Otto von Guericke University of Magdeburg

Faculty of Computer Science

Department of Knowledge Processing and Language Engineering

Master's Thesis

Analysis of Electroencephalographic DWT Features for Classification and Regression of Visual Field Charts

Author:

Christoph Doell

August 26, 2013

Supervisors:

Prof. Dr. Rudolf Kruse
Otto von Guericke University
Faculty of Computer Science
Universitätsplatz 2
39106 Magdeburg, Germany

Prof. Dr.-Ing. Andreas Nürnberger
Otto von Guericke University
Faculty of Computer Science
Universitätsplatz 2
39106 Magdeburg, Germany

Advisor:

Christian Moewes
Otto von Guericke University
Faculty of Computer Science
Universitätsplatz 2
39106 Magdeburg, Germany

Doell, Christoph:
doell@iws.cs.uni-magdeburg.de
Analysis of Electroencephalographic DWT Features for Classification and Regression of Visual Field Charts
Master's Thesis, Otto von Guericke University
Magdeburg, 2013.

This work is dedicated to my dear friend Christian Welz.

I hope you have found the peace you were looking for.

Abstract

This thesis is about with the analysis of correlations between brain activities of visually impaired subjects, expressed electroencephalographies ($EEGs$) and clinical variables relating to the visual perception, especially with the visual field of patients. While recent research on this topic interpret EEGs as brain graphs based on the synchronization likelihood, the primary goal of this work is to create a comparison to this approach using classical time series analysis on the EEGs. Secondary, we examine emerging questions by giving a detailed view on the data analysis process. We apply model-based machine learning techniques and cross validation to measure the prediction quality of our generated models. We utilize discrete wavelet transformation (DWT) and study different combinations of mother wavelets for feature generation. Within the presented approach, various regression algorithms including ordinary least squares, regression trees and support vector regression are compared. Further, two filtering concepts are investigated; the first depends on the noise rate of the EEG, the second on clinical variables, limiting the reliability of the underlying tests of the visual field. We show that clinical variables are well predictable by using our features and a support vector regression model.

Zusammenfassung

Die Arbeit beschäftigt sich mit der Analyse von Zusammenhängen zwischen Hirnaktivität von visuell geschadigten Probanden, dargestellt durch Elektroencephalogramme ($EEGs$) und klinischen Variablen bezüglich der Sehfähigkeit, insbesondere mit dem Sichtfeld von Patienten. Aktuelle Forschungen in diesem Themenbereich interpretieren die EEGs als Gehirn-Graphen basierend auf dem Synchronisations-Likelihood-Maß. Das Hauptziel dieser Arbeit besteht darin, dazu einen Vergleich zu schaffen, der klassische Zeitreihenanalyse auf den Electroencephalogrammen verwendet. Sekundär beschäftigt sich die Arbeit mit dem Datenanalyseprozess, an Hand dessen aufkommende Fragen gestellt und behandelt werden. Wir verwenden modellbasiertes maschinelles Lernen und Kreuzvalidierung, um die Vorhersagegüte des Modells zu messen. Beim Erstellen der Features wird die diskrete Wavelet Transformation (DWT) verwendet, wobei verschiedene Kombinationen von Mutterwavelets untersucht werden. Im verfolgten Ansatz werden verschiedene Algorithmen zur Regression verglichen, darunter die Methode der kleinsten Quadrate, Regressionsbäume und Support Vector Regression. Darüber hinaus werden zwei Möglichkeiten von Filterung eruiert. Die erste filtert nach der Stärke des Rauschens im gemessenen EEG, die zweite an Hand von klinischen Variablen, die die Glaubwürdigkeit der zugrundeliegenden Sehtests einschränken. Es wird gezeigt, dass mit unseren Features und einem Support-Vector-Regressionsmodell gute Vorhersagen über klinische Variablen möglich sind.

Acknowledgements

Contents

List of Figures

List of Tables

List of Acronyms

CART   Classification and Regression Tree

CMF   Human Cortical Magnification Factor

CVars   Clinical Variables

dB   Decibel

DFT   Discrete Fourier Transformation

DWT   Discrete Wavelet Transformation

EC   Eyes Closed

ED   Entrance Diagnosis

EEG   Electroencephalogram

EEG   Electroencephalography

EO   Eyes Open

ET   Epileptiform Transient

HRP   High Resolution Perimetry

ICA   Independent Component Analysis

IQR   Inter Quartile Range

LogMAR   Logarithm of the Minimum Angle of Resolution

LOO-CV   Leave One Out Cross-Validation

MSE   Mean Squared Error

NI   Node Impurity

OLS   Ordinary Least Squares

PCA   Principal Component Analysis

POM   Primary Outcome Measure

RBF   Radial Basis Function

RSS   Residual Sum of Squares

SL   Synchronization Likelihood

SVM   Support Vector Machine

VAR   Vector Autoregression

VF-CV   V Fold Cross-Validation

Chapter 1

Introduction and Motivation

'If the brain were so simple we could understand it, we would be so simple
we couldn't.'

(Lyall Watson)

Understanding the human brain and especially its adaptivity is a topic of research for
decades. There is scientific evidence that after the loss of one sense, the other senses
improve. Gougoux et al. [18] found improved audio perception at blind people. Differ-
ent to seeing people, their visual cortex did not receive visual information and adapted
by enhancing the ability to locate sounds. We focus on the adaption of the brain af-
ter a damage to the visual system. Visual disorders can occur for example after an
apoplexia. The diagnosis to what extend the field of view is limited can be determined
by conducting a test using a perimeter device. In some cases, for this test the patient
needs to look at one point with full concentration for more than one hour. As result
of the test, a grid of the field of view can be created, which shows to what extend the
sight is limited. These visual field charts can be processed to real values of so called
clinical variables which describe different defects of the visual system.
Recent studies [25, 34] investigate the possibility to use an EEG to predict the outcome
of this test and therefore get a hint on measurable adoptions of the brain. Creating an
EEG for a patient needs less time and is less exhausting for him. The usual method
to predict the disorders magnitude with EEG data is to calculate and analyze average
values; Held et al. [21] generated dynamic graphs, averaged them and used character-
istics of this graph, at the cost of the temporal dynamics. The approach of Varela et
al. [43] uses the temporal information and interpret the EEG data directly as dynamic
graph. This approach was later used by Moewes et al. [34]. They applied several global
graph measures on EEG data and used machine learning methods on these measures,

to build models to predict clinical variables. We compare the results of the dynamic brain graph approach from Moewes et al. [34] to those results we get when creating prediction models by interpreting the data as time series.

As data source, there are data from different studies available. These data are aggregated to deliver as many examples as possible to the machine learning algorithm. Some questions are important in this approach: How can thousands of measured points of the EEG signal be transformed to a set of few and stable features for learning? Does filtering improve the results? Which types of clinical variables like global vision or peripheral vision can be predicted the best?

This thesis is structured as follows: Chapter 2 contains information on medical, computer scientific and further related work, which will be referred to later on. Chapter 3 provides information about the available data, the preprocessing and reports the data understanding. Chapter 4 describes the data analysis process, the questions which come up along this process and the experiments designed to answer them. The experiments' results are shown and discussed in Chapter 5. The work concludes with summing up the main results and presenting possible suggestions for future research in Chapter 6. Readers, who prefer not to jump between chapters, can read this thesis in the given order. Curious readers or experts can continue with Chapter 4 and turn back, whenever background information is necessary.

Chapter 2

Related Work

The following chapter provides basic information, required to understand the experiments. The first section of this chapter focuses on the the biological background, the second section introduces related methods from computer science and the third and last section presents recent scientific works in the field.

2.1 Background in Biology and Neurosciences

In the following section, we first show how damages of the optic nerve can influence the visual field. Then we present procedures measuring different aspects of visual perception. After a short introduction to electroencephalography we reveal some problems of EEG analysis, which leads to the section about computer science.

2.1.1 The Eye

The eye is one of the most important organs for humans. According to [24] the rate of information, sent from eye to brain is approximately $875.000\,^{\text{bits}}/_{\text{sec}}$. There the information is processed. The way the data take and how different damages in optical nerves can limit the visual field of a subject, is displayed in Figure 2.1. An early damage of *nervus opticus*, as in *a*, can cause one eye to completely lose function, while a damage in *chiasma opticum*, as in *b*, results in a loss of the visual field of one side in one eye, and the loss of the visual field of the other side in the other eye. When the *tractus opticus* cannot forward information, as in *c*, this generates a loss in the visual field of the same side in both eyes. A nerve damage behind the *corpus geniculatum laterale*, as
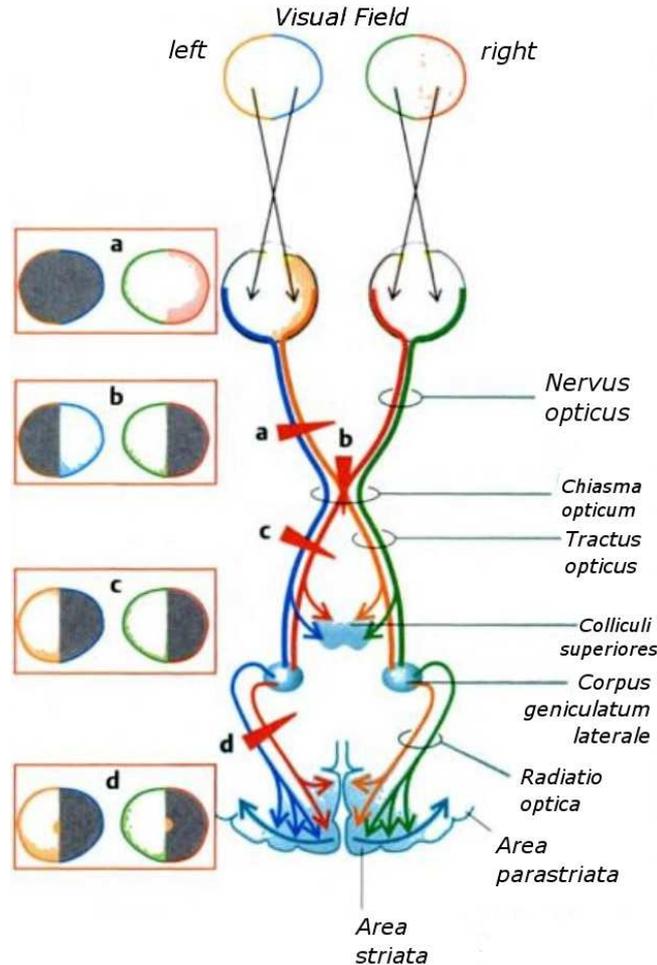
Figure 2.1: Schematic view of possible injuries of the optical nerves and their resulting visual field defect. Gray areas imply (partial) blindness, the content of white areas can be seen. (Translated) Original Source: [39, p. 365]

in d, results in a loss of the visual field of one side in both eyes in which the center of the visual field would not be defect.

Generally, the human visual perception in sensitivity and resolution is highly dependent on the retinal locations. Results in physiological research have shown that functional and structural properties of the visual system are similar and closely related across the whole retina [45]. It has been evaluated that the sensitivity and resolution decreases from the visual field center to the peripheral visual field. The *Human Cortical Magnification Factor (CMF)* has been measured, giving a possibility to weight all areas the same thus compensating the overweight of the center. The factor only depends on the angular distance to the center of the visual field. In order to benchmark visual field deficits taking this natural behavior into account, it is possible to use the *Human Cortical Magnification Factor* as normalization.

2.1.2 Optometry

The word optometry has its origins in the Greek words "opsis" → view and "metron" → measure; there are several possibilities for measuring a person's visual perception. The so called *logMAR chart*, as depicted in Figure 2.2, is frequently used as test chart for visual acuity. *LogMAR* stands for Logarithm of the Minimum Angle of Resolution. It has been developed to provide less fluctuating results in repeated measurements than earlier charts exhibited, which is shown in [36]. During a test, the subject reads characters from the chart, starting with large ones, ending with small ones until they become indecipherable. Usually this testing method is applied separately for near vision at a distance of 40 cm and far vision at distance of 3 m to 6 m from the chart. A result value of 3 shows that the subject was unable to correctly identify any of the characters, a value of 0 shows a normal acuity of a healthy person, values smaller than 0 stand for outstanding visual acuity. The logMAR test measures central vision like visual acuity thus the subject intuitively focuses the character to be read. Measuring the visual field is called perimetry; for the different abilities of the eye – distinguishing contrasts and recognizing moving objects – and for the measurement of the visual field's size, different kinds of perimetry have been developed, e.g. static perimetry, high resolution perimetry (HRP) and kinetic perimetry (see [11] for an overview).

The primary goal of kinetic perimetry is to determine the size of the visual field. In the past, physicians used finger perimetry to roughly determine a patient's peripheral visual field. The patient fixates centrally, while a stimulus like a patch of light or earlier a finger is slowly moved across the visual field from a non-seeing area until it is detected. The detection point is recorded. Repeated measurements for several angles build a 360° chart for the visual field. One disadvantage of kinetic perimetry is that the measurement is influenced by the patient's reaction time, as a slower reaction time leads to a smaller value. It is further known that a moving stimulus can be detected more easily than a static stimulus. Nevertheless, kinetic perimetry is one of the fastest methods for measuring the size of the peripheral visual field.

Figure 2.2: LogMAR chart for measuring the visual acuity, source: [48]

In static perimetry, the goal is to determine the light sensitivity across the entire visual field. Therefore light stimuli with different intensities are presented across the visual field and the lowest intensity measured for each spot builds the visual field map. The

Figure 2.3: Humphrey Field Analyzer, source: [27, p. 395]

Figure 2.4: High Resolution Perimetry (HRP), from [37]
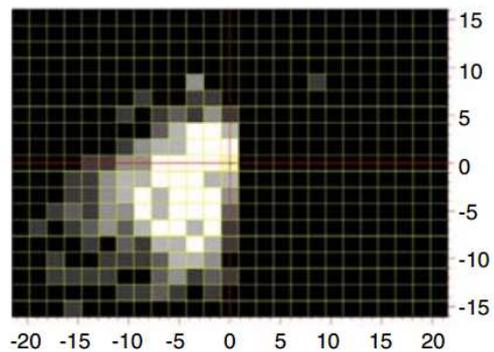
Figure 2.5: Result of a HRP test of a right eye, from [37]

measurement is performed for each eye separately. While the first eye is tested the other one is completely blocked by an eye patch.

Nowadays, the visual fields are assessed with computer based high-resolution perimetry (HRP) [22, 37]. When the patient recognizes a light stimulus, he pushes a button. If the button is pushed, although there was no point shown, this counts as a false positive hit. Assuming a patient wants to improve his test results by pushing the button randomly, this would be detected in a high false positive ratio. Generally, when the visual field is measured, in order to gain a reliable result, the patients have to focus on a central point. Moving the eye, searching for points, might influence the results. For measuring the fixation accuracy there are several possibilities; some automatic methods use a camera, observing the patient's pupil, others change the color of the central point, whereupon the patient reacts, pushing the button attesting correct fixation.

A Humphrey-Field-Analyzer is depicted in Figure 2.3 as an example of a computer assisted perimeter. The execution of a high resolution perimetry is painted in Figure 2.4.

The result of a high resolution perimetry can be seen in Figure 2.5. A differently colored 25×19 grid is shown, where the colors indicate to what extend the patient was able to see the light at the specific point. White means that each test on that point was

successful, gray means that some some tests were successful, while some were not, and black means that the patient could not recognize any of the points in the sector. The fixation spot in the center is shown by the vertical and horizontal axes.

So far, different kinds of measurements for the visual field and visual acuity were presented. Yet it is unclear which ones are best predictable by using EEGs. The following sections focus on this subject.

