Investigation of feature selection optimization for EEG signal analysis for monitoring a driver

Author: Stefan Danielsson
sdn08003@student.mdh.se

Examiner
Shaina Begum
shahina.begum@mdh.se

Supervisor
Mobyen Uddin Ahmed
mobyen.ahmed@mdh.se
Abstract

Electroencephalogram (EEG) is a well known, and well used method for studying brain activity, and it's possibilities have lately stretched into the car industry, were it's capabilities of detecting sleepiness in drivers are currently being put to the test. When performing EEG signal analysis on the brain, standardized signal bands exists that are characteristic to specific states of mind, such as when a driver is feeling sleepy. However, EEG as a method for studying the brain has major problems. The signal contains a lot of information that can be redundant or irrelevant, and the result is easily influenced and deviant by other parameters, that can cause incorrectness and inaccuracy in the final prediction and classification of the signal frame. One of the important methods for reducing this inaccuracy of EEG, and also reducing the computational cost of the diagnose, is feature selection. Finding key features in the signal, that can support a reliable diagnosis of a specific state of mind, is of great importance. Especially since learning systems, incorrectly predicting or interpreting a signal in the classification stage, can lead to incorrect triggering of safety features in futuristic cars, such as cruiser control. There are many existing feature selection algorithms available, and features that has been tried in different research project. The goal of this research was to help gather more accurate inputs from EEG, through an optimization study, and to increase the reliability of EEG. And by doing so, hopefully improve safety systems in cars, that in turn could help preventing sleepiness-related accidents on roads in the future. This was realized through a study of features, and feature selection algorithms. By determining key features that could distinguish sleepiness from a signal, as well as performing accuracy tests for different feature selection algorithms, the motivation for an optimal selection, based on the used parameters, could be made. However limited this research was, it concluded that Information Gain as a method for selecting features, was the most accurate algorithm, and that some features were better to use then others, such as Huguchi's fractal dimension, and the Hjorth complexity.

Mälardalens högskola, Box 883, 721 23 Västerås

15-11-04

2/40
# List of contents

Abstract .................................................................................................................................................. 2

Acknowledgments .................................................................................................................................. 6

1. Introduction ........................................................................................................................................... 6
   1.1 Problem formulation ....................................................................................................................... 6
   1.2 Goals .................................................................................................................................................. 7
   1.3 Research questions ....................................................................................................................... 7

2. Background ............................................................................................................................................ 8
   2.1 Theory .............................................................................................................................................. 8
       The scientific phenomenon of EEG ............................................................................................... 8
       How electroencephalogram is generated and measured ................................................................. 8
       What electroencephalogram is ......................................................................................................... 8
       Pre-processing and artifact removal ............................................................................................... 9
       EEG processing and interpretation ................................................................................................. 9
   Feature Extraction ............................................................................................................................. 9
   Feature selection ............................................................................................................................... 10
   Change of domain – access new features ......................................................................................... 10
   Sleepiness ........................................................................................................................................... 11
   Predicting sleepiness through learning systems ............................................................................... 12
   Why using EEG? ................................................................................................................................... 13
   2.2 State of the art ............................................................................................................................... 14

4. Method ................................................................................................................................................ 17
   4.1 Approach ....................................................................................................................................... 17
   4.2 Design ........................................................................................................................................... 18
   4.3 The classification algorithm ........................................................................................................ 20
   4.4 Feature selection .......................................................................................................................... 20

5. Materials and tools ............................................................................................................................. 22

6. Implementation ..................................................................................................................................... 24

7. Evaluation ........................................................................................................................................... 25
List of figures

Illustration 1: Illustration showing how Fourier transformations can be used to decompose any signal in the time domain (red), into its individual frequency components (blue graph).

Source: http://i.stack.imgur.com/Y5EAf.png

Illustration 2: Example of plot of a test signal in MATLAB, where the time domain of a signal, has been transformed into its corresponding frequency domain. This transform gives access to, and enlightens, new features for extraction. Here, the magnitude of each frequency band is clearly visible, that was hard to extract from the time domain of the signal.

Illustration 3: Example of standard brain wave patterns, that can be observed.

Illustration 4: The overall design of the feature optimization process in this research.

Illustration 5: The 10-20 international system for EEG electrode setup on a scalp.

Illustration 6: MATLAB program flow.

Illustration 7: Table showing the average accuracy of each individual feature.

Illustration 8: Diagram showing the average accuracy of each individual feature. (Red) less than 60 %, (Yellow) less than 62 %, (Blue) less than 63% and (Green) greater than 63 %.

Illustration 9: Table showing the average accuracy of selection methods.


Illustration 11: Illustration: Table showing the average accuracy of the manual selection methods and the best selection algorithm.

Illustration 12: Diagram showing the average accuracy of the manual selection methods and the best selection algorithm.

Illustration 13: Table of the relative error of the reduced manual selection, and the best selection algorithm.

Illustration 14: Diagram visualizing the relative error for the reduced manual selections, and the best selection algorithm.
Acknowledgments

This research was a part of the VMD-project\(^1\), and it took place at Mälardalens Högskola. The project aims to “strengthen the competitiveness of the Swedish automotive industry, and provide tools for development of active safety systems”. Special thanks to the following collaborators, for making this project possible:

- Volvo Vehicles AB
- VTI - The Swedish research institution for roads and transportation
- The Academy for Innovation, Design and Science - Mälardalens högskola

1. Introduction

The concept of using EEG signals to study the brain, has been a wide field of study since the middle of the 20th century, and it has been used widely medically, to diagnose brain related diseases such as epilepsy, sleeping disorders and to determine coma or brain death. [1][2]. But there has also been studies about the possibilities of using EEG signal analysis to detect sleepiness in drivers[3][4]. In this section, the problem of using EEG analysis on the driver's brain, and the problem this research tries to solve, will be identified, and the overall goals of the research will be mentioned. In order to reach the goals, some research questions has been formulated, and they can also be viewed in this section.

1.1 Problem formulation

In 2005, VTI claimed that about 10 % of all car accidents, with lethal outcome in Sweden, were the results of drivers, driving under the effect of sleepiness. In accordance with the Swedish “Vision Zero project”\(^2\), the total number of car accidents with lethal outcome, has fortunately been greatly reduced\(^3\) by nearly 47 %, since it's introduction in 1995. But in order to reduce the number of car accidents per year even further, active safety systems might be required to be implemented in cars and trucks, traveling on Swedish roads.

To prevent this type of accidents, this research project focused on studying the possibilities of using EEG signal analysis on the driver's brain. The EEG signal however contains huge amounts of information, and the characteristics of the signal are usually stochastic, and can consist of both continuous patterns, as well as discontinuous patterns. So in order to be able to obtain useful data from the EEG signal, and make a proper diagnose from it, machine learning algorithms, learning systems, and numerical methods have been introduced in order to refine, pattern match, and simplify the signal interpretation, when an EEG signal specialist isn't an option. These methods and algorithms are crucial for the accuracy of the EEG, and one of these important concepts, that increases the credibility and accuracy, is feature selection. It focuses on reducing computational the

---


Mälardalens högskola, Box 883, 721 23 Västerås

15-11-04
cost of the total diagnose calculation, as well as removing unnecessary features, that are either irrelevant, or redundant for the diagnose.

1.2 Goals

The goal of this research is to suggest an optimal algorithm for selection of features, to be used when interpreting artifact-free electroencephalogram data. This optimization will be done by performing a research study, to see what features that have been used for classification of EEG earlier, and to see what kind of selection algorithms that have been implemented in supervised learning system. To actually see the potential of these factors, MATLAB will be used to generate prediction models from sets of training data. By performing accuracy measurements on this model, the ability to distinguish sleepiness from a signal can be correlated to the features used, and the algorithm that selected them.

1.3 Research questions

Existing methods and features

- What features have been used for classifying EEG signals in earlier research?
- What feature selection algorithms have been used for classifying EEG signals in earlier research?

Possible optimizations:

- How closely related to sleepiness are the features, and how good are the features for distinguishing sleepiness from the signal?
- Can the number of features be reduced, and yet keep the prediction accuracy?
- What is the most optimal selection of features?
- What is the most optimal feature selection algorithm?
2. Background

In this section, the overall and necessary theory needed to understand this research will be listed. The background is a review of the electroencephalogram (EEG), stating what it is, how it works, and why it is being used. The review will hopefully also future clarify, and give an overview over the specific task, to be done in this research. After the theory section, a SOTA can be viewed, were features and feature selection algorithms identified from the literature study can be found.

2.1 Theory

The scientific phenomenon of EEG

In short, neuroscience states that the human brain consists of approximately 86 billion neurons, also called brain cells. The neurons are able to communicate by sending electrical impulses, triggered by chemical reactions in the cells. The neurons are constantly active, even when a person is asleep, and the process could naively be considered as the brain “thinking” or reacting on events in the body.

