	Without DC				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	36%	12%	22	8	59
	41%	18%	22	9	40
	45%	17%	22	10	50
	50%	26%	22	11	31
	55%	32%	22	12	26
Mean	45.4%	21.0%	22	10.0	41.2
Median	45.0%	18.0%	22	10.0	40.0
STD	7.4%	7.9%	0	1.6	13.5

Features used:	Without CORR				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	55%	38%	22	12	20
	59%	29%	22	13	32
	59%	32%	22	13	28
	73%	33%	22	16	32
	73%	36%	22	16	29
Mean	63.8%	33.6%	22	14.0	28.2
Median	59.0%	33.0%	22	13.0	29.0
STD	8.6%	3.5%	0	1.9	4.9

Features used:	Without PER				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	59%	22%	22	13	45
	64%	22%	22	14	49
	64%	25%	22	14	42
	64%	29%	22	14	34
	73%	27%	22	16	43
Mean	64.8%	25.0%	22	14.2	42.6
Median	64.0%	25.0%	22	14.0	43.0
STD	5.1%	3.1%	0	1.1	5.5

Features used:	Without FREQ				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	41%	27%	22	9	24
	50%	34%	22	11	21
	50%	34%	22	11	21
	55%	24%	22	12	37
	59%	37%	22	13	22
Mean	51.0%	31.2%	22	11.2	25.0
Median	50.0%	34.0%	22	11.0	22.0
STD	6.7%	5.4%	0	1.5	6.8

Features used:	Without highest FREQ				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	55%	23%	22	12	40
	68%	26%	22	15	43
	73%	27%	22	16	44
	73%	28%	22	16	41
	73%	36%	22	16	28
Mean	68.4%	28.0%	22	15.0	39.2
Median	73.0%	27.0%	22	16.0	41.0
STD	7.8%	4.8%	0	1.7	6.5

Features used:	Without sensor 1				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	45%	14%	22	10	59
	50%	22%	22	11	38
	55%	15%	22	12	70
	55%	16%	22	12	64
	59%	24%	22	13	42
Mean	52.8%	18.2%	22	11.6	54.6
Median	55.0%	16.0%	22	12.0	59.0
STD	5.4%	4.5%	0	1.1	14.0

Features used:	Without sensor 2				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	23%	12%	22	5	37
	27%	17%	22	6	29
	27%	17%	22	6	29
	32%	20%	22	7	28
	36%	17%	22	8	38
Mean	29.0%	16.6%	22	6.4	32.2
Median	27.0%	17.0%	22	6.0	29.0
STD	5.0%	2.9%	0	1.1	4.9

Features used:	Without sensor 3				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	73%	37%	22	16	27
	73%	38%	22	16	26
	73%	48%	22	16	17
	77%	38%	22	17	28
	77%	41%	22	17	24
Mean	74.6%	40.4%	22	16.4	24.4
Median	73.0%	38.0%	22	16.0	26.0
STD	2.2%	4.5%	0	0.5	4.4

Features used:	Without sensor 1 and 2				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	59%	13%	22	13	89
	64%	12%	22	14	100
	68%	14%	22	15	93
	73%	13%	22	16	103
	86%	14%	22	19	118
Mean	70.0%	13.2%	22	15.4	100.6
Median	68.0%	13.0%	22	15.0	100.0
STD	10.3%	0.8%	0	2.3	11.2

Features used:	Without sensor 2 and 3				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	55%	27%	22	12	33
	55%	27%	22	12	32
	59%	26%	22	13	37
	59%	27%	22	13	35
	59%	29%	22	13	32
Mean	57.4%	27.2%	22	12.6	33.8
Median	59.0%	27.0%	22	13.0	33.0
STD	2.2%	1.1%	0	0.5	2.2

Features used:	Without sensor 1 and 3				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	41%	22%	22	9	32
	41%	23%	22	9	31
	45%	23%	22	10	33
	50%	25%	22	11	33
	64%	25%	22	14	42
Mean	48.2%	23.6%	22	10.6	34.2
Median	45.0%	23.0%	22	10.0	33.0
STD	9.6%	1.3%	0	2.1	4.4

Features used:	Using feature differentials				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	95%	30%	22	21	49
	95%	39%	22	21	33
	95%	43%	22	21	28
	95%	47%	22	21	24
	100%	32%	22	22	46
Mean	96.0%	38.2%	22	21.2	36.0
Median	95.0%	39.0%	22	21.0	33.0
STD	2.2%	7.2%	0	0.4	11.0

Features used:	Optimal set				
Cluster analysis:	No clustering				
PLS-DA variance:	100%				
Classifier:	Old method, Prior prob. weights [1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	4%	22	18	476
	82%	4%	22	18	476
	82%	4%	22	18	463
	82%	4%	22	18	458
	82%	4%	22	18	448
Mean	82.0%	4.0%	22	18.0	464.2
Median	82.0%	4.0%	22	18.0	463.0
STD	0.0%	0.0%	0	0.0	12.0