2.1.3 Electroencephalography

Electroencephalography records electrical activity along the head. Several electrodes are placed on the scalp, measuring the electric potential differences and their changes over time. Each electrode records one channel. Their distribution is depicted in Figure 2.6. The purpose is to have a non-invasive method for measuring brain activities. The resulting electroencephalogram, which is also abbreviated *EEG*, is used for example for the medical examination of epilepsy or in the diagnosis of brain death. It is also frequently used in neuroscience as a non-invasive, cheap and fast tool to retrieve data of a brain at work. Achermann [1] shows differences in the Fourier transformed EEG signal during different states of sleep. Already in 1975, Kelsey [23] presented significant alterations in the EEG of blind adults.

The signal can also be disturbed by muscle activities like eye movement, heartbeat or nearby technical devices. Parts of the signal, which do not show the electrical activity of the brain, are called artifacts. Although the data are recorded in shielded rooms, artifacts can usually only be reduced, and therefore have to be filtered, corrected or rejected [14], increasing the quality of the EEG data for later automatic processing.

The signal of one channel can be split in a way revealing underlying waves of different frequencies. Details on how this works are presented later in this chapter. Specific frequency ranges change, depending on the subject's state of activity. As mentioned earlier, different states of sleep can be observed as well as whether the subject is relaxed or engaged. Examples of raw EEG and several filtered frequency bands are portrayed in Figure 2.8. Changes in these frequency bands are frequently viewed separately, e.g. in [16, 25, 34, 33].

2.1.4 EEG in Data Analysis

EEG data are usually analyzed by experienced neurologists by means of visual inspection, where channels as shown in Figure 2.7 are used. When learning how to interpret
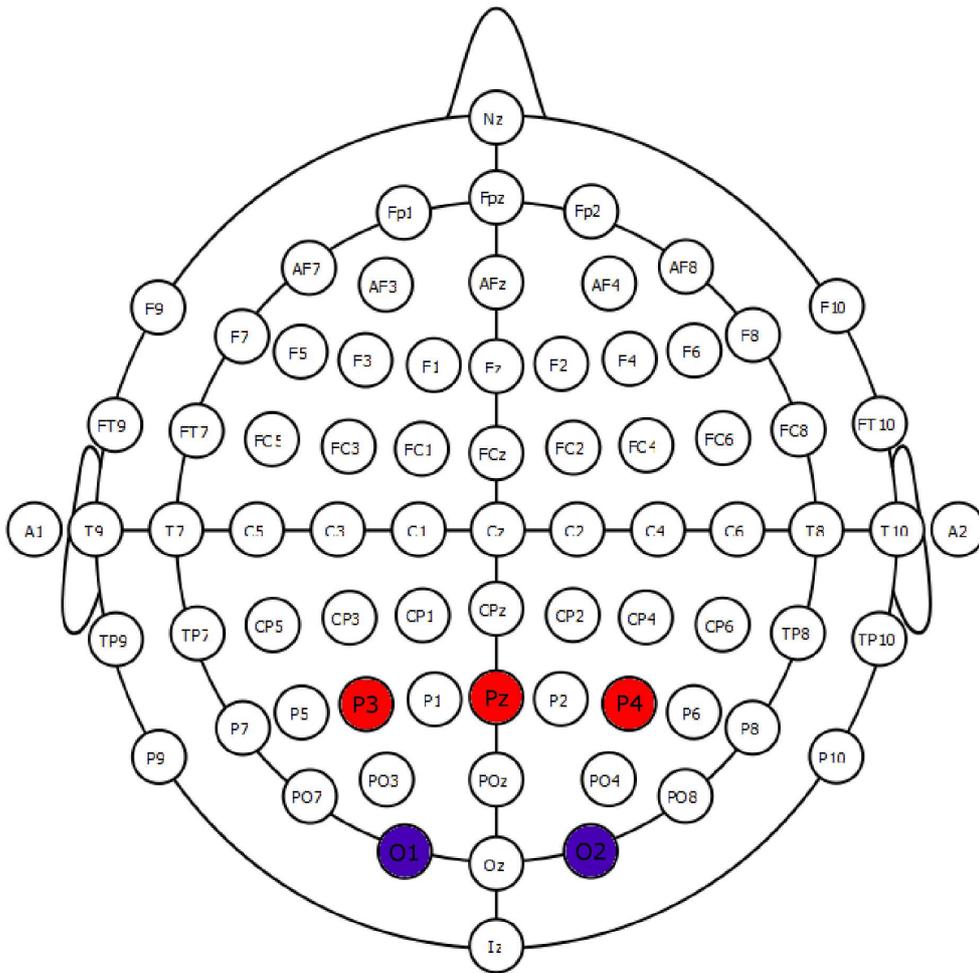
Figure 2.6: Schematic diagram of the 10-20-EEG distribution of electrodes on the skalp [49]. Electrodes of special interest in the data analysis part are colored. The occipital electrodes O1 and O2 are colored blue, parietal electrodes P3, Pz, P4 are highlighted red.

the EEG correctly there are several aspects to consider. All electrodes are displayed and the signal change in time is shown just in real time. The doctor's task is to find patterns which imply a disease, or if not finding any signals for diseases, treat the patient as healthy. From the data analysis view, these data contain several synchronized time series, one for each electrode, which in our case have a length of approximately 100 seconds, taken with a frequency of 250 Hz.

$$20\,[electrodes/patient] \times 100\,[seconds] \times 250\,[datapoints/second] =$$

$$500.000\,[datapoints/patient]$$

So we have half a million data points per patient and the task is to find features used later to predict the degree of a disease. From now on we use the term *data point* in

Figure 2.7: Example EEG measurement with channel titles and time marks, source: [51]

order to describe one patient's entire EEG measuring in a high dimensional feature space. *Dataset* stands for a set of data points. Beyer et al. show in [7] that usual distance measures are not meaningful when the dimensionality is too high, which is the case with time series, when each measurement is interpreted as one dimension. Note that this approach, directly applied to EEG data, does not cope with the autocorrelation effect; usually the signal does not differ much from one measured value to the next, which results in (highly) correlated dimensions.

So the first goal is to process the signals in a way that generates some few features for the machine learning algorithms to produce reliable results. There are several approaches available to increase the information density in the EEG signal before features are used for learning: *principal component analysis (PCA)*, see also Section 2.2.7, as extension the kernel PCA, the *regularized common spatial patterns algorithm* [29] or *independent component analysis (ICA)* [14]. All these algorithms cope with the problem of a low sample size at a high dimensionality. Nevertheless, when this gap gets too large, like in this case with ca. 100 samples with 0.5 million dimensions, expert knowledge is necessary to reduce this divergence. We abstracted the following procedures for dimension reduction from earlier approaches:

- Omitting
- Averaging
- Transforming

Such dimensions, where it is probable or even known that they are unimportant in, can be omitted. Sabel et al. [37] used only two channels and the frequency range from 7.5 to 13 Hz (alpha-rhythm). Focusing on few brain areas was done earlier e.g. by Barry et al. [4] and Kriegseis et al. [26]. Averaging can be used if the data are probably im-

Frequency Band Name	Frequency Bandwidth	State Associated with Bandwidth	Example of Filtered Bandwidth
Raw EEG	0–45 Hz	Awake	
Delta	0.5–3.5 Hz	Deep Sleep	
Theta	4–7.5 Hz	Drowsy	
Alpha	8–12 Hz	Relaxed	
Beta	13–35 Hz	Engaged	
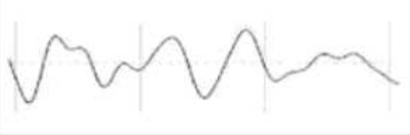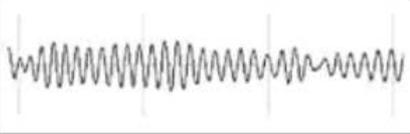
Figure 2.8: Frequency bands separation, from [9, p. 87]

portant, but the number of dimensions is still too high and it is not known in detail, which ones, from a given set of dimensions, are needed. At this point it is important to carefully choose which and how many dimensions are averaged when the original data are omitted. The calculation of the standard deviation can be seen as extension of averaging. If there is also interest in the number of outliers, they can also be counted before omitting the raw dimensions. These extensions try to cope with the knowledge that averaging a whole set of n dimensions down to only one mean value, surely loses lots of information. In this case the reduction is performed from n to 3 values, at which all 3 contain a habit of the signals information that cannot be reconstructed even if the others are available. Averaging electrodes was performed e.g. in [16], extensions of averaging (using mean, standard deviation, maximum and minimum) have been applied e.g. in [47, 17].

Transforming data points into another space is often complex and intensive in computation; it is used when the transformed space suits better than the original one for interpretation or further processing. The *Fast Fourier Transformation (FFT)* has been performed for example in [16, 37, 26]. The solution we finally use for experimentation is explained in Section 4.1.1.

2.2 Used Methods of Computer Science

This section contains several methods used in computer science, which are relevant for the experiments. First classification and regression are explained, followed by model-based machine learning techniques. Then a general problem for the prediction task is stated, which can occur, when combining data from several studies, which is not generated exactly the same way. Discrete Fourier Transformation and Discrete Wavelet Transformation are introduced. Finally, different methods for cross validation and the principal component analysis and their purposes are described.

2.2.1 Supervised Learning: Classification and Regression

In supervised learning, a set of training examples is given, each containing a pair of input and desired output values. Algorithms learn these examples, generate models and can then predict output values on given input. The algorithms are validated by their predictions on a test dataset. Results from heterogeneous test datasets show to what extent generalizing was possible or if the algorithm just learned the examples by heart. In the classification task, the goal is first to learn a model and then give class predictions for the data points in the validation dataset. The easiest form of classification distinguishes only two classes, e.g. class $\ominus$ and class $\oplus$, where class $\oplus$ means that the given points belongs to the points with a given attribute and class $\ominus$ means that the point does not have this attribute. So there are two types of correct decisions. First: the predicted class is $\ominus$ and the actual class is $\ominus$, so the point is predicted not to have the attribute and the decision is correct, this is called true negative decision. Second: the predicted class is $\oplus$ and the actual class is $\oplus$, so the point is predicted to have the attribute and in fact has it, which is called true positive. Similarly, predicting the positive class $\oplus$ while the real one is class $\ominus$ is called false positive, predicting the negative class $\ominus$ when it is the positive class $\oplus$ is called false negative. The distribution of the learning dataset is important at this point, because an uneven distribution between the classes leads to an uneven prediction. Given that there are three times more examples for class $\ominus$ than for class $\oplus$, using only this distribution for the prediction, a general prediction of $\ominus$, independent from the given data point, will lead to a correct prediction rate of 75%. This is one of the reasons, why there are other measurements for the performance of the model, which take the distribution of the datasets into account. One method applies weights to each of the four cases described above, so that for example false predictions have negative weights while correct ones have positive weights. The goal is then to maximize the sum of the scored weighted values.

The classification task predicts a class variable. The regression task differs here and predicts a real value instead of a class. Example: Assume that explanatory variables about a set of people are given, then predicting the gender of a person on the one hand is a classification task, because there are only two possibilities available for humans. On the other, predicting the age of a person is a regression task, because the value to be predicted is real. In contrast to classification there are no more four cases, which show if the prediction was positive or negative and correct or not. For regression there is just the difference in the predicted value and the real one. If the difference is 'small' the prediction was good, otherwise not. This leads to the question, what 'small' means in this case. It is dependent on the application and of course very dependent on the distribution of the dependent variables. When predicting two different variables, where one has a much higher variance than the other one, it is obvious that predicting the mean value in the case of the small variance might be a good estimation, while in the other case this would result in much worse outcomes. So when comparing the outcomes of regressions, it is necessary to have a deeper look at the dependent variable's distribution or first normalize both to the same distribution. For an automatic, computer supported evaluation process, the first method is not applicable. That is why in this work the dependent variables will be normalized to standard normal distribution. This can be done by using the standard *Z-score* instead of the raw variable X. When X is the measured random variable with mean value μ and standard deviation σ, then be $z = \frac{x-\mu}{\sigma}$.

More detailed information about the classification task and the regression task can be found in [6, Ch.8].

2.2.2   Model-Based Machine Learning Techniques

After discussing supervised learning, we focus on machine learning techniques which will be used in the experimentation for classification or regression. We define for the following chapter: Be S a set of k examples in an n dimensional space, with $\forall i \in 1 \ldots k, s_i = (x_i, y_i)$, where for each $x_i = (v_1 \ldots v_n)$ denotes the input vector, and $y \in \{\oplus, \ominus\}$ builds the classifications output. A_j builds a dimension for each $j \in 1 \ldots n$, $dom(A_j)$ is the set of all possible values of A_j and $S_{A=v} \subseteq S$ for which attribute A has value v.

Linear Regression

A regression model takes all the given (high dimensional) points given for learning and fits a function so that the summed error between the actual point and the function value is minimized. In a linear regression the only functions taken into account are straight lines. Lots of different measurements of the error can be found in literature. One of the best interpretable ones, which is used later on in this work, is *ordinary least squares (OLS)* . The value to be minimized is the sum of the euclidean distance between predicted and true value. For details, the exact definition, and problems that might possibly occur when working with linear regression, see [5]. Linear regression is one of the simplest methods which can be applied to a dataset and the interpretation of the result is simple. It is usually used for the first experiments or for comparison with more difficult models.

Decision Tree

A decision tree learner is a typical example for a divide and conquer algorithm, details can be found in [31]. It splits the dataset regarding given attributes. The decision to which class the elements belong, shall be easier after. Performing a task with a collection of examples S to be classified, the question can be asked to what extent they are already separated. The measure, *entropy*, is defined to describe this behavior. When the ratio of elements of the negative class is $p_\ominus$, and the ratio of the elements of the positive class $p_\oplus$, then

$$Entropy(S) := -p_\oplus log_2(p_\oplus) - p_\ominus log_2(p_\ominus)$$

Note that in the calculation $0 \, log(0)$ is defined to be 0. *Entropy* has its maximal value at 1, when $p_\ominus = p_\oplus$. The minimal value is 0 for $p_\ominus = 1$, and $p_\oplus = 0$ or vice versa. So, if there is only one class represented, the entropy is 0. If the number of elements in S is the same for both classes (and at least one), then the value is 1. The definition of *entropy*, as described, holds only when the target classification is boolean. In the case that the target attribute can take on c different values, the following definition is applied:

$$Entropy(S) := \sum_{i=1}^{c} -p_i log_2(p_i)$$

Getting back to the idea of splitting the dataset, the splits improvement can be measured by the *information gain*, which describes the expected reduction in entropy caused by the split. When using A_j as split attribute, the *information gain* is defined as

$$Gain(S, A_j) := Entropy(S) - \sum_{v \in dom(A_j)} \frac{|S_v|}{|S|} Entropy(S_v)$$

The first term contains the entropy before the split, the second term the summed entropies after the split so that the information gain when splitting at A_j is the difference of these two terms. The information gain is the measure which was used when decision trees were first presented in 1986 by Quinlan [35]. The attribute used for the split is chosen by trying all remaining attributes and taking the one ensuring the highest *information gain*. This procedure is iterated for all subtrees and for the datasets in that tree, until there are no further attributes remaining or the entropy is not increasing for all remaining splits. As implied earlier, every attribute is only used once as split attribute on every path from the root to the leaves.