The interaction and communication between neurons are possible through the help of specialized connections between the neurons, known as synapses. Neurons triggering electrical signals repeatedly, or repeated exchange of electrical signal in between several neurons, causes oscillating activity in the brain. This phenomenon is know as neuron oscillation [5].

How electroencephalogram is generated and measured

The neuron oscillation can be measured by using patterns of electrodes arranged in pairs, attached to a human scalp. The difference in electrical potential between each pair can then be registered, and amplified. However, the electrical potential generated by an individual neuron, is far too small to be measured. So in order to register a measurable potential between a pair of electrodes, summation of the synchronous activity of millions of neurons might be needed. In other words, neurons must have similar purpose and spatial orientations in order to generate electrical impulses strong enough to be detectable by the instrument [6].

What electroencephalogram is

Electroencephalogram, or short EEG, is the resulting measured signal, from the summation and generalization of the electric potentials from the electrodes. The recorded potential between two electrodes is referred to as a channel. One standardized system for the placement and wiring of the electrodes are called the international 10-20 system.

This signal channel can be observed as microscopic oscillations, or frequencies, that are standardized into frequency bands. The most interesting one's for this research are known as the delta-, theta-, alpha-, beta and gamma frequency bands. These frequency bands lies between 0-32 Hz [6][7].
Pre-processing and artifact removal

The signal is usually pre-processed through different noise filters, before any kind of analysis of the signal takes place, such as low-, high- and band-pass filters. There are also specific methods for dealing with faulty signals due to recorded activity, that is not of cerebral origin. These types of noise are known as artifacts, and some examples of artifacts could be faulty recordings due to broken EEG electrodes, blinking and movement of eyes, dried electrodes, stretching of muscles, excessive electrode gel, etc. Algorithms in learning systems exists, in order to learn and remove specific subject's brains contribution of artifacts[8].

EEG processing and interpretation

The next step in processing the EEG signal, is the step of transforming the artifact-free EEG signal into relevant data, leading to a better human interpretation. This relevant data, is commonly referred to as a desired feature. A feature is an individual measurable property of a phenomenon being observed, and they are commonly displayed as vectors, an n-dimensional vector representation of a numerical feature. The vector space associated with these vectors is often called the feature space or the domain. In other words, the feature, is a derived set of data from the initial set of data, or a unique property, that is distinguishable from a data set (the particular measurable property we are interested in). In the case of this research project, the data is represented by the recorded EEG signal itself, and the feature we are looking for, is the sub data or data set, that can tell us something about the mental state of the driver. This sub data must be located, wisely selected, and be used to create interpolated graphical representations of the neuron oscillation in the brain. In other words, the goal is to map data to a lower dimensional space, in order to discard uninformative variance in the data set, or in order to detect the subspace, were the desired feature lies. This is a crucial step in terms of time consumption and memory requirements, and in order to visualize the data for better human interpretation. This step is often generalized through data analysis, using numerical algorithms and methods, and consists of two parts; feature extraction and feature selection [9][10].

Feature Extraction

Feature extraction is the process of reducing the amount of data required to describe a large set of data, by transforming the initial data set into a new and approximated data set. When analyzing huge amounts of EEG data, one of the major problems comes from the fact that the initial data set, have a huge amount of variables involved. Analysis with a large number of variables generally have larger complexity, and can even contribute to higher inaccuracy.

So feature extraction is a methods that changes and constructs new combinations of the initial data set, that are smaller, in order to get around the problem of huge data sets, and large computational complexity, and yet generating models with sufficient accuracy[10]. In addition, the extraction process also involves changing the domain of the signal in order to access other features, that were not available or visible in the original time domain. The signal can be viewed in several domains, such as time- and, frequency domain. The process of transforming the signals domain in known as time-frequency transformations.
Feature selection

Feature selection is the process of selecting a subset of relevant features in a feature set, and to use it for a prediction model. By doing this, the model will be simplified, and the computational cost of training the model will be reduced, and also the generalization abilities of the model and prevention of over-training will be enhanced. As mentioned above, the main problem when performing signal interpretation, is the fact that the signal contains too many features that are either redundant, or irrelevant to the final result. It can even result in decreasing the accuracy of the final model. Feature selection uses this fact in order to removed these unwanted features from the signal, or rather just selecting the ones that are relevant, without losing too much information or accuracy. The feature selection is done by looking at features or sub sets of features in the signal, and assigning them scores. Depending on the score, the feature can then be removed or saved in the final signal representation. Feature selection does not change the data set, it only picks out/selects the informative parts of the signal. There are three types of algorithms for feature selection. Filter methods, wrapper methods and embedded methods, but the features can also be selected manually by an EEG signal expert[11][10][12].

Change of domain – access new features

The EEG signals are usually represented as time series. A sequence of data points, typically consisting of successive measurements made over a time interval. Any signal that can be represented as an amplitude that varies with time has a corresponding frequency spectrum. Time-frequency transforms simplifies a signals information, by transforming and splitting a time series into its consistent components: phase and amplitude, on a frequency series/domain. This is useful because it visualizes each independent frequency wave very clearly (such as the alpha wave contribution to the signal). For this research, the Discrete Fourier Transform – Fast Fourier Transform, was used in order to change the time domain into frequency domain for additional extraction of features[13], such as PSD - Power spectral density.

Illustration 1: Illustration showing how Fourier transformations can be used to decompose any signal in the time domain (red), into it’s individual frequency components (blue graph).

Source: http://i.stack.imgur.com/Y5EAf.png
Sleepiness

It is important to separate sleepiness from fatigue, and that sleepiness is meant as the phenomenon known as micro-sleep\(^4\). Not as the kind of sleepiness that is felt when laying in bed. When are drivers at a risk of being sleepy in terms of EEG-characteristics? It is quite hard to say exactly[15]. But from studying EEG signals on human brains, it is quite clear that in the frequency bands of alpha, theta and delta, the mental state of a human have been observed to be more or less in a state of sleep. When the neuron oscillation in the brain slows down, the risk of the driver falling asleep increases. The characteristics of such frequencies, are observed as amplitudes with slow oscillation. It is well to remember, that the EEG signal interpretation can be one part of the entire active safety system in a car. In addition, a camera can be installed and check for eye blinking, and medical instruments could measure heart rate, and other physical properties to help determine if a driver is at risk of being sleepy accurately. But for EEG signals, the following frequency bands can be considered as indicators that the driver is more or less in a state of sleep[16].

• **Alpha waves** (7–13 Hz) indicate a state of mind, that could be relaxed, reflecting, or indicating closing of the eyelids. In some cases it is also considered as the pre-state of sleeping, where a person starts losing consciousness about the surroundings. In other words, alpha waves are usually the characteristics of when the safety system should react.

• **Theta waves** (4–7 Hz) indicates a very relaxed or meditative state of mind, and can also be observed in the early stages of the sleeping-cycle. The subject is not aware of it's surroundings.

• **Delta waves** (< 4 Hz) indicates a state of mind in deep and dreamless sleep, in the later stages of the sleeping-cycle.

Illustration 3: Example of standard brain wave patterns, that can be observed.

*Source: http://hubpages.com/education/What-are-Brain-Waves-Theta-Delta-Alpha-Beta-brain-waves-What-is-brainwave-entrainment*

**Predicting sleepiness through learning systems**

How can sleepiness be determined without a signal analysis expert, diagnosing the signal? Learning systems are used to make predictions on signal data, and they are mainly generalized through unsupervised, or supervised learning. In unsupervised learning, the system uses the already collected information in order to predict the next solution based on previous values and trends, and since there exists no labeled data, there is no error feedback or reward scoring that can have impact on the predicted solution. An example of such methods is regression, that uses previous values to calculate the next value. On the other hand there is supervised learning, that instead uses labeled data for training a model, that in turn can be used to classify unlabeled data [17][18].
Classifying data, is the problem of identifying to which class, or category, an observation belongs to, on the basis of a training set of data, containing observations/features, whose category membership is known/labeled. In this research, this technique is used to classify segments of a signal, into their proper class. The response class is the state of the subjects brain. Is the driver tired, or is the driver focused?

**Why using EEG?**

There are three main reasons why EEG is useful in these type of applications. Firstly, compared to other techniques for studying the brain, such as MRI (Magnetic resonance imaging) or MAG (Magnetoeencephalography), EEG is relatively tolerant of subject movement and outer disturbances, and the equipment is small, portable, and flexible. Of course, there are differences in what the different methods can do and measure, but still it is important that the method is easy to use, and that it is relatively small, since it is needed to be implemented in a car, with normal people operating the sensors[19]. Secondly, the necessary hardware to preform EEG-measurements have costs that are significantly lower the other tools for measuring brain activity. And thirdly, the method is noninvasive, meaning that the brain can be studied, without any instruments actually entering the body. In a study with the aim to evaluate commercially available devices for driver sleepiness monitoring, Sommer concluded that “adaptive signal analysis of EEG/EOG in combination with computational intelligence methods, resulted in highest detection performance” [20].