Features used:	Optimal set				
Cluster analysis:	No clustering				
PLS-DA variance:	100%				
Classifier:	Old method, Prior prob. weights [1 2] open radius 10				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	73%	10%	22	16	138
	73%	11%	22	16	136
	73%	11%	22	16	133
	73%	11%	22	16	127
	82%	12%	22	18	130
Mean	74.8%	11.0%	22	16.4	132.8
Median	73.0%	11.0%	22	16.0	133.0
STD	4.0%	0.7%	0	0.9	4.4

A.3 Patient F1

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	93%	28	27	2
	96%	93%	28	27	2
	93%	93%	28	26	2
	93%	90%	28	26	3
	89%	93%	28	25	2
Mean	93.4%	92.4%	28	26.2	2.2
Median	93.0%	93.0%	28	26.0	2.0
STD	2.9%	1.3%	0	0.8	0.4

Features used:	All features				
Cluster analysis:	Kmeans, 5 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	93%	28	27	2
	96%	87%	28	27	4
	93%	93%	28	26	2
	93%	93%	28	26	2
	93%	90%	28	26	3
Mean	94.2%	91.2%	28	26.4	2.6
Median	93.0%	93.0%	28	26.0	2.0
STD	1.6%	2.7%	0	0.5	0.9

Features used:	All features				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	90%	28	27	3
	96%	90%	28	27	3
	96%	84%	28	27	5
	96%	93%	28	27	2
	96%	93%	28	27	2
Mean	96.0%	90.0%	28	27.0	3.0
Median	96.0%	90.0%	28	27.0	3.0
STD	0.0%	3.7%	0	0.0	1.2

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	56%	28	27	21
	89%	56%	28	25	20
	89%	53%	28	25	22
	82%	50%	28	23	23
	79%	49%	28	22	23
Mean	87.0%	52.8%	28	24.4	21.8
Median	89.0%	53.0%	28	25.0	22.0
STD	6.7%	3.3%	0	1.9	1.3

Features used:	All features				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	53%	28	27	24
	93%	48%	28	26	28
	93%	48%	28	26	28
	89%	52%	28	25	23
	89%	50%	28	25	25
Mean	92.0%	50.2%	28	25.8	25.6
Median	93.0%	50.0%	28	26.0	25.0
STD	3.0%	2.3%	0	0.8	2.3

Features used:	All features				
Cluster analysis:	Gm hard, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	60%	28	27	18
	96%	54%	28	27	23
	96%	54%	28	27	23
	89%	58%	28	25	18
	86%	44%	28	24	30
Mean	92.6%	54.0%	28	26.0	22.4
Median	96.0%	54.0%	28	27.0	23.0
STD	4.8%	6.2%	0	1.4	4.9

Features used:	All features				
Cluster analysis:	Gm hard, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	93%	53%	28	26	23
	93%	47%	28	26	29
	96%	53%	28	27	24
	96%	48%	28	27	29
	93%	47%	28	26	29
Mean	94.2%	49.6%	28	26.4	26.8
Median	93.0%	48.0%	28	26.0	29.0
STD	1.6%	3.1%	0	0.5	3.0

Features used:	All features				
Cluster analysis:	Max value				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	93%	81%	28	26	6
	93%	74%	28	26	9
	93%	72%	28	26	10
	93%	72%	28	26	10
	93%	72%	28	26	10
Mean	93.0%	74.2%	28	26.0	9.0
Median	93.0%	72.0%	28	26.0	10.0
STD	0.0%	3.9%	0	0.0	1.7

Features used:	All features				
Cluster analysis:	Min value				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	79%	28	27	7
	96%	75%	28	27	9
	93%	76%	28	26	8
	93%	70%	28	26	11
	93%	68%	28	26	12
Mean	94.2%	73.6%	28	26.4	9.4
Median	93.0%	75.0%	28	26.0	9.0
STD	1.6%	4.5%	0	0.5	2.1

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	50%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	90%	28	28	3
	100%	88%	28	28	4
	96%	90%	28	27	3
	93%	96%	28	26	1
	93%	93%	28	26	2
Mean	96.4%	91.4%	28	27.0	2.6
Median	96.0%	90.0%	28	27.0	3.0
STD	3.5%	3.1%	0	1.0	1.1

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	93%	28	27	2
	96%	93%	28	27	2
	96%	87%	28	27	4
	96%	84%	28	27	5
	89%	83%	28	25	5
Mean	94.6%	88.0%	28	26.6	3.6
Median	96.0%	87.0%	28	27.0	4.0
STD	3.1%	4.8%	0	0.9	1.5

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 1				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	93%	28	27	2
	96%	90%	28	27	3
	93%	90%	28	26	3
	93%	87%	28	26	4
	89%	89%	28	25	3
Mean	93.4%	89.8%	28	26.2	3.0
Median	93.0%	90.0%	28	26.0	3.0
STD	2.9%	2.2%	0	0.8	0.7

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	90%	28	27	3
	96%	87%	28	27	4
	96%	84%	28	27	5
	93%	93%	28	26	2
	93%	87%	28	26	4
Mean	94.8%	88.2%	28	26.6	3.6
Median	96.0%	87.0%	28	27.0	4.0
STD	1.6%	3.4%	0	0.5	1.1