Using an earlier learned decision tree for classification works as follows. Starting at the root, the split attribute of the current node is checked on the given data point. The search is continued at the child, which holds only such elements with the attribute identical to the one of the wanted data point. When a leaf of the tree is reached, the value, which the learning datasets in this leaf implied, is predicted. For the regression task, with real value attributes, the split attribute itself and the split point are again chosen in a way that optimizes the *node impurity (NI)*. Node impurity defines a measurement like *entropy* where no fixed number c of classes is needed. For non-categorical attributes *entropy* can no more be used as measurement, due to the fact that there are no classes available. An extension solving this problem is the *residual sum of squares (RSS)*. It is implemented in the *Classification and Regression Tree (CART)*, which is described in [28].

$$RSS(S) = \sum_{i=1}^{k} (v_i - mean(S, i))^2$$

v_i describes the ith attribute of the learned point v and $mean(S, i)$ represents the mean value of attribute i for all points in S. This means that the tree will grow up until a leaf exists for each predicted value; for m different points this means m leaves. So this would result in an over-fitted model, which is able to memorize the learned datasets but unable to generalize. This is why the tree has to be pruned at the end, which enforces a generalization. Merging leaves is executed in a way that the costs for regression

errors are minimal under the condition that leaves have to contain a certain amount of learning datasets and therefore are able to perform regression predictions. When the learning process is complete and the prediction task with a new dataset is performed, the search performs similar to the search in classification trees. When a leaf is reached the mean value stored in this leaf will be returned as predicted value. Interpreting the decision tree in a geometrical way splits the feature space iteratively along axis parallel hyperplanes. This feature allows to construct a human interpretable set of rules, which contains splits of fixed values of given attributes.

Support Vector Machine

As described above, Splitting the search space is performed parallel to the given dimensions along the split attributes. This results in the possibility to generate association rules from a decision tree. Generalizing this idea, a freely chosen separating hyperplane cannot reduce the prediction results, because axis parallel splits could still be used. When freely choosing the plane, it is possible to separate the space in a way so that at least the same number of points are correctly predicted. This idea is implemented in the *Support Vector Machine (SVM)*. Interested readers may have a look at [41, 32]. If there is such a hyperplane $x : \langle w, x \rangle + b = 0$, with $w \in \mathbb{R}^n$ being the norm of the vector and $b \in \mathbb{R}$ is the bias of the hyperplane p separating the points correctly, then they are called linear separable. The decision function for a separation by a hyperplane is given by:

$$f(x) = sgn\left(\langle w, x \rangle + b\right).$$

Note that $\oplus$ and $\ominus$ are replaced by 1 and -1 in the decision formula. Then, one might be interested in a plane which performs this task in a way that the margin between the two sets of points is as large as possible. It can be shown that, if the points are linear separable, there exists a unique hyperplane, which maximizes the margin. Some points might be far away from p, not influencing it at all. The other points, which influence the separation, are called *support vectors* and give the name to the *SVM*. This plane can be found by solving the quadratic optimization problem with linear inequality constraints:

$$\underset{w,b}{\text{minimize}} \quad \tau(w) = \frac{1}{2}\|w\|$$
$$\text{subject to} \quad y_i\left(\langle w, x_i \rangle + b\right) \geq 1, \forall i = 1, \ldots, k$$

In order to create a linear separability, where none is naturally given, support vector machines use a function $\phi : X \mapsto H$ which raises the data points from the original space X into a higher dimensional space H, look for a separating hyperplane p there, and

afterwards project the points back to X. In order to find the separating hyperplane only the similarities in H are needed, not the actual (expensive) transformation. Accessing the similarities directly $\Phi : X \times X \to \mathbb{R}$ is called *kernel trick*. The resulting separation borders are not necessarily hyperplanes in X and are less intuitive than the decision boundaries of a decision tree, for example. There is a huge amount of kernel functions for any kind of applications, see e.g. [38]. Most prominent in application are *linear* and *radial basis function (rbf)* kernels.

Naïve Bayes Classifier

Bayes classifiers use the probability theory based on the work of *Thomas Bayes* for their classification. Well explained introductions to Bayes classifiers can be found in [8, ch. 6.1] and in [6, ch. 8.2]. Given a domain of classes, $\text{dom}(C) = \{c_1, \ldots, c_m\}$ and a set of attributes $\{A_1 \ldots, A_n\}$, an object instantiation can by written using the attribute values $(a_1, \ldots, a_n)$ of the attributes $A_1, \ldots, A_n$. The naïve Bayes classifier computes the conditional probability $P\left(C = c_i \mid A_1 = a_1 \ldots, A_n = a_n\right)$ for all classes c_i. The class with highest probability will be predicted. In practice it is not possible to store all probabilities for all possible given values, because their number increases exponentially with the number of attributes n. Assuming that the attributes are independently distributed, applying Bayes' rule simplifies the calculation to

$$P\left(C = C_i \mid A_1 = a_1, \ldots, A_n = a_n\right) = \frac{f\left(A_1 = a_1, \ldots, A_n = a_n \mid C = c_i\right) \cdot P\left(C = c_i\right)}{f\left(A_1 = a_1, \ldots, A_n = a_n\right)}.$$

The denominator is only dependent on the object; when using this formula for one object and all classes, in order to find the most probable class, the object is constant and so the denominator can be ignored. Using the chain rule of probability results in the formula:

$$P\left(C = C_i \mid A_1 = a_1, \ldots, A_n = a_n\right) = \frac{P\left(C = c_i\right)}{p_0} \cdot \prod_{j=1}^{n} f\left(A_j = a_j \mid C = c_i\right).$$

At this point p_0 is a normalization constant, $P\left(C = c_i\right)$ describes the empirical probability which is stored for every class. For a symbolic class attribute A_j the conditional probabilities $P\left(A_j = a_j \mid C = Ci\right)$ can be stored in a table. For numeric real value attributes, there are two possibilities: numeric attributes can be discretized and then handled like symbolic ones, or the probability density can be stored.

In practical applications, a normal distribution is often assumed and only the expected values $\mu_j\left(c_i\right)$ and the variances $\sigma_j^2\left(c_i\right)$ need to be stored. When for example the predicted, most probable class only has a probability of 30%, it is still predicted, because of

the lack of alternative probable possibilities. This behavior often leads to surprisingly good results in practice for classification, although the probability estimations may be poor (cf. [46]). Naïve bayes classifiers can be interpreted as simple bayesian networks, which principially do not need the independence assumption [8].

2.2.3 Predict the Device or the Disease

A general problem can occur, when datasets from several studies shall be jointly analyzed. For example, imagine there are sick people being studied in one study and the control group data comes from another one. Furthermore, the studies might have taken place in different cities, by different doctors, with different EEG devices, EEG enhancers and different electrodes.

When there are so many differences in the studies, to which extend will the result be influenced by all these differences? Maybe the prediction is good at finding the EEG device or the electrodes and not so good at finding the disease, the study will then show a strong prediction ratio although one cannot be sure what exactly has been predicted. So one could try to use these data to predict the devices instead of the disease. How would variables for this experiment look like? The input data would be the same EEG data and the predicted variable – lets say device X is used for the sick people and device Y for healthy people – would have a 100% coincidence. So the prediction will be – independent from the applied learning method – the same for both experiments. This means when using data which have this coincidence it is principially impossible to distinguish between the two predictions. This has to be avoided to get a clear conclusion.

2.2.4 Discrete Fourier Transformation

A time series is a sequence of data points, successively measured, usually at uniform time intervals. Every equally-spaced numerical time series may be interpreted as a discrete function from the time point to the given value. It can be shown that every such given time series of length n can be written as sum of (at most n) sine functions of different frequency and amplitude. This process is called *Discrete Fourier Transformation (DFT)*. It was first researched by the French mathematician *Jean Baptiste Joseph Fourier* in the early 19th century, when he was investigating heat transfer and vibrations. Honoring his work, this transformation bears his name. For an overview of this topic, see [15].

In the transformed representation, every original data point can be restored by summing up all the function values for the given point. Trying to minimize a given stream of data, one might for example omit the second half of the stream. This would reduce the memory usage by the half, of course, but would also lose the second half completely. In the Fourier-representation, leaving a random point out would change the entire time series, but it would only change it slightly, because all the other summands would still be remaining. When leaving out the values with the smallest amplitude, then the general shape of the signal could be mostly conserved while the amount of data would be reduced.

This idea has been employed in several fields of signal processing, such as image processing or audio compression. With regards to the periodicity of the sine function, this means that the Fourier-representation does not just reconstruct the given time series, but the given time series plus its periodic continuation.

A conventional plot of the original time series uses the x-axis as the axis for time and the y-axis for the values of the time series. In contrast, the transformed time series uses the frequency of the sine function in the x-axis and the amplitude as value of the frequency. So one can see, in a common way, the frequencies, which are included in the signal by plotting the transformed time series and having e.g. a look at the frequencies, which show high amplitudes. One feature of the *DFT* is that every value of a single sine function affects the restoration of every point. This might not be desired, for example, in a compression task. In the current scenario of analyzing EEG time series, temporal changes of the signal are also of interest. They would be untraceable in a frequency-representation.

2.2.5   Discrete Wavelet Transformation

Trying to dispose of the behavior of the discrete Fourier transformation that every change of the transformed time series affects the whole reconstructed time series, it is necessary to find another basis function for the transformation than the sine function. Due to the scientific work of *Ingrid Daubechies*, there are several continuous functions known, allowing a transformation to be done without losing any data. Additionally less transformed points are needed to reconstruct an original data point. Some examples of such transformation functions, so called mother wavelets or basis wavelets, are shown in Figure 2.9. Honoring their first creator, some of the wavelets developed by her are called Daubechies Wavelets, for example DB2, DB4, DB5, DB20. An introduction into the topic of wavelets, about wavelet composition and decomposition can be found in [44]. Further details are summed up in [42].
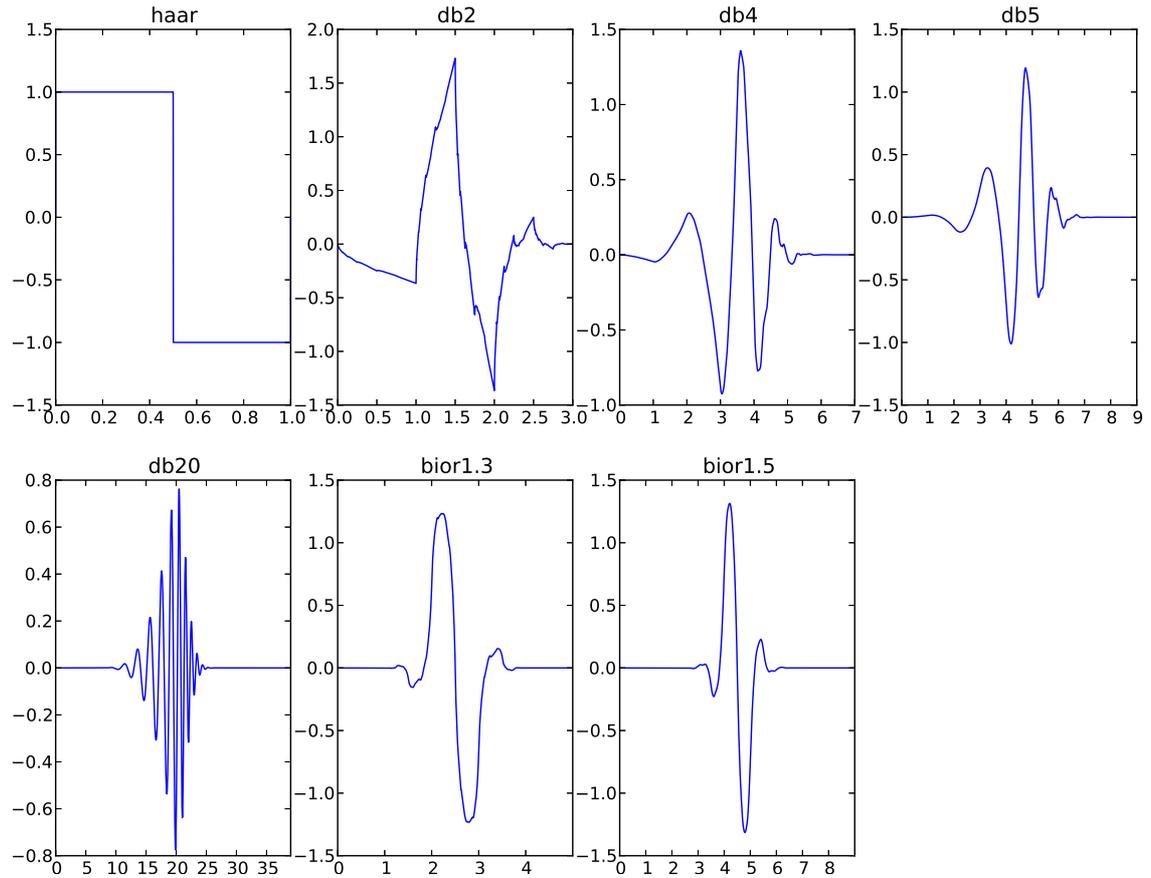
Figure 2.9: Transformation functions for the transformation of the wavelets: Haar, DB2, DB4, DB5, DB20, biorthogonal wavelet 1.3 and biorthogonal wavelet 1.5

When performing a discrete wavelet transformation, the wavelet coefficient for the current signal are calculated and afterwards the signal is split at the center, so in the next iteration the process will be repeated for both splits. This results in a tree of splits and wavelet coefficients. In the simplest case, when the Haar wavelet acts as basis function, and the original time series has a length of 2^n for an $n \in \mathbb{N}_0$ the tree always splits binaryly and each level $k \in \mathbb{N}_0$ of the tree contains 2^k elements.

When using the Haar wavelet, the values at the nodes of the tree represent the mean value of the underlying level. The root contains the mean value of the whole time series. In the second level there are two values, one containing the mean value of the first split, one for the second split and so on. In the last level of the tree only the differences of each two neighbored points are stored. This means, that omitting the lowest level of the tree for compression purposes would retain a time series with a pretty similar shape as the original, it would just be more coarse-grained. When looking at the wavelet transformed data, it is not possible to easily distinguish the frequencies contained in the signal, as it was possible after the Fourier transformation. When traversing the wavelet

tree in level-order, there are temporal characteristics in this plot, most coarse-grained at the left most side, getting more and more detailed to the right. Whenever one level ends and the next lower level of the tree starts we can see another frequency band. This way the wavelet transformation delivers a natural split into disjoint frequency bands of the time series. Note that a change of a single data point in the original time series only affects one transformed value per level of the tree, which sums up to $\log_2(n)$ values in the entire wavelet transformed signal, while it might affect all values of the Fourier transformed signal.

2.2.6   Cross Validation

In a machine learning scenario, it is mostly difficult to validate the learned models and their results. When using the same datasets for learning and for validation, it is simple to get a perfect result. This is why datasets for validation and learning have to be disjunctively chosen in way to assure independence. Furthermore, in order to get the most valid result, the machine learning algorithm should get a real chance of learning the entire data space and should therefore get uniformly distributed samples for learning. When there are many datasets available, the best method for a validation is to split the dataset into two disjunct, uniformly distributed parts, where one is used as learning dataset and the other is used for validation. But also when there are only few data available for learning and validation of a model, there is a need for reliable validation results. Cross validation addresses this need and uses the same dataset several times, performing different splits and for each split independently applying learning and validation. A summary of different cross validation methods with advantages and disadvantages can be found in [3].

The *Leave One Out Cross-Validation (LOO-CV)* always splits n data points in a way that there is exactly one data point used for the validation, while all the other $n-1$ data points are utilized for learning. There are n possible splits, for all of them the validation measure is computed and the results will be averaged.

The *V Fold Cross-Validation (VF-CV)* splits the data once to several equally sized, disjoint, preferably identically distributed parts, which are called folds. Let the number of the folds be v. Then $v-1$ folds are used for learning, while the last one remains for validation. The results will again, be averaged. Note that for $v = n$, this becomes a *LOO-CV*. When $v < n$, the computational cost of a *VF-CV* is lower.