However, EEG is subject to noise and blurry signal outputs, and artifacts. Extended data analysis and signal interpretation is needed, in order to extract useful information from the signal, and a large number of test-subjects, and learning algorithms, in order to make clear assumptions about brain patterns. In cars, safety systems using a camera for watching the eyes of a driver have the advantage of being accurate. If your eyes are moving slow, or are not blinking fast enough, the system reacts. However, just because your eyes are open, does not necessarily mean that you are focused and alert when driving. EEG is at the moment not accurate enough to be implemented in mass produced cars, but it has the potential of detecting sleepiness earlier in a driver, compared to a eye-camera[21].
2.2 State of the art

In order to have redundant and fail-safe system, it is highly beneficial to have more than just one observable quantity, when determining sleepiness in a driver. AlZu'bi recommends that more than just one sleepiness monitoring system is used, in order to increase reliability of the safety system[22]. Such monitoring systems could consist of a camera for detecting signs of sleepiness visually, an EEG instrument neurologically, and a heart-rate instrument medically. Few of these systems however, have been implemented in mass produced cars, and those who have, do not apply, or do not depend on EEG. This is mainly due to high error rate in accuracy of the EEG, incorrectly triggering events in the car. The fact that brains are different, is also a contributing factor to making good and accurate readings, and learning algorithms needs to learn how the subject specific brain works, in order to give accurate decisions[23].

Until relatively recently, a problem in making trustworthy research on the subject, has been the resources necessary to test the methods in real life. These resources include special cars, engineered and modified to record EEG, and of course to react on events, such as sending out warning signals to the driver or cruiser systems, taking control over steering and/or speed of the vehicle. Other required resources are safe testing-environments were to drive the cars, and a huge number of test subjects, to mine extensive and supporting data from. Fortunately in our today's modern computer-society, relatively non expensive, and minimal-resource-demanding simulations of EEG studies, are possible through virtual environments, and can be performed on a computers in a classroom, increasing the opportunities for making large scale research on the area. For example, Fei Wang and his research-team performed EEG studies on a BCI-device (Brain-computer Interaction Device), assisted by a driving-simulator, in order to study sleepiness, and record EEG signals on students [24]. Many research teams have also built advanced simulator facilities for making extensive test. Among them are VIT, one of the collaborators for this research, who is one of the major research institutions on transportation in Sweden5.

This type of research is important, since it builds a foundation that motivates the specs and structure for how actual test-cars could be built. These test-cars can then hopefully acknowledge the research performed in virtual environments, and hopefully lead to mass production of cars with active sleepiness safety systems, using EEG in the near future. One of the available safety systems on the market today, is the one released by Volvo, who was the first car manufacturer to develop an active safety system in a car (in 2007), and it is known as DAC, Driver Alert Control6 7. However, it is not based on EEG analysis.

As mentioned before, a limiting factor of EEG is it's accuracy. There are many challenges and factors that affects the accuracy of the EEG, such as the overall design of the signal processing system, the artifact handling and cleaning of the signal, and the amount of training the learning system receives. But it is also important to select relevant features for the construction of the learning system. And that is what this research will be focusing on. Features can be found in different domains, such as in the amplitude domain, time domain, velocity domain, frequency domain and within the time-frequency domains combined.

6 http://support.volvocars.com/uk/cars/pages/owners-manual.aspx?mc=Y555&my=2015&sw=14w20&article=2e82f6fc0d1139c2c0a801e800329d4e, Last visited:2015-09-19
There are many features that has been implemented in similar research, and below some of them are listed, though it is certain that many more exists.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute Value</td>
<td>[25]</td>
</tr>
<tr>
<td>Autoregressive coefficients</td>
<td>[26]</td>
</tr>
<tr>
<td>Derivative</td>
<td>[26]</td>
</tr>
<tr>
<td>Band power</td>
<td>[27]</td>
</tr>
<tr>
<td>Fall length</td>
<td>[28]</td>
</tr>
<tr>
<td>Fall value</td>
<td>[28]</td>
</tr>
<tr>
<td>Fall velocity</td>
<td>[28]</td>
</tr>
<tr>
<td>Geometric mean</td>
<td>[29]</td>
</tr>
<tr>
<td>Half−rise length</td>
<td>[28]</td>
</tr>
<tr>
<td>Half length</td>
<td>[28]</td>
</tr>
<tr>
<td>Half−fall length</td>
<td>[28]</td>
</tr>
<tr>
<td>Half-fall velocity</td>
<td>[28]</td>
</tr>
<tr>
<td>Half-rise length</td>
<td>[28]</td>
</tr>
<tr>
<td>Half−rise velocity</td>
<td>[28]</td>
</tr>
<tr>
<td>Harmonic mean</td>
<td>[29]</td>
</tr>
<tr>
<td>Higuchi fractal dimension</td>
<td>[30]</td>
</tr>
<tr>
<td>Hjorth Complexity</td>
<td>[31][31.5]</td>
</tr>
<tr>
<td>Hjorth Mobility</td>
<td>[31][31.5]</td>
</tr>
<tr>
<td>Hurst exponent</td>
<td>[31]</td>
</tr>
<tr>
<td>Integrated amplitude</td>
<td></td>
</tr>
<tr>
<td>Interquartile range</td>
<td>[26]</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>[26][31.5]</td>
</tr>
<tr>
<td>Logarithm of energy</td>
<td>[30]</td>
</tr>
<tr>
<td>Maximum amplitude</td>
<td>[27]</td>
</tr>
<tr>
<td>Mean</td>
<td>[32][27][33][26]</td>
</tr>
<tr>
<td>Mean absolute deviation</td>
<td>[25][26]</td>
</tr>
<tr>
<td>Mean absolute value</td>
<td>[25][34]</td>
</tr>
<tr>
<td>Median</td>
<td>[32][27]</td>
</tr>
<tr>
<td>Minimum amplitude</td>
<td>[27]</td>
</tr>
<tr>
<td>Most frequent value</td>
<td></td>
</tr>
<tr>
<td>Peak amplitude</td>
<td>[28]</td>
</tr>
<tr>
<td>Peak counting</td>
<td>[26]</td>
</tr>
<tr>
<td>Peak frequency</td>
<td>[33]</td>
</tr>
<tr>
<td>Mean curve length</td>
<td>[31]</td>
</tr>
<tr>
<td>Petrosian's fractal dimension</td>
<td>[30]</td>
</tr>
<tr>
<td>Power spectral density</td>
<td>[25][33]</td>
</tr>
<tr>
<td>Rise length</td>
<td>[28]</td>
</tr>
<tr>
<td>Rise value</td>
<td>[28]</td>
</tr>
<tr>
<td>Rise velocity</td>
<td>[28]</td>
</tr>
<tr>
<td>Root Mean Square</td>
<td>[25][27][26]</td>
</tr>
<tr>
<td>Shannon entropy</td>
<td>[35]</td>
</tr>
<tr>
<td>Skewness</td>
<td>[31]</td>
</tr>
<tr>
<td>Slope sign changes</td>
<td>[34]</td>
</tr>
<tr>
<td>Spectral centroid</td>
<td>[26]</td>
</tr>
<tr>
<td>Spectral entropy</td>
<td>[26][31]</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>[32][25][27][26]</td>
</tr>
<tr>
<td>Variance</td>
<td>[32][25][31][26]</td>
</tr>
<tr>
<td>Mode</td>
<td>[32][27]</td>
</tr>
<tr>
<td>Waveform length</td>
<td>[34]</td>
</tr>
<tr>
<td>Zero Crossing Rate</td>
<td>[34][33][26]</td>
</tr>
</tbody>
</table>
Another important problem to solve when dealing with signal analysis of EEG, is how to select the features, using learning systems, when a signal analysis expert is not available to select them. Algorithms are then needed in order to pick an optimal selection of features to use, in the construction of the prediction model. Below are some of the identified algorithms, that has been used in feature selection in other research.

- Correlation Based Filter [36].
- Fast Correlation Based Filter[37].
- Chi Square Feature Evaluation[38].
- Fisher score/criterion[39].
- Gini-Index selection[40].
- Information gain[41][40].
- Kruskal-Wallis test of variance[42].
- Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation [43].
- Student distribution test [44].
- HillClimbing[45].
- Genetic algorithm[46].
- Ant collony [47].
- Kolmogorov-Smirnov test [44].
- Kullback Leibler Divergence test[44].
- Expected Cross Entropy [40].
- Weight of Evidence[40] [65].
- Principal feature analysis [48].
- Sequential feature selecion[49].
- Correlation-based Feature Selection[50].
4. Method

In this section, the method for this research will be described in detail. Among these details are the work flow and the overall development process for the research, described in it's five separate stages. Also the selected approach for the research, and how the research questions are meant to be answered, and the design for the implementation will be reviewed. This information along with a complete listing and description of all the tools and signals that were used, will give a survey of how the results of this research was obtained and realized.