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	82%	28	28	6
	100%	82%	28	28	6
	100%	78%	28	28	8
	100%	78%	28	28	8
	100%	72%	28	28	11
Mean	100.0%	78.4%	28	28.0	7.8
Median	100.0%	78.0%	28	28.0	8.0
STD	0.0%	4.1%	0	0.0	2.0

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [10 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	80%	28	28	7
	100%	80%	28	28	7
	100%	76%	28	28	9
	100%	74%	28	28	10
	96%	75%	28	27	9
Mean	99.2%	77.0%	28	27.8	8.4
Median	100.0%	76.0%	28	28.0	9.0
STD	1.8%	2.8%	0	0.4	1.3

Features used:	All features				
Cluster analysis:	Kmeans, 4 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights [1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	82%	28	28	6
	100%	80%	28	28	7
	100%	78%	28	28	8
	100%	74%	28	28	10
	96%	79%	28	27	7
Mean	99.2%	78.6%	28	27.8	7.6
Median	100.0%	79.0%	28	28.0	7.0
STD	1.8%	3.0%	0	0.4	1.5

Features used:	Without VM and MAMD				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	87%	28	27	4
	96%	90%	28	27	3
	96%	93%	28	27	2
	96%	93%	28	27	2
	96%	93%	28	27	2
Mean	96.0%	91.2%	28	27.0	2.6
Median	96.0%	93.0%	28	27.0	2.0
STD	0.0%	2.7%	0	0.0	0.9

Features used:	Without SMA and MAMD				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	88%	28	28	4
	100%	90%	28	28	3
	100%	90%	28	28	3
	100%	93%	28	28	2
	100%	93%	28	28	2
Mean	100.0%	90.8%	28	28.0	2.8
Median	100.0%	90.0%	28	28.0	3.0
STD	0.0%	2.2%	0	0.0	0.8

Features used:	Without SMA and VM				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	93%	90%	28	26	3
	96%	90%	28	27	3
	96%	90%	28	27	3
	96%	90%	28	27	3
	96%	93%	28	27	2
Mean	95.4%	90.6%	28	26.8	2.8
Median	96.0%	90.0%	28	27.0	3.0
STD	1.3%	1.3%	0	0.4	0.4

Features used:	Without DC				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	84%	28	27	5
	96%	87%	28	27	4
	96%	90%	28	27	3
	96%	90%	28	27	3
	100%	88%	28	28	4
Mean	96.8%	87.8%	28	27.2	3.8
Median	96.0%	88.0%	28	27.0	4.0
STD	1.8%	2.5%	0	0.4	0.8

Features used:	Without CORR				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	90%	28	27	3
	96%	93%	28	27	2
	96%	93%	28	27	2
	96%	93%	28	27	2
	96%	93%	28	27	2
Mean	96.0%	92.4%	28	27.0	2.2
Median	96.0%	93.0%	28	27.0	2.0
STD	0.0%	1.3%	0	0.0	0.4

Features used:	Without PER				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	87%	28	27	4
	96%	90%	28	27	3
	96%	93%	28	27	2
	96%	93%	28	27	2
	96%	96%	28	27	1
Mean	96.0%	91.8%	28	27.0	2.4
Median	96.0%	93.0%	28	27.0	2.0
STD	0.0%	3.4%	0	0.0	1.1

Features used:	Without FREQ				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	84%	28	27	5
	96%	87%	28	27	4
	96%	87%	28	27	4
	96%	87%	28	27	4
	100%	85%	28	28	5
Mean	96.8%	86.0%	28	27.2	4.4
Median	96.0%	87.0%	28	27.0	4.0
STD	1.8%	1.4%	0	0.4	0.5

Features used:	Without highest FREQ				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	93%	28	27	2
	96%	93%	28	27	2
	96%	93%	28	27	2
	96%	93%	28	27	2
	96%	93%	28	27	2
Mean	96.0%	93.0%	28	27.0	2.0
Median	96.0%	93.0%	28	27.0	2.0
STD	0.0%	0.0%	0	0.0	0.0

Features used:	Without sensor 1				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	93%	90%	28	26	3
	96%	84%	28	27	5
	96%	90%	28	27	3
	96%	90%	28	27	3
	100%	90%	28	28	3
Mean	96.2%	88.8%	28	27.0	3.4
Median	96.0%	90.0%	28	27.0	3.0
STD	2.5%	2.7%	0	0.7	0.9

Features used:	Without sensor 2				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	93%	93%	28	26	2
	93%	93%	28	26	2
	93%	93%	28	26	2
	96%	90%	28	27	3
	96%	90%	28	27	3
Mean	94.2%	91.8%	28	26.4	2.4
Median	93.0%	93.0%	28	26.0	2.0
STD	1.6%	1.6%	0	0.5	0.5

Features used:	Without sensor 3				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	93%	28	28	2
	100%	93%	28	28	2
	100%	93%	28	28	2
	100%	93%	28	28	2
	100%	93%	28	28	2
Mean	100.0%	93.0%	28	28.0	2.0
Median	100.0%	93.0%	28	28.0	2.0
STD	0.0%	0.0%	0	0.0	0.0