Dependent on n, v and the validation's purpose, it can be reasonable to choose the folds randomly and repeat the entire procedure several times in order to gain a faster calculation than *LOO-CV* and still retrieve a reliable validation result. It is known that

the folded version leads to more pessimistic, less variable, better generalizing results than *LOO-CV* [19]. Nevertheless, *LOO-CV* models the real learning scenario better, taking more of the (few) datasets for learning. Showing the results' distribution instead of averaging the results handles the high variance by displaying it.

2.2.7 Principal Component Analysis

Principal component analysis (PCA) is a statistical method, projecting a set of data points from the original high m dimensional space to a lower or equal q dimensional linear subspace, while preserving as much variance as possible. When interpreted geometrically, the data are a cloud of points. It is first centered at the origin and then rotated so that the first axis contains the most variance and so on. The axes are called *principal components*. When the first q principal components contain 100% of the variance, this means that the points are located on a q dimensional plane in the m dimensional space. The variance distribution among the dimensions indicates how many dimensions are nessecary to cover most of the variance. At this point the analyst has to decide how much variance suffices. Note that it can be seen as disadvantage that the resulting subspace can not generally be interpreted as easily as the original dimensions. Further details, also on the mathematical background, can be found in [6].

2.3 Related Former Studies

Zhou et al. show in [47] that for patients with epileptic diseases the EEG-Analysis performs well when trying to predict artifacts in EEG signals correctly as epileptiform or non-epileptiform. Epilepsy is clearly known as a brain defect. Looking at the EEG signal specialists can see epileptiform transients (ETs). The ET's visual nature is already known, but they vary in morphologies and they can be similar to normal background activities. In their approach, a 4-level-wavelet decomposition was applied for several basis wavelets. The resulting wavelet coefficients were used as one feature vector to predict whether or not an EEG example contains an ET or not. They combined several daubechies and biorthogonal wavelets as mother wavelets, leading to improved results. Good results appeared for the single usage of daubechies wavelets DB2, DB4, DB5, DB20 and biorthogonal 1.3 and 1.5. The best combination was found to be DB2 + DB4.

Moewes and Kruse [33] interpret EEG data as dynamic graphs and thereby generate network features describing their dynamics. These features correlate with the extent of

the vision loss. In the graph each electrode is a node. There are two variants for edges connecting nodes: First, all nodes are connected to each other, for nodes A and B the edge weight $w(A, B)$ gets the similarity $s(A, B)$, for a similarity measure s. Second, an arbitrarily chosen threshold t is given, all edges are unweighted, edges exist between A and B if and only if $s(A, B) \geq t$. Note that as similarity measure *Synchronization Likelihood (SL)* is used here. For the described graphs several graph measures like *density* are computed receiving again a multivariate time series. Vector autoregressive (VAR) models are learned from them, building a predictor for clinical variables. The results of the predictions are listed in Table 2.1, showing the medians of the *Mean Squared Errors (MSE)* from LOO-CV. The median squared prediction errors are mostly above 1.0, meaning that rough predictions are possible, but no exact result can be expected. Surprisingly the thresholded version for the left eye shows better results in all cases, while for the right eye it is mostly the other way around. Most stable results for the eyes appear at the best overall results for the thresholded graph predicting detection accuracy in HRP, with values smaller than 0.9. This implies that dynamic graph models are better able to represent the complex, mean measurements from high resolution perimetry than the less complex ones like visual acuity.

Q1: *Which approach leads to better results? EEG-Interpretation as classical time series or as brain graph?*

Clinical Variable	weighted		thresholded	
	right eye	left eye	right eye	left eye
HRP: Detection accuracy	1.160	1.466	0.878	0.896
Static perimetry: foveal threshold	1.077	1.232	1.307	1.009
Static perimetry: mean threshold	0.906	1.182	1.050	0.836
Kinetic Perimetry: Mean Eccentricity	1.112	1.307	1.318	1.064
Visual Acuity: near vision (logMAR scaled)	1.449	1.332	1.610	1.087
Visual Acuity: far vision (logMAR scaled)	1.151	1.314	1.610	1.021

Table 2.1: Mean squared error regression results for both eyes from [33], 50 Hz lowpass EEG filtered, using weights or thresholds in brain networks. As validation LOO-CV was performed. OLS was used for regression.

Köppen [25] analyzed the phase lag index as similarity measure and came to the conclusion that there were no systematical correlations between the field of view and the EEG. He suggests separating the two halves of the field of view, and at the same time only taking the EEG data from those electrodes of the brain areas where the signal is processed. As the available data in this work do not support a split in the two halves, we can only follow the second suggestion and focus on brain areas which process visual information.

Chapter 3

Data Description

All data used for this work were collected in studies, that have been approved by an ethical standards committee on human experimentation. All the studies were registered at `http://clinicaltrials.gov/`, the according detail information can be found when following the links of Table 3.1. Written informed consent was obtained from all patients participating in the studies (consent for research). All patients were treated according to the Declaration of Helsinki.

Study	Link to clinical trials details
NO 2	`http://www.clinicaltrials.gov/ct2/show/study/NCT01270126`
NO 3	`http://www.clinicaltrials.gov/ct2/show/study/NCT01282827`
MCS	`http://www.clinicaltrials.gov/ct2/show/study/NCT01280877`

Table 3.1: Clinical Studies and links to their documentation

One goal of these studies is to capture temporal changes of EEG and clinical variables, documenting the results of an experimental therapy. Clinical criterion for participation in NO3 was low vision, for NO2 and MCS patients were suffering from optic nerve diseases, optic nerve injuries or optic neuropathies. In this work, we focus on the *entrance diagnosis (ED)*, which means we always use the first measurement available of each patient, before any kind of therapy was applied. When only looking at these data, the different studies can be combined, generating a bigger pool of datasets for analysis. Table 3.2 shows numbers of patients, for which EEGs or clinical variables were available.

Study	NO2	NO3	MCS	$\sum$
Participants	21	14	98	133
EEG available eyes open	11	14	0	25
EEG available eyes closed	19	14	72	105
Left eyes in CVars	21	14	82	117
Right eyes in CVars	20	14	82	116
Both eyes in CVars	20	14	82	116
Left eye and EEG	18	14	68	100
Right eye and EEG	19	14	68	101
Both eyes and EEG	18	14	68	100

Table 3.2: The number of the participants, for which EEG data, clinical variables or both were available, grouped by study. Clinical variables (CVars) are counted if at least one of them is accessible, so it cannot be assumed that tests on all variables can be performed on all of the datasets listed.

3.1 Clinical Variables

When building a predictor for visual field damage, we first specify how to measure the damage. Generally in this work, clinical variables mean data, specifying a patient's ability to see. They were measured with the methods explained earlier. Section 2.1.2 describes several of these methods, which are used in this study. After the measuring has been performed, the data have to be preprocessed. In this particular case this means deriving numerical clinical variables from all gathered data, which can later be used for experiments, e.g. for correlating EEG features with them.

The collected data vary dependent on the focus of the studies. The clinical variables could be grouped by the type of measurement performed when obtaining them: static perimetry, kinetic perimetry, high resolution perimetry and visual acuity. Another grouping focuses on the nature of the variable itself: does it contain data about the visual field defect or about the reaction time? Because of the high number of different clinical variables, in the data understanding phase, preprocessing needs to be performed. This includes grouping and filtering them in order to remove redundancies and correlations before the primary experiments of this work are conducted. Details on this issue follow later in this chapter.

3.1.1 Grouping clinical variables

Remembering the goal, to find correlations between the EEG data and the visual field defect, we want few or even only one variable, describing all the visual field defects. So as first step we created six groups of variables, dependent on different content:

1. **Experimentation validity** collects values which challenge the credibility of the patient's visual test results. In perimetric tests in which a patient fixates one spot, it is a general assumption of the test that the correct fixation is given. Another assumption is that the patients always act honestly, not trying to cheat or to manipulate the results in any way.

   - fixation accuracy in HRP [% ED]
   - fixation accuracy in static perimetry [% ED]

2. **Central vision** contains the variables describing the center of the visual field or those describing the general visual perception. Humans focus on the point they want to see in detail and because the eye's focus point is at the center of the visual field, measurements for visual acuity can be counted for central vision. Light sensitivity in the fovea, which is located centrally in the eye, is included as well as values for visual acuity, which are for physiological reasons normally the best in the central area. Fixation accuracy could also quantify the quality of the central vision, although there are other plausible reasons for a low fixation ratio, like lack of concentration. However these values are already in the group about experimentation validity.

   - foveal threshold in static perimetry [dB ED]
   - far vision (LogMAR scale) [ED]
   - near vision (LogMAR scale) [ED]

3. **High resolution perimetry (HRP)**, normalized with respect to the *Cortical Magnification Factor* from Chapter 2.1.1 builds the next group. All sectors in the HRP result were multiplied with their corresponding CMF factor, the products are summed up, grouped by sector color. Although values for totally defected sectors were also available, we did not use them because of the redundancy to the variables used.

   - relative size of partly defected sectors using CMF
   - relative size of intact sectors using CMF

4. **Reaction times** from HRP are contained in the following set. It is known that humans' reaction times are generally influenced by several factors such as the age or the state of rest. Here they are measured between the light stimulus and the following reaction. They were calculated with slightly different methods, depending on the underlying perimetric test and the study. Therefore, the first three variables were only available for NO2 and NO3, while the others existed only for MCS. A high similarity is assumed.

- reaction time for intact sectors
- reaction time for partly defected sectors
- reaction time for partly defected sectors (normalized with respect to eccentricity)
- reaction time in HRP [ms ED]
- reaction time in HRP defective visual field sectors [ms ED]

5. **Global vision** comprehends measures from HRP. It contains aspects globally describing the visual perception. For NO2 and NO3 the measure was called "POM whole field", with POM standing for *primary outcome measure*, for MCS it was named "detection accuracy in whole HRP visual field [% ED]". A priori it was not clear if these two variables described exactly the same information, if they were calculated the same way or to what extent they were alike. Both values describe the number of gray sectors divided by the number of all sectors in HRP.

   - detection accuracy in whole HRP visual field [% ED]
   - POM whole field

6. **Peripheral vision** describes values about the periphery of the visual field. Fhe average angle distance from the center to the outer most point is taken from kinetic perimetry. From static perimetry, in a fixed angle distance of 30° the threshold is measured and averaged.

   - mean eccentricity in kinetic perimetry [degrees ED]
   - mean threshold in static perimetry (whole 30 degrees visual field)[dB ED]

3.1.2   Preprocessing Clinical Variables for Machine Learning

One important step in data analysis is data understanding, having a look at the data, grouping and normalizing them. It can be assumed that some of these variables are correlated with each other, because they — at least partially — measure the same issues. For example: relative sizes of defected, partially defected and intact sectors sum up to one. Having the grouped data available, we first have to determine those clinical variables, which shall be used for experimentation.

As described in Section 2.1.2, a low fixation accuracy during the test might lead to falsified results. The same fact applies for false positive reactions. It can be assumed that false positive reactions might occur, when nerve cells, which just healed, now need much longer to react. The general distribution of the reaction times does not support this hypothesis, which can be seen in Figure 3.1 for the left, and in Figure 3.2 for
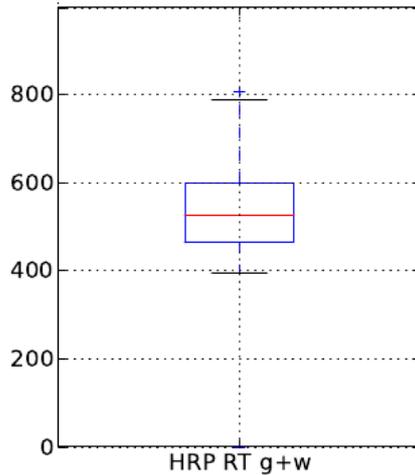
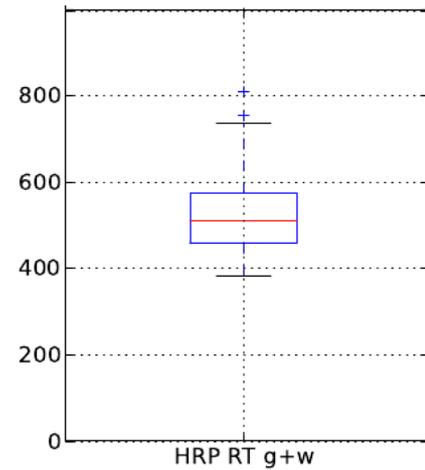Figure 3.1: Boxplot of reaction times in HRP for left eyes

Figure 3.2: Boxplot of reaction times in HRP for right eyes

the right eye. As illustrated, the general reaction times are much lower than the time threshold $1000\,\mathrm{ms}$ between two tests of the testing device. From the subject's surveys it is known that some subjects tried to improve their testing results by pushing the button, although they had not seen a point. Further reasons for false positive reactions are e.g. that a subject sees a point, where actually none was displayed, which can occur after a longer time of concentration on the test. It is also possible that subjects pushed the button by mistake.

For all of these cases, it can be assumed that a high false positive ratio shows that the reliability of the experiment can be mistrusted. Therefore for further analysis, we use the values from the experimentation validity group for filtering purposes only. In consultation with experts we ignore all subjects with a false positive rate above 5% — knowing that for each of the 475 points in the grid at least three tests haven been performed, this means that at least 72 false positive clicks have been made, before the filter declines the patient. In addition we filter all subjects with a fixation accuracy below 90%, because again, the reliablity of the experiment can be mistrusted, when the fixation accuracy is too low.