4.1 Approach

The research process was sequentially executed in the following stages; Literature study, Design, Implementation, Feature selection, and lastly Evaluation. Since one of the major questions to be answered in this research, was to find out what features in an EEG signal, that related to sleepiness, a literature study was initially performed, with intention to find features that could be extracted from EEG signals, and that had been used in earlier research studies.

By looking at these features, the next challenge was then be to find out how much the individual features actually related to sleepiness. The approach for finding relationships between features and sleepiness, was realized through designing a supervised learning system, that was trained with labeled EEG data, in order to predict unlabeled EEG data into it's correct classes.

The classes the classifier predicted (the responses), were states of the subject's mind, scored with sleepiness score such as Wake, Sleep-stage 1, and Sleep-stage 2, that were provided by the database were the signal was found. This prediction score, in combination with an array containing the extracted features of an individual frames of the signal, could then be sent frame by frame, to a classifier as labeled data. The result from training the classifier on this labeled data, yielded a model of a certain accuracy. This model could have been used for classifying unlabeled data, but that was out of the scope for this research.

The accuracy of the model was realized and confirmed by a hold out validation algorithm, that calculated the potential accuracy of the trained model. This accuracy was then analyzed in order to see what selection of features that yielded the best accuracy, and what features that single-handedly gave the best accuracy and distinct properties. Several cases were taken into consideration, such as manually selecting features based on their individual accuracy, and by testing feature selection algorithms, and see what accuracy their selection could contribute to.
4.2 Design

In this phase, the design was set, and it was determined that the system would evaluate feature’s relevance and correspondence to sleepiness, through a supervised learning system, using a classifier to predict signal patterns into classes. The process from EEG signal, to classification, is illustrated in illustration 4 below. Note that this design however describes the whole process, including predicting unlabeled data. Predicting unlabeled data was not a part of this research. The design also assumes that one frame at a time is observed (how it would work in reality). In this research however, and for the sake of simplicity, frames are not received one at a time, but all frames were received instantly as one big matrix. Everything starts when a signal is received from the pre-processing stage of the EEG. It is now considered to be free from artifact. Then the signal goes into the interpretation stage, where the signal is divided into frames of a specified duration. From these frames, features from the different domains are extracted. The features are then assembled into one big matrix, that is sent to the classifier. If we are constructing the prediction model, the frames gathered would be labeled, and the labeled arrays would be sent to the classify learner, that constructs a model form the data by using a decision tree classifier. Then either manual selection, or an algorithm for feature selection could be used to get feedback on the models output accuracy.

Illustration 4: The overall design of the feature optimization process in this research.
4.3 The classification algorithm

The “Medium tree classifier” in MATLAB was selected for this study\(^8\). Decision trees classify predictors (dependent variables, or features) on their ranges (an example of range could be a feature called: “weight”, having the range from 1.3 grams, to 1.8 grams). So each feature has it's own classification tree, where the leafs of the tree are the classes the EEG frames can be predicted to, and the nodes are the ranges of the features. This classifier was selected mainly due to it's fast prediction calculation speed, and it's low memory usage\(^9\). The prediction was calculated using a 10-k-fold cross-validation algorithm\(^10\). The classification itself was performed in an interactive app environment called “Classification Learner App”, included in the Statistics and Machine Learning Toolbox for MATLAB. However, the model could be exported to the work space of MATLAB if necessary.

4.4 Feature selection

The results in this research were gathered by the help of the Classification Learner App, in the Statistics and Machine Learning Toolbox of MATLAB. It had many ways of displaying and visualizing results from the trained prediction model, such as the accuracy of the trained model, graphs, scatter plot, ROC curves, and confusion matrix diagrams. It also provided tools for feature selection, like the possibility to manually select, or deselect features to be present in the feature arrays, that were used to train the model. By selecting and deselecting features to include in the training of the prediction model, different values of accuracy were achieved.

The feature selection algorithms on the other hand were scripted directly in MATLAB. Each algorithm had 2 inputs, were the first one was the training data for it's own built in classifier, and the second input was the corresponding labels/responses for the frames. Each algorithm had it's own way of determine the importance of the features. The simplest algorithm for determining what features to select, is the sequential algorithm, that sequentially selects features, measure their individual accuracy, and then selecting another feature, until there is no improvement in prediction. The output of the algorithms were either a list of all features, were the features had been sorted by their scores, or list, containing the features that had been deemed useful.

One thing to remember though, is that in this test, compared to a real application in an actual car, the amount of training the classifier received, was probably insufficient, and the gathered signal alone, did not contain sufficient data. However, the intention of this research, was not to construct a optimal prediction model, or to use it in a real situation. The research this research focused on, was to find optimal setup of features to be used in the prediction model, in order to detect sleepiness. So no matter what accuracy the model received, it is well to remember that the actual accuracy of the model is irrelevant for this research, and that many limiting factors were contributing to it, mentioned in the limitation section below. What does matter for the sake of this research, is what features that made the accuracy to increase or decrease. The relevance of the features, were relative to the amount of training the classifier received, and the quality of the data that the features were extracted from.

\(^8\) http://se.mathworks.com/help/stats/choose-a-classifier.html#bunt0ky
\(^9\) http://se.mathworks.com/help/stats/choose-a-classifier.html
\(^10\) http://se.mathworks.com/discovery/cross-validation.html

Mälardalens högskola, Box 883, 721 23 Västerås
15-11-04
The features that were located in the literature study were reduced to the following 31 features to be used in this research [Number] [ID] [Name] [Domain]:

<table>
<thead>
<tr>
<th>Number</th>
<th>ID</th>
<th>Name</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AV</td>
<td>Absolute Value</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>2</td>
<td>MAV</td>
<td>Mean Absolute Value</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>3</td>
<td>MAD</td>
<td>Mean Absolute Deviation</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>4</td>
<td>MNF</td>
<td>Mean frequency</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>5</td>
<td>GEOMN</td>
<td>Geometric mean</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>6</td>
<td>HARMN</td>
<td>Harmonic mean</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>7</td>
<td>RMS</td>
<td>Root Mean Square</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>8</td>
<td>STD</td>
<td>Standard Deviation</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>9</td>
<td>VAR</td>
<td>Variance</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>10</td>
<td>KUR</td>
<td>Kurtosis</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>11</td>
<td>INQUR</td>
<td>Interquartile range</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>12</td>
<td>ZCR</td>
<td>Zero Crossing Rate</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>13</td>
<td>HURST</td>
<td>Hurst exponent</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>14</td>
<td>HIGUCHI</td>
<td>Higuchi Fractal Dimension</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>15</td>
<td>MFV</td>
<td>Most frequent value</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>16</td>
<td>PC</td>
<td>Peak counting</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>17</td>
<td>MAVALPHA</td>
<td>Mean Absolute Value of Alpha component</td>
<td>(Freq. domain)</td>
</tr>
<tr>
<td>18</td>
<td>MAV DELTA</td>
<td>Mean Absolute Value of Delta component</td>
<td>(Freq. domain)</td>
</tr>
<tr>
<td>19</td>
<td>MAVTHETA</td>
<td>Mean Absolute Value of Theta component</td>
<td>(Freq. domain)</td>
</tr>
<tr>
<td>20</td>
<td>MIN</td>
<td>Minimum amplitude</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>21</td>
<td>MAX</td>
<td>Maximum amplitude</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>22</td>
<td>TRAPZ</td>
<td>Integration of amplitude</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>23</td>
<td>HJORTHCOMP</td>
<td>Hjorth Complexity</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>24</td>
<td>HJORTHMOB</td>
<td>Hjorth Mobility</td>
<td>(Time domain)</td>
</tr>
<tr>
<td>25</td>
<td>BP</td>
<td>Band power</td>
<td>(Freq. domain)</td>
</tr>
<tr>
<td>26</td>
<td>TRAPZDELTA</td>
<td>Integration of Delta component</td>
<td>(Freq. domain)</td>
</tr>
<tr>
<td>27</td>
<td>TRAPZALPHA</td>
<td>Integration of Alpha component</td>
<td>(Freq. domain)</td>
</tr>
<tr>
<td>28</td>
<td>TRAPZTHETA</td>
<td>Integration of Theta component</td>
<td>(Freq. domain)</td>
</tr>
<tr>
<td>29</td>
<td>APSD</td>
<td>Alpha Power Spectrum Density</td>
<td>(Freq. domain)</td>
</tr>
<tr>
<td>30</td>
<td>DPSD</td>
<td>Delta Power Spectrum Density</td>
<td>(Freq. domain)</td>
</tr>
<tr>
<td>31</td>
<td>TPSD</td>
<td>Theta Power Spectrum Density</td>
<td>(Freq. domain)</td>
</tr>
</tbody>
</table>
The feature selection algorithms identified in the literature study were reduced to the following nine algorithms to be used and tested in this research:

1. Correlation Based Filter
2. Fast Correlation Based Filter
3. Chi Square Feature Evaluation
4. Fisher score
5. Gini-Index selection
6. Information gain
7. Kruskal-Wallis test of variance
8. Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation
9. Student distribution test

5. Materials and tools

For this research, MATLAB2015b (trial version) was used. "MATLAB® is the high-level language and interactive environment used by millions of engineers and scientists worldwide. It lets you explore and visualize ideas and collaborate across disciplines including signal and image processing, communications, control systems, and computational finance”. For more info, visit their web page11. MATLAB contained many useful extensions to realize the desired design, such as extensions for classification learning, and EEGLAB for pre-processing and selecting EEG data. MATLAB is generally well-used in signal processing and signal interpretation, and the software has extensive supporting documentation, and help forums. This goes for the selected extensions as well, even if other types of EEG software exists. Another great advantage of using MATLAB as a tool, was the fact that the pysionet databases were 100% compatible with MATLAB, enabling direct import of signals and variables directly into the MATLAB studio through the EGGLAB extension, which made the process of choosing software even simpler.

Extensions used:

- **Statistics and Machine Learning Toolbox** - "Statistics and Machine Learning Toolbox provides functions and apps to describe, analyze, and model data using statistics and machine learning”. For more info, visit their web page12.

- **Signal Processing Toolbox** - "Signal Processing Toolbox provides functions and apps to generate, measure, transform, filter, and visualize signals”. For more info, visit their web page13.

- **EEGLAB** - “EEGLAB is an interactive MATLAB toolbox for processing continuous and event-related EEG, MEG and other electrophysiological data incorporating independent component analysis (ICA), time/frequency analysis, artifact rejection, event-related statistics, and several useful modes of visualization of the averaged and single-trial data”.

---

11 http://se.mathworks.com/products/MATLAB/
12 http://se.mathworks.com/products/statistics/
13 http://se.mathworks.com/products/signal/

Mälardalens högskola, Box 883, 721 23 Västerås

15-11-04

21/40
For more info, visit their web page\textsuperscript{14}.

**Signal**

- The signal was acquired from physionet.org, though the “CAP sleep database”\textsuperscript{[51]}. The test subject was 29 year old female, suffering from narcolepsy (narco1). The following recorded channels were available and used in the research, in accordance with the 10-20 international system: CH1: Fp2-F4, CH2: F4-C4, CH3: C4-P4 and CH4: P4-O2. This system refers to the positioning and connections of recording electrodes on the subject scalp, and these positions corresponds to the middle-right-half, of the head. See illustration 5 below. Expert neurologists trained at the a sleep center provided the scoring of the sleep macro structure.

\textit{Illustration 5: Illustration 6: The 10-20 international system for EEG electrode setup on a scalp.}

\textit{Source: http://www.hindawi.com/journals/aai/2011/384169/fig2/}

**Feature selection algorithms**

- An open source library for MATLAB was used to simulate the feature selection algorithms, provided by Arizona state university\textsuperscript{[15]}. Information, instructions, and the mechanism for each algorithm was also available in this library.

\textsuperscript{14} http://scen.ucsd.edu/eeglab/
\textsuperscript{15} http://featureselection.asu.edu/software.php

Mälardalens högskola, Box 883, 721 23 Västerås
15-11-04

22/40
6. Implementation

This section describes how the learning system, for classifying EEG, was implemented in MATLAB. First, labeled segments of the original signal was extracted, using EEGLAB. For example, if the signal was scored “wake” in the time interval 20 seconds to 5 minutes and 30 seconds, this section could be extracted as a “wake-segment”. The particular segments of interests for this research, were segments of the signal that had been scored “Wake”, “Sleep-stage 1”, or “Sleep-stage 2”. These segments were extracted out of four separate EEG channels, for the same time intervals.

Next, MATLAB code was written to decompose each of these segments of the signal, into one second frames (each containing 512 samples), to be used as samples, that the classifier could learn from. The original signal was recorded at a frame rate of 512 Hz, so each frame was designed to contain 512 samples. So in total, 2298 labeled seconds (frames) were extracted from the original signal, covering about 40 minutes of EEG. Each frame in turn then had it's features extracted and saved into arrays. These arrays were then put together into one big matrix, were each feature had it's corresponding column, and each frame had it's corresponding row.

The arrays were then sent to the classifier. The classifier uses features (called predictors), in order to train and produce a prediction model, and guess the label of the data (called the response). The response variable was included as a feature in the matrix that was sent to the classifier, and corresponded to the sleep-stage, which that particular frame corresponded to. Below in illustration 6, this program flow in MATLAB is visualized.
7. Evaluation

In this section, the practical tests made in this research are described in detail. These tests were constructed in a way, so that they would answer the research questions as clearly as possible. The results of these tests can be found in the corresponding “Result” sections.

7.1 Test 1 – Individual feature accuracy
In order to find out what features that were closely related to sleepiness, the first test was to train the prediction model, based only on one feature at a time. By doing this, each feature's accuracy could be determined. Each feature's accuracy was weighted over all four channels. The result can be viewed in the table “Result 1”. All features were divided into four groups, based on their individual accuracy. The groups of features were separated in the following way:

- Group 1 (red) – Features with individual accuracy less then 60 %.
- Group 2 (yellow) - Features with individual accuracy greater then 60 %, but less then 62 %.
- Group 3 (blue) - Features with individual accuracy greater then 62 %, but less then 63%.
- Group 4 (green) - Features with individual accuracy greater then 63 %.

7.2 Test 2 – Optimal selection algorithm
The second test was performed in order to determine the most optimal feature selection algorithm for this setup. Each algorithm was tested in MATLAB, and each algorithm gave output in the following manner:

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Output Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation Based Filter (CBF)</td>
<td>List of features, ordered based on their scoring</td>
</tr>
<tr>
<td>Fast Correlation Based Filter (FCBF)</td>
<td>List, containing features deemed useful.</td>
</tr>
<tr>
<td>Chi Square Feature Evaluation (ChiSquare)</td>
<td>List of features, ordered based on their scoring</td>
</tr>
<tr>
<td>Fisher score (Fisher)</td>
<td>List of features, ordered based on their scoring</td>
</tr>
<tr>
<td>Gini-Index selection (Gini)</td>
<td>List of features, ordered based on their scoring</td>
</tr>
<tr>
<td>Information gain (InfoGain)</td>
<td>List of features, ordered based on their scoring</td>
</tr>
<tr>
<td>Kruskal-Wallis test of variance (KruskalWalli)</td>
<td>List of features, ordered based on their scoring</td>
</tr>
<tr>
<td>Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation (SBMLR)</td>
<td>List of features, ordered based on their scoring</td>
</tr>
<tr>
<td>Student distribution test (Ttest)</td>
<td>List, containing features deemed useful.</td>
</tr>
</tbody>
</table>
The feature lists were ordered in a way so that the most important feature evaluated had earlier index in the list. The output of each algorithm, were then tested in the classifier, in order to see the corresponding accuracy. The eight most important features from each algorithm were used, if not, fewer features were selected in the algorithm. The result for this test can be viewed in “Result 2”.

7.3 Test 3 – Accuracy of all selection methods
The third test consisted of testing the feature groups accuracy (from test 1), and how the accuracy of the prediction model changed, when the group with the current lowest accuracy, successively was discarded from the prediction model. So at first, the selection consisted of all features, then of all the features from group 2-4, then 3-4, and finally 4 alone. In addition, the accuracy of the best selection algorithm from test 2, was included in the comparison. The result of this test can be viewed in “Result 3”, were the average accuracy of all selection methods, on all channels, are evaluated and compared to each other.