Features used:	Without sensor 1 and 2				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	96%	82%	28	27	6
	96%	84%	28	27	5
	96%	84%	28	27	5
	96%	84%	28	27	5
	96%	87%	28	27	4
Mean	96.0%	84.2%	28	27.0	5.0
Median	96.0%	84.0%	28	27.0	5.0
STD	0.0%	1.8%	0	0.0	0.7

Features used:	Without sensor 2 and 3				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	93%	90%	28	26	3
	93%	90%	28	26	3
	93%	90%	28	26	3
	93%	90%	28	26	3
	96%	87%	28	27	4
Mean	93.6%	89.4%	28	26.2	3.2
Median	93.0%	90.0%	28	26.0	3.0
STD	1.3%	1.3%	0	0.4	0.4

Features used:	Without sensor 1 and 3				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	93%	87%	28	26	4
	96%	87%	28	27	4
	96%	90%	28	27	3
	96%	90%	28	27	3
	100%	90%	28	28	3
Mean	96.2%	88.8%	28	27.0	3.4
Median	96.0%	90.0%	28	27.0	3.0
STD	2.5%	1.6%	0	0.7	0.5

Features used:	Using feature differentials				
Cluster analysis:	Kmeans, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	86%	67%	28	24	12
	86%	75%	28	24	8
	89%	66%	28	25	13
	89%	68%	28	25	12
	96%	64%	28	27	15
Mean	89.2%	68.0%	28	25.0	12.0
Median	89.0%	67.0%	28	25.0	12.0
STD	4.1%	4.2%	0	1.2	2.5

Features used:	Optimal set				
Cluster analysis:	No clustering				
PLS-DA variance:	100%				
Classifier:	Old method with prior prob. weight [1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	93%	40%	28	26	39
	93%	42%	28	26	36
	93%	43%	28	26	35
	93%	43%	28	26	34
	93%	44%	28	26	33
Mean	93.0%	42.4%	28	26.0	35.4
Median	93.0%	43.0%	28	26.0	35.0
STD	0.0%	1.5%	0	0.0	2.3

Features used:	Optimal set				
Cluster analysis:	No clustering				
PLS-DA variance:	100%				
Classifier:	Old method with prior prob. weight [1 10^{20}]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	86%	71%	28	24	10
	86%	71%	28	24	10
	86%	71%	28	24	10
	86%	71%	28	24	10
	89%	71%	28	25	10
Mean	86.6%	71.0%	28	24.2	10.0
Median	86.0%	71.0%	28	24.0	10.0
STD	1.3%	0.0%	0	0.4	0.0

A.4 Patient F2

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	77%	17	17	5
	100%	77%	17	17	5
	100%	74%	17	17	6
	94%	80%	17	16	4
	82%	82%	17	14	3
Mean	95.2%	78.0%	17	16.2	4.6
Median	100.0%	77.0%	17	17.0	5.0
STD	7.8%	3.1%	0	1.3	1.1

Features used:	All features				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	94%	17	17	1
	100%	85%	17	17	3
	100%	81%	17	17	4
	94%	94%	17	16	1
	94%	84%	17	16	3
Mean	97.6%	87.6%	17	16.6	2.4
Median	100.0%	85.0%	17	17.0	3.0
STD	3.3%	6.0%	0	0.5	1.3

Features used:	All features				
Cluster analysis:	K means, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	67%	17	14	7
	82%	56%	17	14	11
	82%	50%	17	14	14
	76%	50%	17	13	13
	76%	48%	17	13	14
Mean	79.6%	54.2%	17	13.6	11.8
Median	82.0%	50.0%	17	14.0	13.0
STD	3.3%	7.8%	0	0.5	2.9

Features used:	All features				
Cluster analysis:	K means, 5 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	88%	60%	17	15	10
	82%	64%	17	14	8
	71%	57%	17	12	9
	65%	69%	17	11	5
	59%	50%	17	10	10
Mean	73.0%	60.0%	17	12.4	8.4
Median	71.0%	60.0%	17	12.0	9.0
STD	11.9%	7.2%	0	2.1	2.1

Features used:	All features				
Cluster analysis:	Max values				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	41%	17	14	20
	76%	43%	17	13	17
	76%	43%	17	13	17
	76%	42%	17	13	18
	76%	41%	17	13	19
Mean	77.2%	42.0%	17	13.2	18.2
Median	76.0%	42.0%	17	13.0	18.0
STD	2.7%	1.0%	0	0.4	1.3

Features used:	All features				
Cluster analysis:	Min values				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	88%	38%	17	15	25
	82%	38%	17	14	23
	82%	37%	17	14	24
	76%	35%	17	13	24
	76%	33%	17	13	26
Mean	80.8%	36.2%	17	13.8	24.4
Median	82.0%	37.0%	17	14.0	24.0
STD	5.0%	2.2%	0	0.8	1.1

Features used:	All features				
Cluster analysis:	Gm hard, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	85%	17	17	3
	100%	85%	17	17	3
	100%	81%	17	17	4
	94%	84%	17	16	3
	88%	71%	17	15	6
Mean	96.4%	81.2%	17	16.4	3.8
Median	100.0%	84.0%	17	17.0	3.0
STD	5.4%	5.9%	0	0.9	1.3