A scatter matrix plot showing these two groups' distributions and answering the question for correlations within them is illustrated for the left eye in Figure 3.3, for the right eye in Figure 3.4. The two figures do not significantly differ, so all statements following apply for both eyes. Most of the datasets build one group around 98% fixation accuracy and 2% false positive reactions. It is no correlation visible within the two variables. In the main diagonal, plots show the one dimensional variable distributions, the others the scatterplots. Due to symmetry in the points of the two remaining plots, in the lower left one there are filtering lines added implying that points left of the vertical and above

the horizontal one are culled. In the variable distributions filters are shown by vertical lines, filtering out points left of the line in the upper plot and those right of the line in the lower one. The positions of the filtering line in die distributions imply that that 5% fixation ratio is a moderate filter, while 90% fixation accuracy is rather strict. Similarities in the other groups of clinical variables will be treated in the next section.
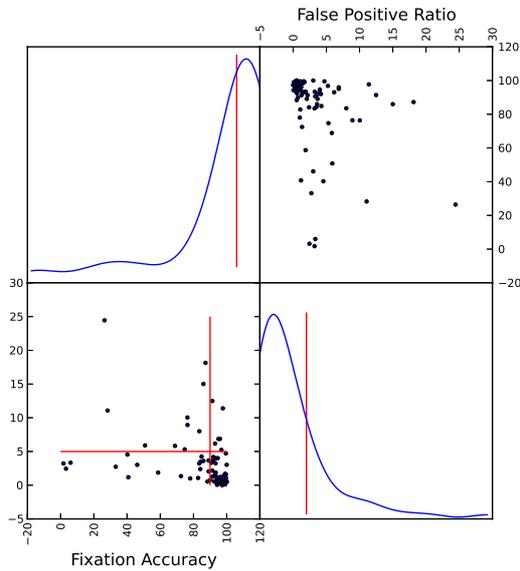
Figure 3.3: Scattermatrix showing the distributions of false positive reactions and fixation accuracy and filtering planes for left eyes
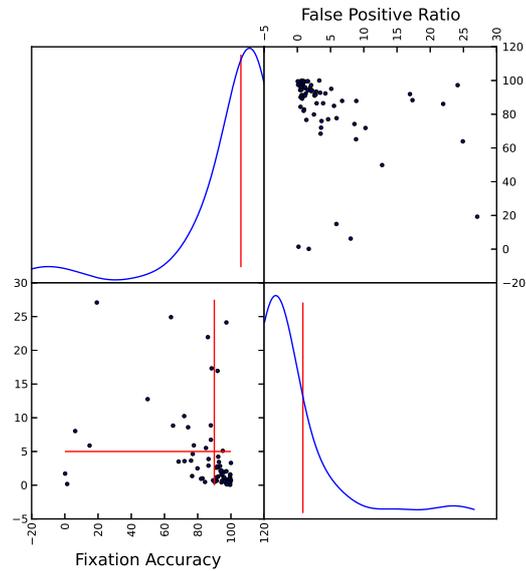
Figure 3.4: Scattermatrix showing the distributions of false positive reactions and fixation accuracy and filtering planes for right eyes

3.1.3   Principal Component Analysis on Clinical Variables

Within the groups it may be assumed that there exist natural correlations. There also might be common reasons for correlations in these groups. Trying to find out which elements of the group contain the most information, a principal component analysis on each group has been performed. All measures for reaction times were available on only 7 subjects for the right eye and only 9 for the left eye. Performing a principal component analysis on so few data points cannot produce conclusive results, therefore a split of the reaction times is performed dependent on the studies, so that two groups were created at which each contains at least 15 subjects for both eyes. The full results of the PCAs can be seen in Table 3.3. For reaction times(1) and CMF there are less than 20 data points available, the analysis results can be considered unreliable and therefore will not be examined further in this thesis. For global vision the number of datasets for which *detection accuracy* is available, is 67 for the left and 70 for the right eye, while *POM whole field* is available for 28 datasets left and 29 datasets for the right eye. Having

Group	left eye			right eye		
	datasets for PCA	principal component	explained variance	datasets for PCA	principal component	explained variance
central vision	73	1st	82%	78	1st	69%
		2nd	10%		2nd	19%
		3rd	8%		3rd	12%
CMF	17	1st	69%	19	1st	63%
		2nd	31%		2nd	37%
reaction times(1)	17	1st	78%	19	1st	74%
		2nd	20%		2nd	23%
		3rd	< 1%		3rd	3%
reaction times(2)	42	1st	97%	45	1st	97%
		2nd	3%		2nd	3%
global vision	16	1st	> 99%	18	1st	> 99%
		2nd	< 1%		2nd	< 1%
peripheral vision	71	1st	84%	73	1st	79%
		2nd	16%		2st	21%

Table 3.3: Groups number of datasets and results of the principal component analysis

99% explained variance in the first principal component it is clear now that both values contain the same variable. For further processing *POM whole field*, having fewer data points, will be omitted. In general, the results of the two eyes are pretty equal, reflecting the similarity of the eyes visual deficits. Further analysis in central vision should contain at least two principal components so that the variance explained by them for both eyes lies above 80%, which is the percentage we decided to keep for each group. The datasets in reaction times(2), show 97% of the variance in the first principal component, thus one component suffices for further analysis. The results for peripheral vision show a variance split around 80%. Keeping in mind that using the principal component, instead of the two original values costs the straight interpretability of all later work, at this point we chose to stay working with the original data, knowing that the results for the two dimensions could be correlated and therefore related results can be expected.

Summing up the results and considering which data shall be used for the main experimentation, the table shows that only central vision and peripheral vision have datasets for more than 70 patients available. One of the five dimensions could be reduced for the cost of the interpretability of the central vision experiments. The interpretation of the main experiments is too important, so in the main experiments the five dimensions from central and peripheral vision, plus the detection accuracy in HRP visual field, as representative for global vision, will be used directly.

3.2   EEG Data

The EEG signals were preprocessed in a way similar to the method described in [34, 33]. This was performed in EEGLAB [13], where the following actions were executed:

- manual removal of noisy frames at the beginning/end of each recording
- removal of uncommon EEG channels across all subjects
- high-pass filtering with cutoff frequency at 1Hz to remove slow movements
- notch filtering 50Hz to cope with European power line frequency
- low-pass filtering with cutoff frequency at 95Hz,
- re-referencing by the average electrode,
- down-sampling to 250Hz to get a unified sampling rate in all datasets,
- removal of biological artifacts using independent component analysis [30].
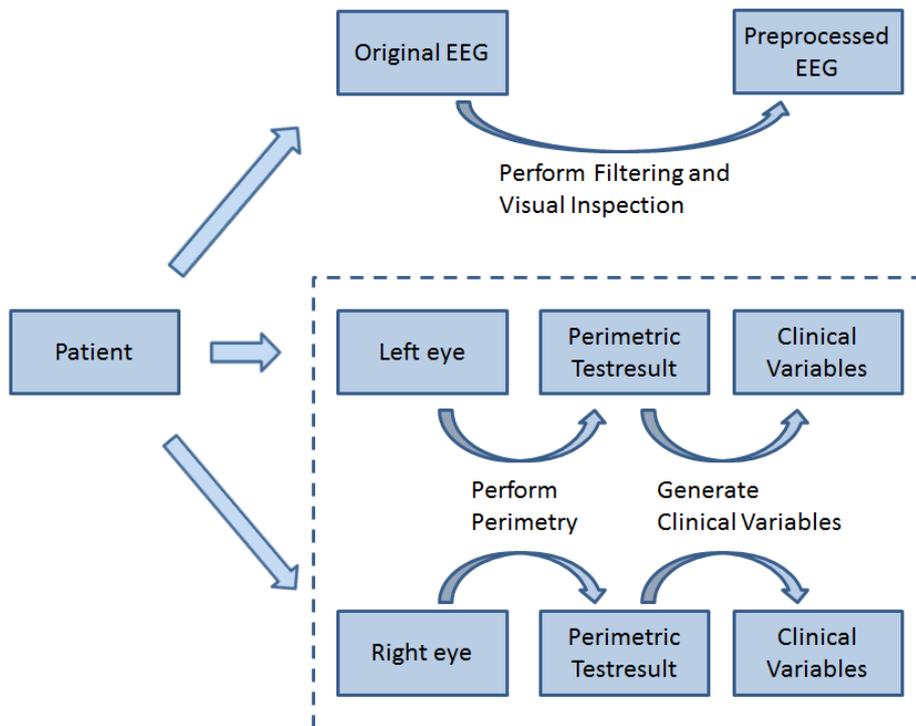
Figure 3.5: Schematic process of preprocessing performed on the data before any further actions were performed.

The schematic view on the preprocessing that was performed right after data acquisition is sketched in Figure 3.5. For each patient, EEGs with closed eyes were recorded, in NO2 and NO3 there were also recordings with open eyes. All of the EEGs were preprocessed in the explained way. Clinical variables were also taken from every patient. At least one eye was measured and recorded, the second eye was measured if it also fulfilled the entrance criteria of the study, when it was neither completely intact nor blind.

Experts defined and calculated the clinical variables from the perimetric test results. For experiments directly predicting clinical variables for both eyes, only such patients were included where measurements for both eyes existed.
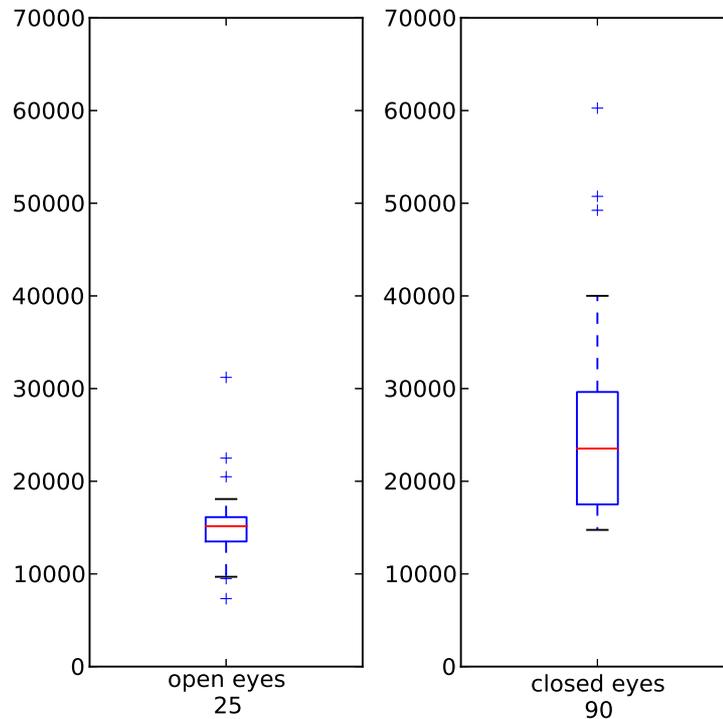
Figure 3.6: Boxplots showing the distributions of numbers of measurements of EEGs (all sampled at 250 Hz) for open and for closed eyes

Chu et al. [10] found out that functional networks in EEGs are highly variable when contemplating short periods of only one second, but they found stable networks when considering data from periods of ∼100 seconds. So for our approach, although not using brain networks, it can be assumed that a period of 100 seconds should be used to gain stable results. The patient's EEGs differ in the recorded length. For transformation purposes, it is helpful that all the EEGs have the same length $l \in \mathbb{N}$ with $l = 2^k$ for a $k \in \mathbb{N}$, see also Chapter 2.2.5. Different approaches are possible; the easiest would be to take the size s of the shortest EEG available and choose k such that $2^k \leq s < 2^{k+1}$ and $l = 2^k$. When considering the length distributions from Figure 3.6 for closed and open eyes, the sizes of the boxes show that most of the EEGs available are much longer than s. Though taking a shorter part might lead to less stable results, we set $l = 2^{k+1}$ as fixed length for the EEGs; this means for open eyes, we use $l = 2^{13} = 8192$ measurements, which equals 32.768 seconds, for closed eyes $l = 2^{14} = 16384$ measurements, meaning 65.536 seconds.

In both cases we do not reach the recommended 100 seconds. EEGs shorter than l are filled up with zeros. For the longer EEG recordings the first, a random, connected part from the middle and the last part are used. We chose to perform experiments on all of these cases separately, comparing them and thereby testing the stability of the prediction. According to [10], because of the shorter EEG length we expect the experiments containing open eyes data to reveal less stable results.

Q2: *Are the experiment's results stable in the part of the EEG taken?*

Knowing now about the data sources, the preprocessing of EEGs, the groups of clinical variables and their correlations, we explain the experiments in the following chapter.

Chapter 4

Experiments

In the following chapter we first present the primary goal of the experiments. After that we elaborate on the data analysis process and describe our feature generation. We explain the problem of predicting the visual field with the EEG, when in fact a human being usually has 2 eyes and hence 2 visual fields. This leads to the two experimentation approaches: classification of eyes open versus eyes closed and the regression task with clinical variables, which builds the basis for the comparison with brain graphs.

Given EEG data and values describing the visual field defect of one patient, the goal is to find a model that finds the differences of the visual field defects in the EEG data. The optimal case would be a model or even simpler a rule, which describes the EEG's differences between a healthy and a visually injured person. The entire topic aims at how visual information is processed on a global brain level. The experiments performed for this thesis use methods of classical time series analysis and the results are compared with the results achieved with brain graphs [34] presented in Chapter 2.3.

4.1 The Data Analysis Process and Its Goals

Although most steps in the data analysis process are simple and already known, there are often details like normalizing to [0,1] or to $N(0,1)$, producing completely different results, which is why we explain the process in detail. A further benefit of this approach is that questions arising in the given context can be named directly, as already done with question **Q1** and **Q2**.

4.1.1   Feature Generation

One main problem of the entire analysis is to find reliable features for a good and stable prediction. We described earlier that for every patient, the EEG data contain approximately 500.000 data points of real values.

Sabel et al. have shown in [37] that after alternating current stimulation, a non-invasive therapy, the EEG changes primarily in the occipital cortex and the visual field improves for patients with nerve injuries. Although the focus in this work is on changes of the visual field; we chose the EEG occipital electrodes near the visual cortex, O1, O2 and P3, Pz, P4 shown in Figure 2.6 from the medial parietal areas, which are also important for visuospatial processing, like hand-eye-coordination [40]. Omitting the other channels and focusing only on the occipital and parietal brain areas is a strategy, which has been pursued earlier, e.g. in [16, 4, 26]. In order to get one signal for each of the two brain areas and because the visual analysis of the data showed noise in some cases, especially in O1 and O2, it is convenient to average these channels. In the following O-Mean and P-Mean are arbitrary channels, containing the mean data from channels O1 and O2 and P3, Pz, P4, respectively. It is unclear if averaging the raw signals and generating the features afterwards or processing the signal vice versa, will result in better results, hence experiments for both procedures were performed. EEG preprocessing and feature generation are depicted in Figure 4.1.

> **Q3:** *Which order leads to better results? Average the raw signal and generate the features afterwards or vice versa?*

Imagine not transforming the EEG time series: When asking which attributes are most important for the prediction, the result will contain a set of points in time and split values for them, which allow a good prediction for the training datasets. This means that single time stamps in the EEG recording generate the decision to what extent the patient is ill. So the result for the learning datasets will be that the value of the nth time stamp after the start of recording must be higher than a value v_n, to predict full health. For the learning dataset, this might be correct, but the goal of the model, namely to be able to generalize from the learning data in a way, which produces good predictions especially in the validation datasets, is obviously not fulfilled, because the measurements are not synchronized. So, aiming on generating a model, which can generalize correctly, the next step in processing is a transformation. Keeping not only the frequencies but also the time domain, the wavelet transformation is performed. This has another advantage: the values of single points or short parts of the signal, like when

cringing, only changes the transformed signal in few points, as explained in Section 2.2.5. This is why we chose a wavelet transformation for the EEG preprocessing.