7.4 Test 4 - Relative error
In the fourth and final test, the relative error of each feature selection method was estimated by comparing all feature selection method's average accuracy from test 3, with the average accuracy of the case when all the features were present in the construction of the prediction model. The result of this test can be viewed in “Result 4”.
<table>
<thead>
<tr>
<th>Feature Number</th>
<th>Feature ID</th>
<th>CH1(accuracy)</th>
<th>CH2(accuracy)</th>
<th>CH3(accuracy)</th>
<th>CH4(accuracy)</th>
<th>Average accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>AV</td>
<td>64,10</td>
<td>58,00</td>
<td>61,70</td>
<td>61,70</td>
<td>61,38</td>
</tr>
<tr>
<td>2</td>
<td>MAV</td>
<td>64,10</td>
<td>58,00</td>
<td>61,70</td>
<td>61,70</td>
<td>61,38</td>
</tr>
<tr>
<td>3</td>
<td>MAD</td>
<td>67,80</td>
<td>60,60</td>
<td>61,00</td>
<td>61,80</td>
<td>62,80</td>
</tr>
<tr>
<td>4</td>
<td>MNF</td>
<td>54,00</td>
<td>61,90</td>
<td>64,10</td>
<td>66,90</td>
<td>61,73</td>
</tr>
<tr>
<td>5</td>
<td>GEOMN</td>
<td>62,40</td>
<td>57,80</td>
<td>60,40</td>
<td>60,10</td>
<td>60,18</td>
</tr>
<tr>
<td>6</td>
<td>HARMN</td>
<td>56,10</td>
<td>54,70</td>
<td>55,40</td>
<td>54,50</td>
<td>55,18</td>
</tr>
<tr>
<td>7</td>
<td>RMS</td>
<td>55,80</td>
<td>55,50</td>
<td>54,40</td>
<td>55,40</td>
<td>55,28</td>
</tr>
<tr>
<td>8</td>
<td>STD</td>
<td>69,00</td>
<td>60,20</td>
<td>60,60</td>
<td>61,40</td>
<td>62,80</td>
</tr>
<tr>
<td>9</td>
<td>VAR</td>
<td>69,00</td>
<td>60,20</td>
<td>60,60</td>
<td>61,40</td>
<td>62,80</td>
</tr>
<tr>
<td>10</td>
<td>KUR</td>
<td>56,20</td>
<td>56,10</td>
<td>54,40</td>
<td>55,30</td>
<td>55,50</td>
</tr>
<tr>
<td>11</td>
<td>INQUR</td>
<td>65,80</td>
<td>58,80</td>
<td>61,60</td>
<td>61,50</td>
<td>61,93</td>
</tr>
<tr>
<td>12</td>
<td>ZCR</td>
<td>62,50</td>
<td>66,10</td>
<td>65,10</td>
<td>68,60</td>
<td>65,58</td>
</tr>
<tr>
<td>13</td>
<td>HURST</td>
<td>57,00</td>
<td>61,20</td>
<td>62,10</td>
<td>64,10</td>
<td>61,10</td>
</tr>
<tr>
<td>14</td>
<td>HIGUCHI</td>
<td>72,30</td>
<td>63,90</td>
<td>66,50</td>
<td>70,00</td>
<td>68,18</td>
</tr>
<tr>
<td>15</td>
<td>MFV</td>
<td>55,70</td>
<td>55,60</td>
<td>55,80</td>
<td>55,90</td>
<td>55,75</td>
</tr>
<tr>
<td>16</td>
<td>PC</td>
<td>67,70</td>
<td>63,90</td>
<td>64,40</td>
<td>66,50</td>
<td>65,63</td>
</tr>
<tr>
<td>17</td>
<td>MAVALPHA</td>
<td>61,90</td>
<td>56,20</td>
<td>59,40</td>
<td>61,30</td>
<td>59,70</td>
</tr>
<tr>
<td>18</td>
<td>MAVDELTA</td>
<td>62,30</td>
<td>61,90</td>
<td>63,20</td>
<td>67,20</td>
<td>63,65</td>
</tr>
<tr>
<td>19</td>
<td>MAVTHETA</td>
<td>65,10</td>
<td>60,70</td>
<td>63,00</td>
<td>64,70</td>
<td>63,38</td>
</tr>
<tr>
<td>20</td>
<td>MIN</td>
<td>63,10</td>
<td>61,60</td>
<td>63,50</td>
<td>63,30</td>
<td>62,88</td>
</tr>
<tr>
<td>21</td>
<td>MAX</td>
<td>67,30</td>
<td>59,50</td>
<td>61,60</td>
<td>62,70</td>
<td>62,78</td>
</tr>
<tr>
<td>22</td>
<td>TRAPZ</td>
<td>55,70</td>
<td>56,20</td>
<td>54,90</td>
<td>60,20</td>
<td>56,75</td>
</tr>
<tr>
<td>23</td>
<td>HJORTHCOMP</td>
<td>73,50</td>
<td>70,20</td>
<td>70,40</td>
<td>73,20</td>
<td>71,83</td>
</tr>
<tr>
<td>24</td>
<td>HJORTRM</td>
<td>59,50</td>
<td>57,80</td>
<td>60,70</td>
<td>66,50</td>
<td>61,13</td>
</tr>
<tr>
<td>25</td>
<td>BP</td>
<td>66,10</td>
<td>60,00</td>
<td>62,30</td>
<td>62,90</td>
<td>62,83</td>
</tr>
<tr>
<td>26</td>
<td>TRAPZDELTA</td>
<td>64,40</td>
<td>60,80</td>
<td>61,90</td>
<td>64,40</td>
<td>62,88</td>
</tr>
<tr>
<td>27</td>
<td>TRAPZALPHA</td>
<td>61,60</td>
<td>57,80</td>
<td>59,10</td>
<td>61,60</td>
<td>60,03</td>
</tr>
<tr>
<td>28</td>
<td>TRAPZTHETA</td>
<td>64,30</td>
<td>60,70</td>
<td>62,40</td>
<td>64,50</td>
<td>62,98</td>
</tr>
<tr>
<td>29</td>
<td>APSD</td>
<td>65,10</td>
<td>58,10</td>
<td>61,20</td>
<td>62,60</td>
<td>61,75</td>
</tr>
<tr>
<td>30</td>
<td>DPSD</td>
<td>62,10</td>
<td>59,60</td>
<td>62,40</td>
<td>64,30</td>
<td>62,10</td>
</tr>
<tr>
<td>31</td>
<td>TPSD</td>
<td>62,10</td>
<td>60,70</td>
<td>63,70</td>
<td>64,30</td>
<td>62,70</td>
</tr>
</tbody>
</table>

Illustration 7: Table showing the average accuracy of each individual feature.
Illustration 8: Diagram showing the average accuracy of each individual feature. (Red) less than 60 %, (Yellow) less than 62 %, (Blue) less than 63% and (Green) greater than 63 %.
7.6 Result 2 – Feature selection algorithms

Average accuracy for all channels

<table>
<thead>
<tr>
<th>Selection method</th>
<th>Accuracy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>83,35</td>
</tr>
<tr>
<td>Manual &gt; 60 %</td>
<td>83,19</td>
</tr>
<tr>
<td>Manual &gt; 62 %</td>
<td>82,89</td>
</tr>
<tr>
<td>Manual &gt; 63 %</td>
<td>81,18</td>
</tr>
<tr>
<td>CFS</td>
<td>81,575</td>
</tr>
<tr>
<td>FCBF</td>
<td>79,15</td>
</tr>
<tr>
<td>ChiSquare</td>
<td>82,50</td>
</tr>
<tr>
<td>Fisher</td>
<td>80,45</td>
</tr>
<tr>
<td>Gini</td>
<td>81,30</td>
</tr>
<tr>
<td>InfoGain</td>
<td>82,55</td>
</tr>
<tr>
<td>KruskalWallis</td>
<td>78,50</td>
</tr>
<tr>
<td>SBMLR</td>
<td>73,23</td>
</tr>
<tr>
<td>Ttest</td>
<td>80,63</td>
</tr>
</tbody>
</table>

Illustration 9: Table showing the average accuracy of selection methods.

Average accuracy of selection methods
(for all channels)

7.7 Result 3 – Accuracy for different selection methods

<table>
<thead>
<tr>
<th></th>
<th>Accuracy CH1</th>
<th>Accuracy CH2</th>
<th>Accuracy CH3</th>
<th>Accuracy CH4</th>
<th>Average accuracy for all channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Features</td>
<td>82,25</td>
<td>81,55</td>
<td>82,68</td>
<td>86,93</td>
<td>83,35</td>
</tr>
<tr>
<td>Manual &gt; 60 %</td>
<td>81,65</td>
<td>81,08</td>
<td>82,63</td>
<td>87,38</td>
<td>83,19</td>
</tr>
<tr>
<td>Manual &gt; 62 %</td>
<td>81,03</td>
<td>80,83</td>
<td>82,83</td>
<td>86,88</td>
<td>82,89</td>
</tr>
<tr>
<td>Manual &gt; 63 %</td>
<td>82,35</td>
<td>79,48</td>
<td>79,15</td>
<td>83,73</td>
<td>81,18</td>
</tr>
<tr>
<td>InfoGain</td>
<td>82,00</td>
<td>80,70</td>
<td>80,50</td>
<td>87,00</td>
<td>82,55</td>
</tr>
</tbody>
</table>

Illustration 11: Illustration 8: Table showing the average accuracy of the manual selection methods and the best selection algorithm.