Features used:	All features				
Cluster analysis:	Gm hard, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	89%	17	17	2
	100%	85%	17	17	3
	100%	85%	17	17	3
	94%	84%	17	16	3
	94%	73%	17	16	6
Mean	97.6%	83.2%	17	16.6	3.4
Median	100.0%	85.0%	17	17.0	3.0
STD	3.3%	6.0%	0	0.5	1.5

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	80%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	77%	17	17	5
	94%	100%	17	16	0
	94%	80%	17	16	4
	94%	76%	17	16	5
	88%	79%	17	15	4
Mean	94.0%	82.4%	17	16.0	3.6
Median	94.0%	79.0%	17	16.0	4.0
STD	4.2%	10.0%	0	0.7	2.1

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	50%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	94%	17	17	1
	94%	89%	17	16	2
	94%	84%	17	16	3
	94%	62%	17	16	10
	82%	82%	17	14	3
Mean	92.8%	82.2%	17	15.8	3.8
Median	94.0%	84.0%	17	16.0	3.0
STD	6.6%	12.2%	0	1.1	3.6

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 5				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	81%	17	17	4
	94%	84%	17	16	3
	94%	80%	17	16	4
	88%	79%	17	15	4
	88%	75%	17	15	5
Mean	92.8%	79.8%	17	15.8	4.0
Median	94.0%	80.0%	17	16.0	4.0
STD	5.0%	3.3%	0	0.8	0.7

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 1				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	94%	84%	17	16	3
	94%	80%	17	16	4
	94%	80%	17	16	4
	94%	76%	17	16	5
	88%	88%	17	15	2
Mean	92.8%	81.6%	17	15.8	3.6
Median	94.0%	80.0%	17	16.0	4.0
STD	2.7%	4.6%	0	0.4	1.1

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights[1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	85%	17	17	3
	100%	85%	17	17	3
	94%	89%	17	16	2
	94%	76%	17	16	5
	94%	73%	17	16	6
Mean	96.4%	81.6%	17	16.4	3.8
Median	94.0%	85.0%	17	16.0	3.0
STD	3.3%	6.8%	0	0.5	1.6

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights[10 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	81%	17	17	4
	94%	84%	17	16	3
	94%	76%	17	16	5
	94%	73%	17	16	6
	88%	75%	17	15	5
Mean	94.0%	77.8%	17	16.0	4.6
Median	94.0%	76.0%	17	16.0	5.0
STD	4.2%	4.5%	0	0.7	1.1

Features used:	All features				
Cluster analysis:	Gm fuzzy, 4 clusters				
PLS-DA variance:	70%				
Classifier:	QDA, Prior prob. weights[1 10]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	100%	94%	17	17	1
	100%	85%	17	17	3
	100%	85%	17	17	3
	88%	88%	17	15	2
	88%	83%	17	15	3
Mean	95.2%	87.0%	17	16.2	2.4
Median	100.0%	85.0%	17	17.0	3.0
STD	6.6%	4.3%	0	1.1	0.9

Features used:	Without VM and MAMD				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	70%	17	14	6
	88%	83%	17	15	3
	88%	94%	17	15	1
	94%	73%	17	16	6
	100%	81%	17	17	4
Mean	90.4%	80.2%	17	15.4	4.0
Median	88.0%	81.0%	17	15.0	4.0
STD	6.8%	9.4%	0	1.1	2.1

Features used:	Without SMA and MAMD				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	94%	76%	17	16	5
	94%	80%	17	16	4
	100%	74%	17	17	6
	100%	81%	17	17	4
	100%	81%	17	17	4
Mean	97.6%	78.4%	17	16.6	4.6
Median	100.0%	80.0%	17	17.0	4.0
STD	3.3%	3.2%	0	0.5	0.9

Features used:	Without SMA and VM				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	88%	83%	17	15	3
	100%	81%	17	17	4
	100%	85%	17	17	3
	100%	89%	17	17	2
	94%	80%	17	16	4
Mean	96.4%	83.6%	17	16.4	3.2
Median	100.0%	83.0%	17	17.0	3.0
STD	5.4%	3.6%	0	0.9	0.8

Features used:	Without DC				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	56%	17	14	11
	88%	58%	17	15	11
	88%	65%	17	15	8
	88%	68%	17	15	7
	100%	57%	17	17	13
Mean	89.2%	60.8%	17	15.2	10.0
Median	88.0%	58.0%	17	15.0	11.0
STD	6.6%	5.4%	0	1.1	2.4

Features used:	Without CORR				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	88%	83%	17	15	3
	94%	76%	17	16	5
	94%	84%	17	16	3
	94%	89%	17	16	2
	100%	85%	17	17	3
Mean	94.0%	83.4%	17	16.0	3.2
Median	94.0%	84.0%	17	16.0	3.0
STD	4.2%	4.7%	0	0.7	1.1

Features used:	Without PER				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	88%	71%	17	15	6
	94%	80%	17	16	4
	94%	84%	17	16	3
	94%	84%	17	16	3
	100%	81%	17	17	4
Mean	94.0%	80.0%	17	16.0	4.0
Median	94.0%	81.0%	17	16.0	4.0
STD	4.2%	5.3%	0	0.7	1.2