Frequency Range	Name
2Hz-4Hz	low1
4Hz-8Hz	low2
8Hz-16Hz	high1
16Hz-32Hz	high2

Table 4.1: Overview of the frequency ranges

We keep only these frequency ranges that are known for differences in EEGs of blind people [26]. So, having cut the tree of wavelet features at the leaves and at the root, the center with the frequencies between $2\,\mathrm{Hz}$ and $32\,\mathrm{Hz}$, as described in Table 4.1, remains. It is probable that the information of the 4 frequency ranges are the important ones for the upcoming machine learning step. Anyway, the number of dimensions is still too high, so now, not exactly knowing which parts of the signal are important, an averaging step is performed. As described in Section 2.1.4, we take the mean, the standard deviation and the number of the outliers as features. Note that we count those values as outliers, which are beyond 3 times the signal's standard deviation. These 3 values for each of the 4 frequency bands for the 2 channels result in 24 features per wavelet, which will be used from now on. Note that learning with an *SVM* requires scaled features in order to give all features the same weight, when learning; this is why the 24 features are each scaled to [0,1].
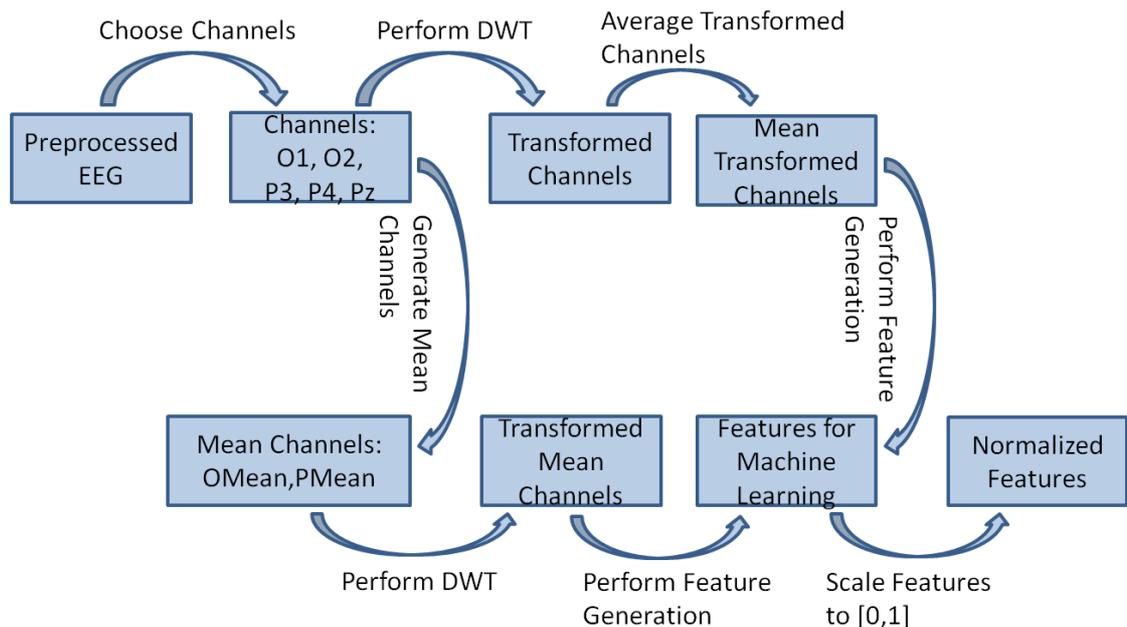

Figure 4.1: Schema, describing the two ways from the preprocessed EEG to normalized features

Thinking of the different shapes of the wavelets as in Figure 2.9, it is unclear which wavelet should be used as mother wavelet. Tests for several mother wavelets were performed for electrocardiograms in [12]. For EEGs a comparison of different mother wavelets was performed by Zhou et al. [47]. They also performed experiments using several mother wavelets for the transformation at the same time, gaining improved results. These results open the question, if different mother wavelets are generally capable of covering different types of information of the signal and therefore the results improve.

> **Q4:** *Which mother wavelet leads to best results? Can the use of several mother wavelets at the same time (keeping all features) significantly improve predictions?*

4.1.2   One EEG, but Two Eyes

The EEG takes data from all over the scalp, while the task is to find differences in the visual field defect depending on the EEG. At this point, one should not forget that humans have two eyes and therefore two visual fields. There might be differences in the brain's processing, dependent on the eye. One should not make the mistake to take each EEG-Signal twice in order to simultaneously fit one model for both eyes not separating them, because this would mean having two dependent variables for the model to fit at the same time for each set of independent variables. Nevertheless, the goal is to have a model for both eyes. Since we have variables given for both eyes, there are several possibilities to combine them: We generate separate, independent models for the left and the right eyes. This way we can perform predictions for every eye of every patient given. It is possible that these models are not well represented in the EEG but that visual perception is somehow better represented in one model for both eyes. The mean value of both eyes might be better predictable or the worse eye or the better. Not knowing which model leads to the most promising results and which is best predictable, we perform experiments for all five possibilities.

Note that the models for left and right eye generally use all datasets of patients whose clinical variables were available for left or right eye respectively. For combined models only such datasets can be used, where both values were given for the patient. This results in a smaller number of datasets in the models for combinations of clinical variables.

The preprocessing on clinical variables is depicted in Figure 4.2. Choosing and grouping of the variables was described earlier in Chapter 3.1.1. Note that a normalization is

performed after the creation of the combined clinical variables in order to make all
results comparable. For all types a standard normal distribution is created.
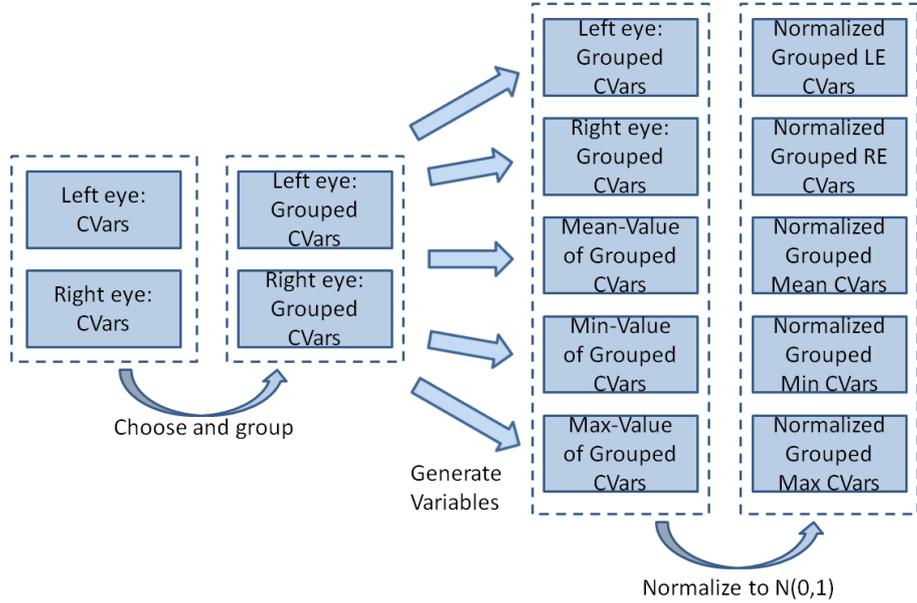
Figure 4.2: Preprocessing and generation of clinical variables

4.2 Experimentation approaches

In the following section we present the two experimentation approaches. We start
with the classification task, predicting only the eye state open or closed, using only
EEGs. In the second part we introduce regression experiments predicting the degree of
the visual field defect which builds the basis for the comparison with the brain graph
interpretation.

4.2.1 Classification: Eyes Open vs. Eyes Closed

As shown in [4], there are differences in the EEG when comparing signals of patients
having their eyes closed (EC) with ones having the eyes open (EO) during the measure-
ment. In the cited work, frequency spectra of the EEG signals, this means data in the
spectral space, after a *DFT* as described in Section 2.2.4 were analyzed for differences.
Although, in this work, the goal of the feature generation was not to be able to separate

EO versus EC, the features might perform well at this task because they do not only contain information about the frequencies in the signal, but also about the time.

Q5: *How well do the features perform when used for classification EO versus EC?*

After all the preprocessing, we use the preprocessed, filtered and grouped datasets for a cross-validation experiment, predicting the classes EO and EC using the *LOO-CV*. We have 25 points available in the constructed feature space for each open and closed eyes, we take 49, learn a classification model, which predicts open or closed for the remaining 50th validation data point. When the prediction is correct the validation error is 0, otherwise 1. The mean prediction error from the 50 possible variants of the validation run forms the result. This process is sketched in Figure 4.3.

Q6: *Which classificator leads to most promising results? Decision tree, support vector classification or naïve bayes?*
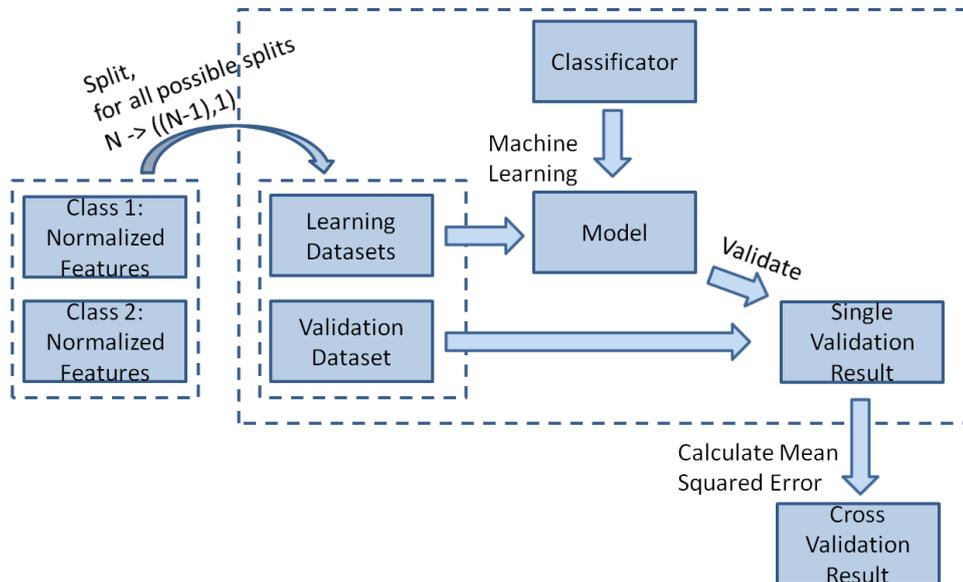

Figure 4.3: Classification process

4.2.2   Regression with Clinical Variables for Closed Eyes

The regression process hardly differs from the classification process, see Figure 4.4. We use *LOO-CV* again. We apply the mean squared error to produce the prediction result. We work with regression trees, ordinary least squares linear regression and

support vector regression as regression algorithms. Instead of using one set of data for each class, the value of the clinical variable is given for every dataset. For different clinical variables, experimentation is performed independently for each of the different eye models.

Q7: *Which eye can be predicted the best? Left? Right? Max? Min? Mean?*
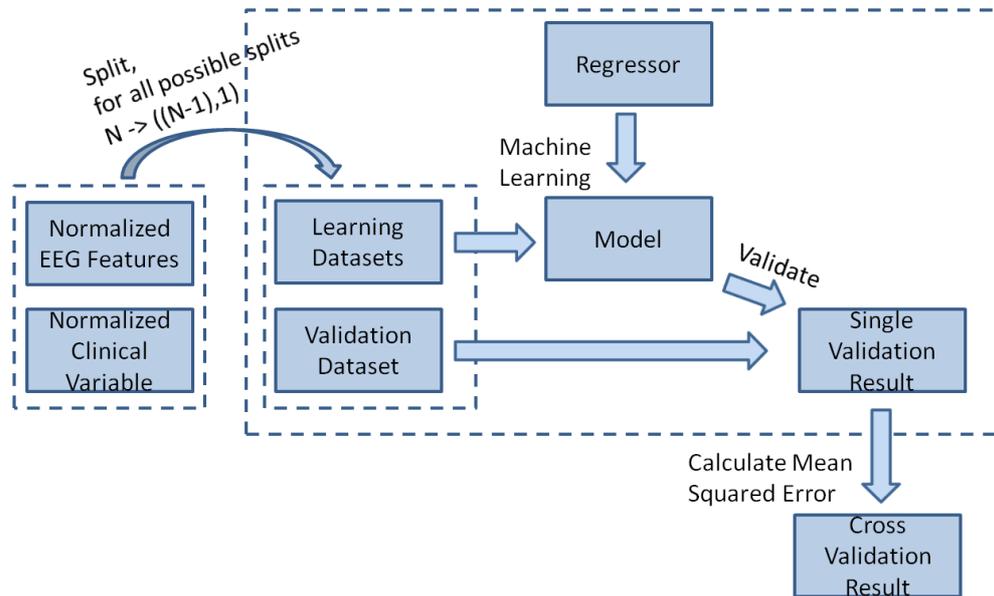
Figure 4.4: Regression process

The clinical variables themselves are also exchangeable in the described regression process. Because we are interested in predicting the visual field defect itself, we use elements of the groups central vision, peripheral vision and global vision. From their *PCA* analysis, we already know roughly how many differences in the results we can expect and so omit the second variable of global vision, which is only available for a small group of patients. This leads to the following set of variables for experimentation:

- mean eccentricity in kinetic perimetry [degrees ED]
- mean threshold in static perimetry (whole 30 degrees visual field)[dB ED]
- far vision (LogMAR scale) [ED]
- near vision (LogMAR scale) [ED]
- foveal threshold in static perimetry [dB ED]
- detection accuracy in whole HRP visual field [% ED]

Note that the groups are included by the order of the list and this order is preserved at the results; staring with peripheral vision, followed by central vision, concluding with global vision.

> **Q8:** *Which clinical variable shows the best results? Are some groups generally better predictable than others?*

> **Q9:** *Which machine learning algorithm leads to the best regression results? Regression Tree, Linear Regression or Support Vector Regression?*

For support vector regression different kernels are possible, thus we compare the results of linear and rbf kernels.

> **Q10:** *Which kernel for the SVM provides better results? Linear or rbf?*

As presented in Chapter 3.1.2, there are different filtering options for patients. We compare no filtering with filtering for noisy EEGs and with filtering using only most reliable results by clinical variables.

> **Q11:** *Do the results improve when filtering for noisy EEGs? Does filtering for clinical variables on fixation ratio and on false positive ratio improve the prediction?*

All the named variations aim only for the main goal, improving our results in order to exceed the approach from Moewes et al. [34] as named before in question **Q1**: Which approach leads to better results? EEG interpretation as classical time series or as brain graph? Only the mean values for the brain graph approach are given, nevertheless they will be integrated into the plot of the best stable results from our approach.

Result Visualization: Boxplots

Boxplots are visualizations for data samples containing numerical attributes. An example for a boxplot of a standard deviated random variable is depicted in Figure 4.5. In this work, the x-axis contains the attributes values, resulting in a horizontal boxplot describing the distribution of the displayed variable. The *median* (or *Q2*) of the sample is marked red inside the box, showing the value, which is bigger or equal than 50% of
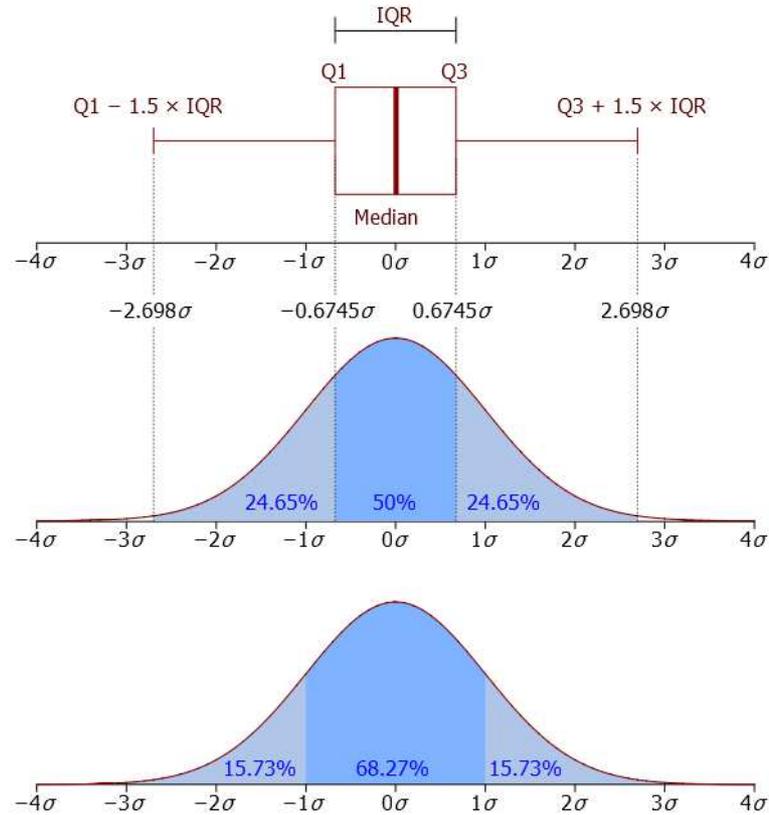
Figure 4.5: Boxplot of a $N(0, \sigma)$ normal distributed variable with the probability density function, source: [50]

the points and at the same time smaller or equal than the other 50%. For an odd number of values, this definition leads to a single value; for an even number every point in the interval between the two inner most points suffices the definition. In this case, the arithmetic mean of those two is used as median. The same procedure, iterated for the two halves, leads to ends of the boxes, indicating that 25% of all samples are left (*Q1*) and 25% are right of the box (*Q3*). The length of the box defines the *Inter Quartile Range (IQR)*. Whiskers at the ends of the box have a maximal length of $1.5 \times IQR$, but they always end at a data point which appears in the sample, so a whisker is shortened until it reaches this data point. Data points outside the whiskers are considered as outliers and are indicated as crosses.