Channel accuracy for different selections

Illustration 12: Diagram showing the average accuracy of the manual selection methods and the best selection algorithm.
7.8 Result 4 – Relative error

<table>
<thead>
<tr>
<th>Selection</th>
<th>Error CH1</th>
<th>Error CH2</th>
<th>Error CH3</th>
<th>Error CH4</th>
<th>Average error for all channels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manual &gt; 60 %</td>
<td>0.73</td>
<td>0.58</td>
<td>0.06</td>
<td>-0.52</td>
<td>0.2125</td>
</tr>
<tr>
<td>Manual &gt; 62 %</td>
<td>1.49</td>
<td>0.89</td>
<td>-0.18</td>
<td>0.06</td>
<td>0.565</td>
</tr>
<tr>
<td>Manual &gt; 63 %</td>
<td>-0.12</td>
<td>2.54</td>
<td>4.26</td>
<td>3.68</td>
<td>2.59</td>
</tr>
<tr>
<td>InfoGain</td>
<td>0.30</td>
<td>1.04</td>
<td>2.63</td>
<td>-0.09</td>
<td>0.97</td>
</tr>
</tbody>
</table>

Illustration 13: Table of the relative error of the reduced manual selection, and the best selection algorithm.

Relative error for different selections, compared to selecting all features

Illustration 14: Diagram visualizing the relative error for the reduced manual selections, and the best selection algorithm.
8. Discussion and limitations

As mentioned earlier, and as shown in the result of this research, feature selection is an important step in EEG analysis. The motivation for the concept, is that features might be redundant, or even irrelevant for the final prediction result. In other words, features can have properties that aren't distinct enough for pointing out unique characteristics of data, such as sleepiness. When testing the accuracy of individual features, this fact was clearly shown in this research, and some features could distinguish sleepiness from the signal better (for example Hjort parameters and Huguchi's fractal dimension), then other features (for example the Harmonic mean and the Root Mean Square features). So when performing prediction in learning systems, the features needs to be picked with respect to this fact, in order to generate a model that yields the most favorable accuracy. Ideal for this, in order to get maximum accuracy in the prediction model, is to have signal interpretation specialist, picking out the features manually. However, this is not always possible, like in the situation when an intelligent system is monitoring a drivers mental state over hours of driving. A specialist cant be sitting for hours, analyzing gigabytes of data from all the drivers, currently driving on the roads. So in order to make decisions, whether a driver is being sleepy or not, the implementation of feature selection algorithms is of great importance. However they usually don't reach the same accuracy in their predictions, as a specialist would achieve. But some of them clearly have the ability to pick relevant features, and to reduce the computational cost of the model. The feature selection method “Information gain” used in this research, clearly showed this fact by only selecting eight features, and still keeping up the accuracy of the model, with less then 1 % relative error compared to the case when all the features were in use.

The extent of this research was limited down based on a number of factors, and so, the actual goals of this research had to be realistic. The research will not, and was not meant to, have enough inputs or outputs, to motivate accurate assumptions on what the most optimal feature selection that exists would be. Some of the factors for why this wasn't possible to realize, are stated in this section. And it would be well to remember the limitations of this study, and that a research study of deeper extent, with more supportive data collection and tests, might be needed in order to increase the credibility of similar research. But is important to remember that this research was an optimization study, not a matter of finding the ultimate solution to the feature selection problem. So the accuracy of the individual features and feature selection algorithms are strictly relative to the parameters that was used in this particular study.

The limitations majorly consisted of the time frame that this research was executed within(1), the lack of human resources and knowledge(2), and the parameters and tools that were used in this study(3).

(1) Time span and extent of research - The research was performed within a time span of 10 weeks, and by one student alone.

(2) Human resources and knowledge- In an extensive research on the subject, several groups might be needed with different areas of expertise, such as knowledge in electronics, neurology, computer science and medical experience in EEG signal interpretation. Also extensive experience in the tools, such as MATLAB, could have had positive impact on the outcome of this research, and speed up
the process.

(3) Parameters – This is probably the biggest limitation for the outcome of the research. Firstly it is important to understand that the results shown in this research are relative to the parameters used for collecting the results. The signal that were used and the parameters from it, the features that were extracted and selected, and the training procedure of the classifier, are all factors that had impact on the accuracy of the prediction model. The accuracy should be considered only for the variables selected in this research. Higher or lower accuracy could be achieved by changing these parameters and variables.

The signal recorded from a real test drive would have given more ideal results, compared to the signal used in this research, that was a recording of a person suffering from narcolepsy. However good signals suited for this study is hard to come by, and was not available in the physionet databases. Also, only one signal was used, and only four channels were observed, and it could be argued that the number of signals, the number of channels, the number of frames, the frame size, etc, were limiting the credibility of this study. Different ages and genders of test subjects, were also factors that weren't concerned. It is also unclear how much artifacts, if any, that were percent in the signal, when the analyzing of it occurred, since no such information was found in the database, were the signal was found.

There were 31 features implemented in this study. However it was impossible to try all the possible combinations of features manually (in total \(2^{31}=2147483648\) possible combinations of the features). The selection was majorly based on firstly selecting the features with the highest single-handedly accuracy, and then combining them, since features with lower accuracy seemingly decreased the overall accuracy of the model. So by implementing other features with individual higher accuracy, a higher accuracy for the prediction model could likely be achieved. Also the features found and used in this study were merely a small amount of all the features that can be extracted and that are available. There might be other features that have higher individual accuracy, as well as other algorithms with better selection of features.

9. Future work

As mentioned in the discussion and limitation section, there were major limitations for this research, and resources that were absent. If the necessary resources would be available, there are a lot of extensive research that could be done on the area.

First of all what could be done, is to actually perform a study, were actual EEG data from more realistic tests are gathered in a driving simulator. People in different ages could participate, and Mälardalens högskola could have it's own EEG database to be used for EEG analysis. Tools for recording EEG data is currently available at Mälardalens högskola, and there are driving simulation studios available on the market, as well as open source solutions on the web to realize it.

This research was an optimization study, but in order to truly support claims of the most optimal feature selection in a realistic situation, future optimization of greater extent would be needed. This could be achieved by taking more, and other features and feature selection algorithms, into consideration. In this research, the complexity of the features that were implemented were strictly limited, and they were extracted from a limited number of domains. Hundreds of features might be needed from several domains in order to truly support claims of the most optimal feature selection, and more complex variants of feature selection algorithms should be considered. This would also
demand optimization of the feature extraction stage, and an extensive research study on optimal feature extraction algorithms. Also an extensive test on classifiers could be performed, to see which one that would be used. In addition, parameters such as increasing the number of channels to analyze, the amount of training the classifier received, etc, could be picked more ideally in a more extensive research study.

10. Conclusion

In this research, an optimization study was made on EEG feature selection. It shows that some features tends to be better to use in the construction of the prediction model, and that some features relate more to sleepiness then others. For example, the study showed that Huguchi’s fractal dimension, and the Hjorth complexity, yielding the highest individual feature accuracy, above 68 %, compared to other features such as the Harmonic mean, and the Root Mean Square, yielding the lowest individual accuracy, below 56 %.

It also shows that by manually reducing the features to be present in the construction of the prediction model, by discarding features having low individual accuracy, more or less the same accuracy can be achieved up to a certain limit. When only manually picking features, having an individual accuracy above 62 %, the average accuracy of the model had a relative error, compared to the case when all features were used, that was less then 0,6 %. And so, the computational cost were theoretically reduced, since only 16 out of 31 feature were used and had to be calculated (52 % of all features in use).