Features used:	Without FREQ				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	67%	17	14	7
	88%	63%	17	15	9
	88%	63%	17	15	9
	88%	68%	17	15	7
	100%	63%	17	17	10
Mean	89.2%	64.8%	17	15.2	8.4
Median	88.0%	63.0%	17	15.0	9.0
STD	6.6%	2.5%	0	1.1	1.3

Features used:	Without highest FREQ				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	88%	79%	17	15	4
	88%	94%	17	15	1
	94%	84%	17	16	3
	94%	94%	17	16	1
	100%	89%	17	17	2
Mean	92.8%	88.0%	17	15.8	2.2
Median	94.0%	89.0%	17	16.0	2.0
STD	5.0%	6.5%	0	0.8	1.3

Features used:	Without sensor 1				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	88%	83%	17	15	3
	94%	59%	17	16	11
	94%	67%	17	16	8
	100%	71%	17	17	7
	100%	77%	17	17	5
Mean	95.2%	71.4%	17	16.2	6.8
Median	94.0%	71.0%	17	16.0	7.0
STD	5.0%	9.2%	0	0.8	3.0

Features used:	Without sensor 2				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	58%	17	14	10
	82%	64%	17	14	8
	88%	58%	17	15	11
	88%	68%	17	15	7
	88%	68%	17	15	7
Mean	85.6%	63.2%	17	14.6	8.6
Median	88.0%	64.0%	17	15.0	8.0
STD	3.3%	5.0%	0	0.5	1.8

Features used:	Without sensor 3				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	76%	48%	17	13	14
	88%	48%	17	15	16
	88%	54%	17	15	13
	94%	52%	17	16	15
	94%	59%	17	16	11
Mean	88.0%	52.2%	17	15.0	13.8
Median	88.0%	52.0%	17	15.0	14.0
STD	7.3%	4.6%	0	1.2	1.9

Features used:	Without sensor 1 and 2				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	88%	45%	17	15	18
	88%	50%	17	15	15
	88%	56%	17	15	12
	94%	53%	17	16	14
	100%	52%	17	17	16
Mean	91.6%	51.2%	17	15.6	15.0
Median	88.0%	52.0%	17	15.0	15.0
STD	5.4%	4.1%	0	0.9	2.2

Features used:	Without sensor 2 and 3				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	38%	17	14	23
	82%	45%	17	14	17
	88%	37%	17	15	26
	88%	41%	17	15	22
	100%	43%	17	17	23
Mean	88.0%	40.8%	17	15.0	22.2
Median	88.0%	41.0%	17	15.0	23.0
STD	7.3%	3.3%	0	1.2	3.3

Features used:	Without sensor 1 and 3				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	40%	17	14	21
	88%	47%	17	15	17
	94%	46%	17	16	19
	94%	46%	17	16	19
	94%	53%	17	16	14
Mean	90.4%	46.4%	17	15.4	18.0
Median	94.0%	46.0%	17	16.0	19.0
STD	5.4%	4.6%	0	0.9	2.6

Features used:	Using feature differentials				
Cluster analysis:	Gm fuzzy, 3 clusters				
PLS-DA variance:	70%				
Classifier:	KNN, k = 3				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	56%	17	14	11
	88%	58%	17	15	11
	88%	58%	17	15	11
	88%	68%	17	15	7
	94%	62%	17	16	10
Mean	88.0%	60.4%	17	15.0	10.0
Median	88.0%	58.0%	17	15.0	11.0
STD	4.2%	4.8%	0	0.7	1.7

Features used:	Optimal set				
Cluster analysis:	No clustering				
PLS-DA variance:	100%				
Classifier:	Old method, Prior prob. weight [1 1]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	71%	11%	17	12	95
	76%	12%	17	13	99
	76%	12%	17	13	93
	76%	12%	17	13	93
	82%	13%	17	14	91
Mean	76.2%	12.0%	17	13.0	94.2
Median	76.0%	12.0%	17	13.0	93.0
STD	3.9%	0.7%	0	0.7	3.0

Features used:	Optimal set				
Cluster analysis:	No clustering				
PLS-DA variance:	100%				
Classifier:	Old method, Prior prob. weight [1 10^{20}]				
	Sensitivity	Selectivity	Num seiz	Found seiz	Found non seiz
	82%	36%	17	14	25
	82%	38%	17	14	23
	82%	38%	17	14	23
	88%	38%	17	15	24
	88%	39%	17	15	23
Mean	84.4%	37.8%	17	14.4	23.6
Median	82.0%	38.0%	17	14.0	23.0
STD	3.3%	1.1%	0	0.5	0.9

Appendix B

Optimal sets for old method

This chapter provide the results from the old method (see sec. 1.3) evaluated using the old evaluation method used in [3].

This evaluation were a little different from the evaluation performed in this thesis. The differences in evaluation methods depends partly on the fact that the older method did not make direct use of time evolutionary behaviors and therefore the data was not as scarce since each time instance inside a seizure was seen as an independent observation.