More details about boxplots in general can be found in [6, p. 43-44]. The cross validation results for the prediction of clinical variables are shown as grid of boxplots, each of them showing the *mean squared error (MSE)* of the prediction. So the squared difference of the predicted and the actual value defines a point in the result set. Note that MSE is a common measure for regression and can also be used for classification. In the case of *LOO-CV* the mean of a single value is calculated, which can be skipped; for classification (where we also applied it), squaring does not change anything.

The grid of boxplots was chosen in order to provide an overview, showing the outcomes of each clinical variable per row and those for the eyes in the columns. Clinical variables are grouped by their content, rows 1 and 2 build the group of peripheral vision, rows 3 to 5 stand for central vision and row 6 represents global vision. Note that the predicted eye means in this case that there were separate models created for the available clinical variables of the left eye, the right eye, the mean value of both eyes, the minimal value of both eyes and the maximal value of both eyes. The number of data points available for the experiment equals the number of results and is shown at the left of each boxplot implying the reliability of the model.

Chapter 5

Discussion

The following chapter presents results of the experiments. On their basis the questions in the previous chapters are answered. Most of the inquiries were independent from each other. For that reason several parameters can be chosen for experimentation. Depending on the chosen parameters the experiments may lead to different results. For reasons of clarity and space we do not present all possible combinations, but few chosen examples; enough to answer the questions. On request, the other results can be viewed as well.

We firstly present results of the classification, then the results of the regression task, resolving all questions listed before. We conclude the chapter with a short summary in table form giving an overview of all questions and answers.

5.1   Classification: Eyes Open vs. Eyes Closed

For machine learning, classification is usually easier than regression, because the number of the different outcomes is smaller. We take data from both studies, NO2 and NO3, use different learning algorithms to predict whether the eyes are opened or closed. 25 data points for open and 33 for closed eyes are available, generating a trivial forecast of $33/58 \approx 57\%$, when always predicting closed eyes. As mentioned before we use support vector machines, naïve bayes and decision trees. We use *LOO-CV* for validation. The mean classification results and the variance are shown in Table 5.1, exposing classification results using a decision tree. None of the predictions exceeds 64 %. The results are unstable in the part of the EEG taken, e.g. DB4 achieves 62 % for first and last parts of the signals, failing at random parts with only 40 %. No combination of wavelets produces results that are clearly better than a coin flip. The variance is always high.

The other learners outcomes can be found in the appendix in detail. The summary is

Wavelet	first		last		random	
	mean	variance	mean	variance	mean	variance
Haar	0.31	0.23	0.57	0.25	0.45	0.25
DB2	0.50	0.25	0.48	0.25	0.45	0.25
DB4	0.62	0.24	0.62	0.24	0.40	0.24
DB5	0.50	0.25	0.41	0.25	0.53	0.25
DB20	0.47	0.25	0.59	0.25	0.59	0.25
bior1.3	0.48	0.25	0.64	0.24	0.50	0.25
bior1.5	0.52	0.25	0.50	0.25	0.41	0.25
DB4+DB20	0.43	0.25	0.40	0.24	0.48	0.25
DB4+bior1.3	0.45	0.25	0.55	0.25	0.45	0.25
DB4+bior1.5	0.48	0.25	0.62	0.24	0.33	0.23
bior1.3+DB20	0.24	0.21	0.59	0.25	0.50	0.25
DB4+DB20+bior1.3	0.33	0.23	0.57	0.25	0.52	0.25

Table 5.1: Decision Tree Classification results for eyes open versus eyes closed for studies NO2+NO3, averaging raw channels. EEG-Length(seconds): 32.768, # datasets: 58, #open eyes: 25, # closed eyes: 33, trivial forecast: 0.57

the same: different learners and the second way of processing, not averaging the raw channels, but the features, neither changes the instability in the used EEG parts, nor the average prediction rate close to the coin flip. Reasons may be the short EEG, the chosen channels, the used frequencies or others. Overall the features are unsuitable for predicting the eye state.

Answer For classification the order of the preprocessing (**Q3**), different combinations for mother wavelets (**Q4**), altering the kernel for the SVM (**Q10**) and filtering noisy EEGs (**Q11**) still generates instable (**Q2**) with results close to random guessing.

Answer (**Q6**): Independent from the classification algorithm the results are not better than a coin flip and show a high variance.

Answer (**Q5**): Classification with eyes open versus eyes closed does not work with the given set of features.

5.2 Regression with Clinical Variables for Closed Eyes

Figure 5.1 shows the prediction results of the support vector regression model, with the good results for far vision. The median of the mean squared prediction error is smaller than 0.1 for the models for single eyes, as well as for the minimal value prediction. The single eye's predictions show a remarkable behavior, for the right eye the IQR is tiny, ending below 0.5 without a whisker, showing no outliers with values between 0.5 and 3.7. Between 3.7 and 5.0 there appear 17 prediction outliers. For the left eye the median is equally low but the IQR is huge, ending above 4. Predicting the minimal values results for this variable in very small variance with an equally low median and only two outliers at 0.9 and 1.5. The variables generated from HRP tests generate median prediction errors around 1.0 for the direct prediction of left, right and max; slightly better results around 0.85 are displayed for the mean value and even better are the results for *min* with median values at 0.6. For static perimetry the foveal threshold predictions show stable results with medians around 0.5 for all eyes, Q1 values close to 0, but Q3 values between 1 for max and 2.5 for min. Kinetic perimetry has small, stable inter quartile ranges with medians around 0.5, lower endings at 0.2, higher endings around 1.5 and some outliers.

Figure 5.1: Experimental results containing data from all studies, last parts of the EEGs, using support vector regression with rbf kernel for prediction, wavelets used: DB4+DB20

Comparing these results with the ones using a regression tree, as can be seen in Figure 5.2, the variances are generally bigger, with mostly higher medians for the regression tree. Looking at Figure 5.3, the results for the linear regression model are even worse,

with wider boxes, showing a mean squared error never below 0.5 and often, especially for the HRP results, so bad that values for Q3 and sometimes even for Q2 are outside the scale of maximally a MSE of 5. Note at this point that although for right and left eye the median was below 0.1 for far vision, thus demonstrating a remarkable result, the drawback for the left eye is a Q3 value above 4.0, resulting in a huge IQR. For the right eye the drawback arises with 17 outliers out of the dataset of 85, with a prediction error of 3.8 or above.
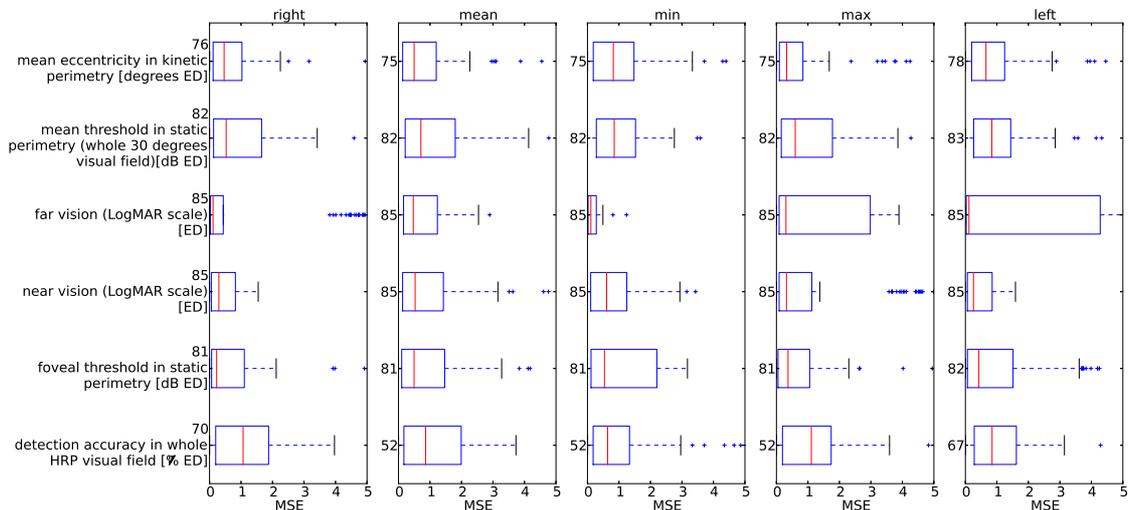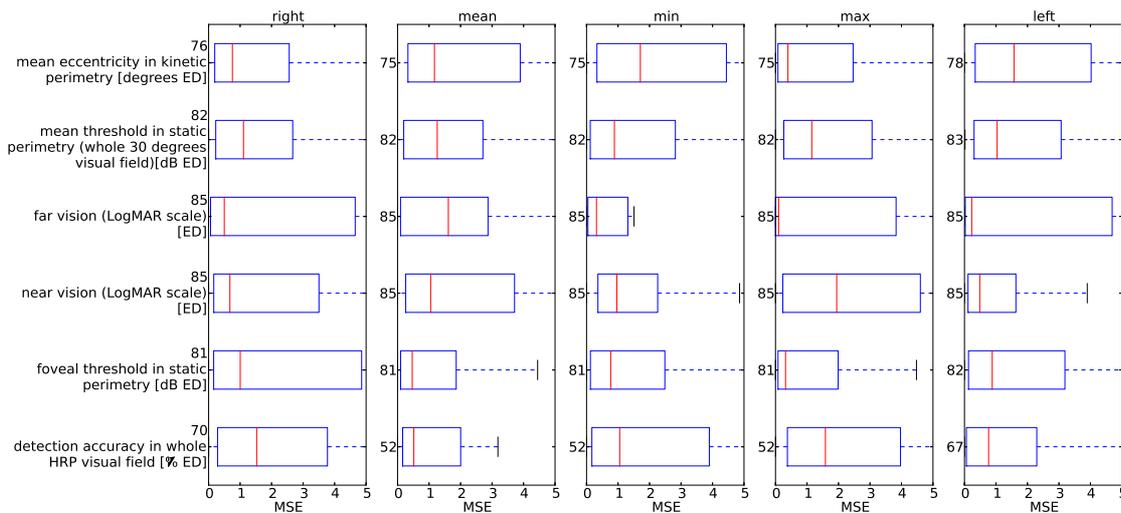

Figure 5.2: Experimental results containing data from all studies, last parts of the EEGs, using regression tree for prediction, wavelets used: DB4+DB20

Answer (**Q9**): The machine learning method producing the overall best regression results, is the support vector regression, followed by the regression tree and the linear regression.

Comparing the Figures 5.1, 2 and 1, shows the prediction's stability. The experiments vary only in the part of the EEG signal used, namely the first part, the last part and a randomly chosen connected part from the middle. The results hardly differ at all.

Answer (**Q2**): For the regression experiments, yes, the features lead to stable prediction results, when varying the parts of the EEG signals.
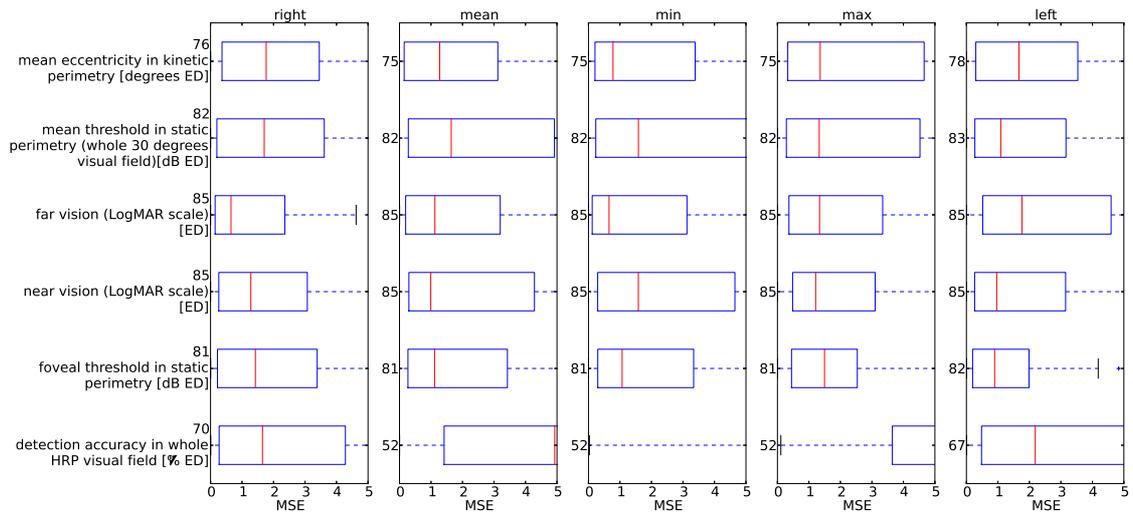
Figure 5.3: Experimental results containing data from all studies, last parts of the EEGs, using linear regression with ordinary least squares for prediction, wavelets used: DB4+DB20

One example for results of experiments for linear and rbf kernel is depicted in Figures 1 and 3, showing that mostly the radial basis function leads to improved results.

> **Answer** (**Q10**): Generally the rbf kernel leads to better results than the linear kernel.

The results for the second way of preprocessing, first generating the features, then averaging and normalizing them, before machine learning is applied is drawn in Figure 5.4. The plot for comparison, only differing in the preprocessing is Figure 5.1. There are only minimal differences visible, like some slight changes in the positions of the outliers. Medians and variances are practically identical.

> **Answer** (**Q3**): The different orders of preprocessing do not show significant differences in the results.

Three examples for comparison of different mother wavelets are depicted in Figures 5.1, 4 and 5 all showing very similar results.

> **Answer** (**Q4**): It does not matter which wavelet or combination of wavelets is used: the results are equal.

For filtering as default example Figure 5.1 is used. Compared with Figure 6 it shows the results when filtering for noisy EEGs. Generally, the results do not differ much.
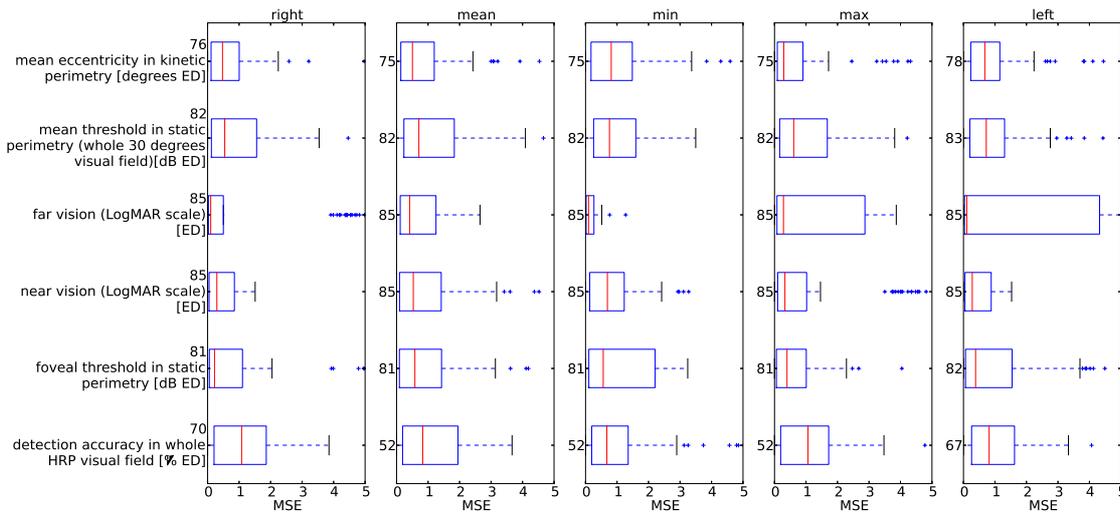
Figure 5.4: Experimental results containing data from all studies, last parts of the EEGs, using support vector regression with rbf kernel for prediction, wavelets used: DB4+DB20, preprocessing first generating the features, then averaging them

Of course the reliability decreases when performing the filter, which does not lead to an increased result here, so it should be avoided. Another interesting point can be seen by watching the outliers, which were already mentioned at the beginning of this chapter, visible at far vision at right and near vision at max. They do not disappear, but obviously some of the well predicted points are filtered, resulting in an increased IQR in both cases. A weak side of the boxplot emerges when focusing on Figure 5.1 on right and far vision: When filtering out more and more well predicted points, the box suddenly increases its size to the point earlier seen as leftmost outlier and adding the whisker to the rightmost outlier. It then looks like a completely different, much worse result distribution. Staying at the same Figure, switching from the prediction of the right eye to the left shows a huge IQR with Q3 above 4.0. This might be caused by the same effect and cannot be decided at this point, when only taking the boxplots as sources of information.