Finally, when comparing some of the available feature selection algorithms, it was concluded that the Information gain method, yielded the best selection of features, corresponding to a average accuracy of 82,55 % in the prediction model (less then 1 % relative error, compared to having all 31 features present in the construction of the prediction model). The algorithm accomplished this accuracy, by only selecting the top eight scored features from the algorithm (26 % of all the features in use). This fact concludes that, when considering computational cost (the number of features that were used in the construction of the prediction model), and the average output accuracy of the model, the Information gain method was the most optimal feature selection technique for this setup.
11. References


By Nuh, M., Jazidie, A and Muslim, M.A.
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2002
Pages: 483 - 486
Print ISBN: 0-7803-7690-0

By: Lim, C.-K.A., Wai Chong Chia and Siew Wen Chin
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2014
Pages: 1 - 5
INSPEC Accession Number: 14947376

[4] A new approach for a safe car assistance system
By: Mejdi Ben Dkhil, Mohamed Neji, Ali Wali, and Adel M. Alimi
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2015
Pages: 217 - 222
INSPEC Accession Number: 15242365


By: Niedermeyer Eand da Silva F.L
Published by: Lippincott Williams and Wilkins
Accessed through google E-books
Year: 2005

By: Ellen R Grass and Tatum, William O
Published by: http://www.readperiodicals.com/
Year: 2014
[8] Effects of eyelid closure, blinks, and eye movements on the electroencephalogram  
By: Iwasaki Masaki, Kellinghaus Christoph, Alexopoulos Andreas V. Burgess Richard C., Kumar Arun N., Han Yanning H.and Lüders Hans O.  
Published by: ScienceDirect - http://www.sciencedirect.com  
Year: 2005

[9] Dimensionality reduction-based phoneme recognition  
By: Zhang and Zhijin Zhao  
Published by: IEEE Xplore - http://ieeexplore.ieee.org/  
Year: 2008  
Page(s): 667 - 670  
Print ISBN: 978-1-4244-2178-7

Publisher: Neorosky. INC - http://neosky.com/  
Year: 2015  

By: Isabelle Guyon and Andre Elisseeff  
Published by: http://www.jmlr.org/  
Year: 2003

[12] Classification and feature selection algorithms for multi-class CGH data  
Jun Liu*, Sanjay Ranka and Tamer Kahveci  
Year: 2008

Miguel Rivera and Laura Salas  
Mälardals högskola IDT Publications - http://www.idt.mdh.se/  
Year: 2013

[15] Countermeasures for fatigue in transportation  
By: Anna Anund, Carina Fors, Göran Kecklund Wessel van Leeuwen and Torbjörn Åkerstedt  
Published by: VTI - http://www.vti.se/publications  
Year: 2015  
Page: 13

[16] Feature Extraction: Foundations and Applications  
By: Isabelle Guyon, Steve Gunn, Masoud Nickravesh and Lotfi A. Zadeh  
Published by: Springer, accessed by google E-books  
Year: 2006
[17] EMD based Temporal and Spectral Features for the Classification of EEG Signals Using Supervised Learning
Riaz, F., Hassan, A., Rehman, S., Niazi, I. And Dremstrup, K.
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2015
ISSN :1534-4320

[18] Semi-automated patient-specific scalp EEG seizure detection with unsupervised machine learning
Smart, Otis and Chen, Michael
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2015

[19] Survey on EEG Signal Processing Methods
M. Rajya Lakshmi*, Dr. T. V. Prasad and Dr. V. Chandra Prakash
Published by: http://www.ijarcsse.com/
Year: 2014

By: Sommer, D. and Golz, M.
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2010

[21] Driver Fatigue Detection Based Intelligent Vehicle Control
By: Zutao Zhang, Jia-shu Zhang
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2006
ISBN: 0-7695-2521-0

[22] EEG based Driver Fatigue Detection
By: AlZu’bi, H.S., Al-Nuaimy, W. and Al-Zubi, N.S.
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2013
ISBN: 978-1-4799-5263-

[23] Presence research and EEG
Alois Schlögl
Published by: Department of Computer Science, University College London
http://www.cs.ucl.ac.uk/research/
Year: unknown
By: Fei Wang, Shaonan Wang, Xihui Wang, Ying Peng, Yiding Yang  
Published by: IEEE Xplore - http://ieeexplore.ieee.org/  
Year: 2014  
INSPEC Accession Number: 14617095

[25] Comparison of Different Time-domain Feature Extraction Methods on Facial Gestures’ EMGs  
Published by: https://piers.org/  
Year: 2012

AvOzgur Yurur, Chi Harold Liu  
Published by: Taylor and Francis group, Page 19-22  
Year: 2015  

[27] Analysis of EEG Signal for the Detection of Brain Abnormalities  
M. Kalaivani, V. Kalaivani and V. Anusuya Devi  
Published by: http://www.ijcaonline.org/  
Year: 2014

[28] Advanced EEG Processing for the Detection of Drowsiness in Drivers  
Griet Goovaerts, Ad Denissen, Milica Milosevic, Geert van Boxtel and Sabine Van Huffel  
Published by: unknown  
Year: 2014

[29] Experimental Comparison of Geometric, Arithmetic and Harmonic Means for EEG Event Related Potential Detection  
year: 2012  
Print ISBN: 978-1-4673-4725-9

Zahra Mardi, Seyedeh Naghmeh Miri Ashtiani, and Mohammad Mikaili  
Published by: http://www.ncbi.nlm.nih.gov/  
Year: 2011

[31] A Novel EEG Feature Extraction Method Using Hjorth Parameter  
Seung-Hyeon Oh, Yu-Ri Lee, and Hyoung-Nam Kim  
Published by: http://www.ijeee.net/  
Year: 2014
[31.5] A Comparative Study on Classification of Sleep Stage Based on EEG Signals Using Feature Selection and Classification Algorithms
Bahaşen & Musa Peker & Abdullah Çavuşoğlu & Fatih V. Çelebi
Published by: http://www.researchgate.net/
year: 2014

Nikita Gurudath, H. Bryan Riley
Published by: http://www.sciencedirect.com/
Year: 2014

[33] EEG/ERP Analysis: Methods and Applications
Nidal Kamel and Aamir Saeed Malik, Page 175-195
Published by: Taylor and Francis group
Year: 2015
ISBN: 978-1-4822-2469

[34] Time domain Feature extraction and classification of EEG data for Brain Computer Interface
P. Geethanjali, Y. Krishna Mohan and Jinisha Sen
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2012
Print ISBN: 978-1-4673-0025-4

[35] Using Shannon Entropy as EEG Signal Feature for Fast Person Identification
Published by: www.elen.ucl.ac.be
Dinh Phung, Dat Tran, Wanli Ma, Phuoc Nguyen and Tien Pham
Year: 2014

[36] Feature Subset Selection: A Correlation Based Filter Approach
Mark A. Hall, Lloyd A. Smith
Published by: http://www.cs.waikato.ac.nz/
Year: unknown

[37] A feature selection based on deviation from feature centroid for text categorization
Jieming Yang and Zhiying Liu
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2011
Print ISBN: 978-1-4577-0813-8

[38] Spam Filtering Based on Improved CHI Feature Selection Method
Zhimao Lu, Hongxia Yu, Dongmei Fan and Chaoyue Yuan
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2009
Print ISBN: 978-1-4244-4199-0

Mälardalens högskola, Box 883, 721 23 Västerås
15-11-04
Feature Selection by Combining Fisher Criterion and Principal Feature Analysis
Sa Wang, Cheng-Lin Liu and Lian Zheng
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2009
Print ISBN: 978-1-4244-0973-0

Complete Gini-Index Text (GIT) feature-selection algorithm for text classification
Heum Park, Soonho Kwon and Hyuk-Chul Kwon
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2010

An Improved Information Gain Feature Selection Algorithm for SVM Text Classifier
Xu, Jiamin and Jiang, Hong
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2015
Print ISBN: 978-1-4673-9199-3

A Multi-task Feature Selection Filter for Microarray Classification
Liang Lan and Vucetic, S.
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2009
Print ISBN: 978-0-7695-3885-3

Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation
Schölkopf, B., Platt, J. and Hofmann, T.
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2007
ISBN: 9780262256919

Feature selection in power transformer fault diagnosis based on dissolved gas analysis
Samirmi, F.D., Wenhu Tang and Wu, H.
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2013
INSPEC Accession Number: 14003228

Large scale feature selection using modified random mutation hill climbing
Farmer, M.E., Bapna, S and Jain, A.K.
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2004

Genetic Algorithm Based Feature Selection Method Development for Pattern Recognition
Ho-Duck Kim, Chang-Hyun Park, Hyun-Chang Yang and Kwee-Bo Sim
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2006
Print ISBN: 89-950038-4-7
[47] An efficient content-based image retrieval with ant colony optimization feature selection schema based on wavelet and color features
Rashno, A., Sadri, S. and SadeghianNejad, H.
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2015
Print ISBN: 978-1-4799-8817-4

[48] Feature Selection Using Principal Component Analysis
Fengxi Song, Zhongwei Guo and Dayong Mei
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2010

[49] A cluster-based sequential feature selection algorithm
Kexin Zhu and Jian Yang
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2013
INSPEC Accession Number: 14320029

[50] Correlation-based Feature Selection Strategy in Neural Classification
Michalak, K. and Kwasnicka, H.
Published by: IEEE Xplore - http://ieeexplore.ieee.org/
Year: 2006
Print ISBN: 0-7695-2528-8

[51] The signal used can be found in the database: http://physionet.org/pn6/capslpdb/
Related research: Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern (CAP) in human sleep.
MG Terzano, L Parrino, A Sherieri, R Chervin, S Chokroverty, C Guilleminault, M Hirshkowitz, M Mahowald, H Moldofsky, A Rosa, R Thomas, A Walters.
Published by http://www.ncbi.nlm.nih.gov/
Year: 2001