First of all, the evaluation utilized re-substitution instead of cross-validation as was used in this thesis (see sec. 3.5). The use of re-substitution introduced an optimistic bias in the results which was believed to be so small that it did not matter. According to the results from this thesis, however, it seems like the bias had a large impact on the results (compare table B.1 with the tables from the old method in sec. 5.2).

Secondly the measures of the results were a little bit different. Sensitivity was defined equally as in this thesis but instead of selectivity a measure called ”false alarm” was used. ”false alarm” was a measure of the average number of false alarm per day that occurred during the classification of the data.

In the thesis, a study of the features that was most important for good classifications were performed. The results of this study on the four patients that have been used in this thesis can be seen below. These results have yielded the feature sets which are used in evaluation of the old method in chapter 5.

B.1 Patient 7

No	Sensitivity	False alarm	Feature #	Feature name
1	0.75	1.08	24	Frequency band 3.75 - 5.25 Hz, Sensor 1
2	0.75	1.08	3	DC, Z axis, Sensor 1
3	0.75	0.54	1	DC, X axis, Sensor 1
4	0.75	0.54	5	DC, Y axis, Sensor 2
5	0.83	0.54	40	Linear correlation, Sensors 1 & 2
6	0.83	1.08	4	DC, X axis, Sensor 2
7	0.83	2.16	44	Linear correlation, Sensors 2 & 3
8	0.92	7.55	26	Frequency band 8.25 - 13.25 Hz, Sensor 1
9	0.83	5.93	25	Frequency band 5.25 - 8.25 Hz, Sensor 1
10	0.92	7.01	39	Frequency band 13.25 - 25 Hz, Sensor 3
11	0.92	8.09	8	DC, Y axis, Sensor 3
12	0.92	8.63	7	DC, X axis, Sensor 3
13	0.92	8.09	6	DC, Z axis, Sensor 2
14	0.92	9.17	9	DC, Z axis, Sensor 3
15	0.92	9.70	37	Frequency band 5.25 - 8.25 Hz, Sensor 3
16	0.92	9.70	46	Linear correlation, Sensor 1 with itself
17	0.92	11.32	48	Linear correlation, Sensor 2 with itself
18	0.92	12.40	49	Circular correlation, Sensor 2 with itself
19	0.92	13.48	27	Frequency band 13.25 - 25 Hz, Sensor 1
20	0.92	14.56	38	Frequency band 8.25 - 13.25 Hz, Sensor 3
21	0.92	15.10	36	Frequency band 3.75 - 5.25 Hz, Sensor 3
22	0.92	16.17	50	Linear correlation, Sensor 3 with itself
23	0.92	16.71	15	VM, Sensor 3
24	0.92	18.87	20	Periodicity, Sensor 2
25	0.92	22.11	41	Circular correlation, Sensors 1 & 2
26	0.92	26.42	35	Frequency band 2.25 - 3.75 Hz, Sensor 3
27	0.92	22.64	2	DC, Y axis, Sensor 1
28	0.92	22.11	34	Frequency band 0.75 - 2.25 Hz, Sensor 3
29	0.92	24.26	18	MAMD, Sensor 3
30	0.92	25.34	45	Circular correlation, Sensors 2 & 3

B.2 Patient 14

No	Sensitivity	False alarm	Feature #	Feature name
1	0.23	50.85	11	SMA, Sensor 2
2	0.32	14.37	2	DC, Y axis, Sensor 1
3	0.64	20.82	45	Circular correlation, Sensors 2 & 3
4	0.64	14.37	9	DC, Z axis, Sensor 3
5	0.82	16.79	6	DC, Z axis, Sensor 2
6	0.82	14.85	19	Periodicity, Sensor 1
7	0.77	12.27	7	DC, X axis, Sensor 3
8	0.86	14.20	15	VM, Sensor 3
9	0.86	13.40	44	Linear correlation, Sensors 2 & 3
10	0.86	15.66	14	VM, Sensor 2
11	0.86	15.01	23	Frequency band 2.25 - 3.75 Hz, Sensor 1
12	0.86	14.85	18	MAMD, Sensor 3
13	0.86	16.79	42	Linear correlation, Sensors 1 & 3
14	0.86	18.24	21	Periodicity, Sensor 3
15	0.91	21.31	1	DC, X axis, Sensor 1
16	0.91	36.32	8	DC, Y axis, Sensor 3
17	0.91	36.80	34	Frequency band 0.75 - 2.25 Hz, Sensor 3
18	0.95	57.47	5	DC, Y axis, Sensor 2
19	1.00	73.12	47	Circular correlation, Sensor 1 with itself
20	1.00	78.93	20	Periodicity, Sensor 2
21	0.91	68.44	4	DC, X axis, Sensor 2
22	0.95	79.26	35	Frequency band 2.25 - 3.75 Hz, Sensor 3
23	0.95	84.75	28	Frequency band 0.75 - 2.25 Hz, Sensor 2
24	1.00	134.30	43	Circular correlation, Sensors 1 & 3
25	0.95	134.62	51	Circular correlation, Sensor 3 with itself
26	1.00	144.63	10	SMA, Sensor 1
27	0.95	112.35	50	Linear correlation, Sensor 3 with itself
28	0.95	141.08	29	Frequency band 2.25 - 3.75 Hz, Sensor 2
29	1.00	207.91	41	Circular correlation, Sensors 1 & 2
30	1.00	179.34	46	Linear correlation, Sensor 1 with itself

B.3 Patient F1

No	Sensitivity	False alarm	Feature #	Feature name
1	0.61	7.21	29	Frequency band 2.25 - 3.75 Hz, Sensor 2
2	0.71	9.61	47	Circular correlation, Sensor 1 with itself
3	0.71	8.65	46	Linear correlation, Sensor 1 with itself
4	0.71	7.21	5	DC, Y axis, Sensor 2
5	0.68	6.73	3	DC, Z axis, Sensor 1
6	0.71	5.77	8	DC, Y axis, Sensor 3
7	0.71	5.29	20	Periodicity, Sensor 2
8	0.71	5.29	40	Linear correlation, Sensors 1 & 2
9	0.71	6.25	6	DC, Z axis, Sensor 2
10	0.71	6.25	48	Linear correlation, Sensor 2 with itself
11	0.71	7.21	49	Circular correlation, Sensor 2 with itself
12	0.79	10.09	43	Circular correlation, Sensors 1 & 3
13	0.79	13.46	50	Linear correlation, Sensor 3 with itself
14	0.79	13.46	45	Circular correlation, Sensors 2 & 3
15	0.86	14.90	9	DC, Z axis, Sensor 3
16	0.82	13.94	42	Linear correlation, Sensors 1 & 3
17	0.82	13.46	41	Circular correlation, Sensors 1 & 2
18	0.96	49.99	13	VM, Sensor 1
19	0.96	53.84	7	DC, X axis, Sensor 3
20	0.93	62.49	23	Frequency band 2.25 - 3.75 Hz, Sensor 1
21	1.00	64.89	27	Frequency band 13.25 - 25 Hz, Sensor 1
22	1.00	68.74	51	Circular correlation, Sensor 3 with itself
23	1.00	71.14	21	Periodicity, Sensor 3
24	1.00	71.62	16	MAMD, Sensor 1
25	1.00	70.18	44	Linear correlation, Sensors 2 & 3
26	0.96	49.99	35	Frequency band 2.25 - 3.75 Hz, Sensor 3
27	1.00	71.62	37	Frequency band 5.25 - 8.25 Hz, Sensor 3
28	1.00	121.61	26	Frequency band 8.25 - 13.25 Hz, Sensor 1
29	1.00	89.89	33	Frequency band 13.25 - 25 Hz, Sensor 2
30	1.00	173.05	22	Frequency band 0.75 - 2.25 Hz, Sensor 1

B.4 Patient F2

No	Sensitivity	False alarm	Feature #	Feature name
1	0.06	4.14	19	Periodicity, Sensor 1
2	0.29	26.89	36	Frequency band 3.75 - 5.25 Hz, Sensor 3
3	0.35	17.24	30	Frequency band 3.75 - 5.25 Hz, Sensor 2
4	0.59	35.86	45	Circular correlation, Sensors 2 & 3
5	0.53	35.86	37	Frequency band 5.25 - 8.25 Hz, Sensor 3
6	0.53	36.55	14	VM, Sensor 2
7	0.65	51.03	40	Linear correlation, Sensors 1 & 2
8	0.59	59.99	35	Frequency band 2.25 - 3.75 Hz, Sensor 3
9	0.59	35.86	38	Frequency band 8.25 - 13.25 Hz, Sensor 3
10	0.65	32.41	5	DC, Y axis, Sensor 2
11	0.65	43.44	46	Linear correlation, Sensor 1 with itself
12	0.88	100.68	7	DC, X axis, Sensor 3
13	0.88	124.12	49	Circular correlation, Sensor 2 with itself
14	0.94	103.43	29	Frequency band 2.25 - 3.75 Hz, Sensor 2
15	1.00	114.47	47	Circular correlation, Sensor 1 with itself
16	1.00	131.71	4	DC, X axis, Sensor 2
17	1.00	188.94	43	Circular correlation, Sensors 1 & 3
18	1.00	168.25	22	Frequency band 0.75 - 2.25 Hz, Sensor 1
19	1.00	195.15	42	Linear correlation, Sensors 1 & 3
20	1.00	259.96	50	Linear correlation, Sensor 3 with itself
21	1.00	278.58	16	MAMD, Sensor 1
22	1.00	326.16	41	Circular correlation, Sensors 1 & 2
23	0.88	99.99	39	Frequency band 13.25 - 25 Hz, Sensor 3
24	1.00	450.28	1	DC, X axis, Sensor 1
25	1.00	448.22	13	VM, Sensor 1
26	1.00	438.56	20	Periodicity, Sensor 2
27	1.00	325.47	24	Frequency band 3.75 - 5.25 Hz, Sensor 1
28	1.00	479.93	48	Linear correlation, Sensor 2 with itself
29	1.00	411.67	44	Linear correlation, Sensors 2 & 3
30	1.00	509.59	51	Circular correlation, Sensor 3 with itself