Filtering for clinical variables leads to Figure 5.5 and to outstanding results, especially at those points recently discussed. For the right eye, far vision, the box lies between 0.015 and 0.125, with the median at 0.06; for the left eye Q1 is 0.026, Q2 = 0.012 and Q3 shows 0.27. Comparing also the other eyes and variables, this filtering usually leads to smaller IQRs and frequently to improved medians, although, because of less subjects being analyzed, the reliability is slightly reduced.
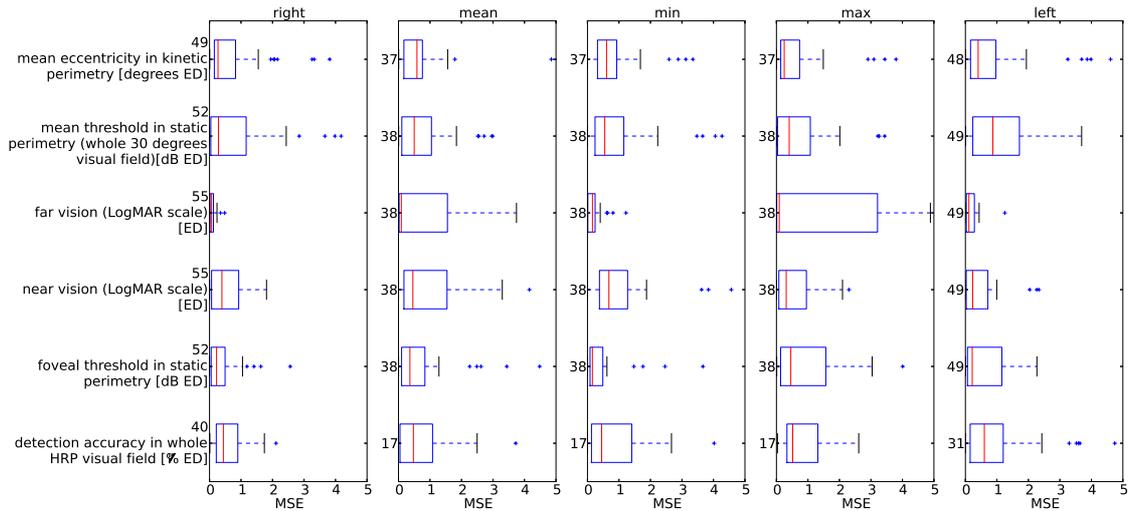
Figure 5.5: Experimental results containing data from all studies, filtering patients with clinical variable dependencies, last parts of the EEGs, using support vector regression with rbf kernel for prediction, wavelets used: DB4+DB20. Patients are filtered, if the fixation ratio is lower than 90% or the false positive ratio is higher than 5%.

> **Answer (Q11):** Filtering generally reduces the number of subjects and thus the tests reliability. Filtering for noisy EEGs does not show a positive effect, but skipping subjects with unreliable test results in clinical variables reduces the number of outliers massively and mostly increases the average prediction.

Figure 5.5 clarifies that values for the group of central vision can be predicted the best, for clinical variables lower prediction errors show far vision, followed by foveal threshold, which is also well predictable, with medians beyond 0.45 and boxes around 0.15 and 0.9.

> **Answer (Q8):** The best predictable clinical variable has shown to be far vision with great results. Still good, but no more outstanding were the results of near vision. Looking at the groups of clinical variables, central vision is clearly best predictable, followed by peripheral vision and the global vision.

> **Answer (Q7):** In the given dataset, models predicting right eyes mostly show best results. The ones for left eyes are also good; predicting *min* follows next. Results for max and mean values form the end.

The results for brain graphs, provided by the authors of [34], are comparable to our results. The mean squared prediction errors from a *LOO-CV* are given for models for

the right and for the left eyes. Due to their processing, two values – one for weighted graphs and one for thresholded graphs – are given. These values are included as magenta (for thresholded) and green (for weighted) lines in Figure 5.6.
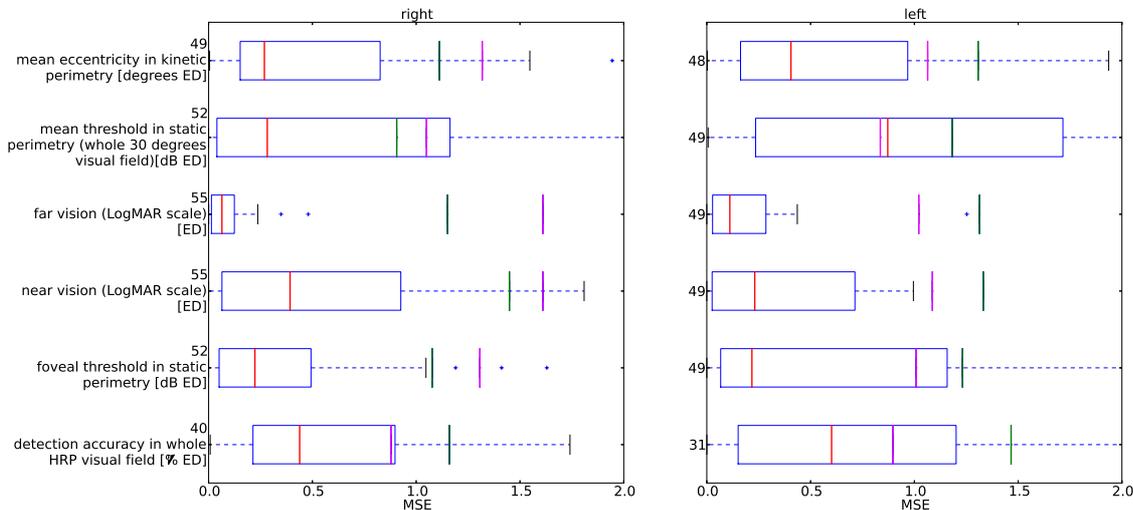

Figure 5.6: Experimental results containing data from all studies, filtering patients with clinical variable dependencies, last parts of the EEGs, using support vector regression with rbf kernel for prediction, wavelets used: DB4+DB20. Patients are filtered, if the fixation ratio is lower than 90% or the false positive ratio is higher than 5%. Results from brain graphs are included: green medians indicate weighted graphs, magenta medians indicate thresholded graphs.

In all cases except mean threshold in static perimetry for the left eye, the time series interpretation clearly outperforms both variants of brain graphs. It is remarkable that there are clinical variables, like far vision, which are well predictable with time series and relatively bad with brain graphs and there are others, like detection accuracy in HRP, where it is the other way around. Time series are still better in this case. When considering that research on brain graphs is still in progress, that only 25 data points were used [34] and that the experiments used only linear models. Future research promises improved results for this approach. Assuming that brain graphs and time series continue to show their best prediction quality for different clinical variables, meta models that combine these two approaches, could lead to excellent results for several groups of clinical variables.

This chapter presented results and answers to all highlighted questions. A short summary is presented in Table 5.2. It gives an overview about all answers at a glance. The following chapter concludes the thesis, sums up the most important results and offers approaches for future work.

	Question	Answer
Q1	Brain graphs or time series, which is better?	Time series mostly reveal better results.
Q2	Are the features stable?	The features are stable for regression but unstable for classification.
Q3	Which preprocessing order is better?	The order of preprocessing does not significantly change the results.
Q4	Which set of wavelets is the best?	Different Wavelets and combinations of them lead to similar results.
Q5	Does classification work?	No.
Q6	Classification: Which machine learning method works best?	There was no difference, classification does not work.
Q7	Which eye models are best predictable?	Right, Left and Min are the best predictable eye models.
Q8	Which clinical variable and which group are best predictable?	Far vision builds the best clinical variable. The best predictable group is central vision.
Q9	Regression: Which machine learning method works best?	Support Vector Regression works best, followed by Regression Tree and Linear Regression.
Q10	Which SVM kernel works better?	In classification both results were negative, for regression rbf kernel outperforms the linear.
Q11	Does filtering help?	Yes, filtering subjects with unreliable tests reduces outliers and mostly increases the average prediction.

Table 5.2: Overview of questions and answers

Chapter 6

Conclusion and Future Work

In the following, results are summarized and a quick review about gained knowledge is given. This section concludes with an outlook about possible extensions, improvements and future research in the topic of predicting visual field damage using electroencephalography.

6.1   Conclusion

This scientific work proposed a method to create predictions for several clinical variables. The average difference between predictions for visual acuity far vision and the actually measured values is less than 0.15 times the signals standard deviation. We showed this to be true for models predicting both eyes separately. We applied several filtering approaches and concluded that filtering out patients with such clinical variables indicating unreliable results of the perimetric tests, leads to a reduced number of outliers and a smaller variance in the predictions. Filtering out patients with noisy EEGs did not improve our findings. We could show the temporal stability of the generated features, by varying the parts of the patient's EEGs and reaching similar results. The general idea of transforming the time series with a wavelet transformation to a spatiotemporal space, which built the basis of our machine learning features, succeeded. Different combinations of mother wavelets produced similar effects. Support vector regression with a radial basis function kernel showed the best outcomes.

The generated features are unsuitable for the classification whether the patient's eyes are open or closed, which might be the case, because of the shorter EEG parts available for the measurements with open eyes.

The primary goal was to compare the results of our approach, which applies classical time series analysis on the EEGs to the approach of generating brain graphs from the EEGs. It turned out that time series analysis is the concept that generates more exact predictions for most of the tested clinical variables. The group of variables with the best forecast was central vision.

6.2 Future Work

The time series approach outperformed the brain graphs, whose disadvantage was the small number of only 25 subjects. So for reliability reasons a filtering was not possible. Future work could extend the comparison by providing a larger number of subjects.
The results imply that the qualities of the two approaches highly depend on the predicted clinical variable. This could be used in building a meta model containing both EEG interpretations and both versions for brain graphs, thus combining the advantages of all models.

The general approach to generating different specific numbers as clinical variables from the grid, which contains the result of a perimetric test, could be altered. Directly predicting the grid itself remains open for future research. In this case it would be possible to analyze parts of the brain in detail, e.g. to what extent they are responsible for specific diseases.

In addition to getting the clinical variables and using them in a more direct way, one could try to get more detailed EEG data from more channels in a higher local resolution or in a higher spatial resolution to get more information which might show more subtle distinctions.

There are several possibilities of nerve injuries which can affect the visual field. In our experiments we did not distinguish between these different diseases because of the low number of subjects in each group. One assumption in [25] is that different diseases cause different changes in the EEG, so that the usually easier task to classify diseases could be applied instead of a regression on the measured variables.

One shortcoming of clinical studies with real patients can be seen in the point that the number of subjects is very small, especially for data mining. For high reliability it is necessary to get more data points and therefore more patients for the machine learning algorithms, especially when the dimensionality of the data points increases. In order to gain a deeper understanding which features actually lead to good prediction results, a feature selection could be performed. This might show which EEG channels, which

frequency bands or which of the three feature sets mean, variance and the number of outliers influence the outcomes.

Bibliography

[1] Peter Achermann. "EEG analysis applied to sleep." In: *Epileptologie* 26 (2009), pp. 28–33.

[2] Ralph G Andrzejak, Klaus Lehnertz, Florian Mormann, Christoph Rieke, Peter David, and Christian E Elger. "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state." In: *Physical Review E* 64.6 (2001), p. 061907.

[3] Sylvain Arlot and Alain Celisse. "A survey of cross-validation procedures for model selection." In: *Statistics Surveys* 4 (2010), pp. 40–79.

[4] Robert J Barry, Adam R Clarke, Stuart J Johnstone, Christopher A Magee, and Jacqueline A Rushby. "EEG differences between eyes-closed and eyes-open resting conditions." In: *Clinical Neurophysiology* 118.12 (2007), pp. 2765–2773.

[5] Ehrhard Behrends. "Beschreibende Statistik." In: *Elementare Stochastik*. Springer, 2013, pp. 269–286.

[6] Michael R Berthold, Christian Borgelt, and Frank Höppner. *Guide to Intelligent Data Analysis*. Vol. 42. Springer, 2010.

[7] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. "When is "nearest neighbor" meaningful?" In: *Database Theory - ICDT'99* (1999), pp. 217–235.

[8] Christian Borgelt. "Data Mining with graphical models." PhD thesis. Otto-von-Guericke-Universtät Magdeburg, Universitätsbibliothek, 2000.

[9] Thomas H Budzynski, Helen Kogan Budzynski, James R Evans, and Andrew Abarbanel. *Introduction to quantitative EEG and neurofeedback: Advanced theory and applications*. Academic Press, 2009.

[10] Catherine J Chu, Mark A Kramer, Jay Pathmanathan, Matt T Bianchi, M Brandon Westover, Lauren Wizon, and Sydney S Cash. "Emergence of stable functional networks in long-term human electroencephalography." In: *The Journal of Neuroscience* 32.8 (2012), pp. 2703–2713.

[11] Robert P Cubbage. *Visual fields*. Butterworth-Heinemann Medical, 2005.

[12] P De Chazal and RB Reilly. "A comparison of the use of different wavelet coefficients for the classification of the electrocardiogram." In: *Pattern Recognition, 2000. Proceedings. 15th International Conference on*. Vol. 2. IEEE. 2000, pp. 255–258.

[13] Arnaud Delorme and Scott Makeig. "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis." In: *Journal of neuroscience methods* 134.1 (2004), pp. 9–21.

[14] Arnaud Delorme, Terrence Sejnowski, and Scott Makeig. "Enhanced detection of artifacts in EEG data using higher-order statistics and independent component analysis." In: *Neuroimage* 34.4 (2007), pp. 1443–1449.

[15] Pierre Duhamel and Martin Vetterli. "Fast Fourier transforms: a tutorial review and a state of the art." In: *Signal processing* 19.4 (1990), pp. 259–299.

[16] A Fedorov, S Jobke, V Bersnev, A Chibisova, Y Chibisova, C Gall, and BA Sabel. "Restoration of vision after optic nerve lesions with noninvasive transorbital alternating current stimulation: a clinical observational study." In: *Brain stimulation* 4.4 (2011), pp. 189–201.

[17] Inan Güler and Elif Derya Übeyli. "Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients." In: *Journal of neuroscience methods* 148.2 (2005), pp. 113–121.

[18] Frédéric Gougoux, Robert J Zatorre, Maryse Lassonde, Patrice Voss, and Franco Lepore. "A functional neuroimaging study of sound localization: visual cortex activity predicts performance in early-blind individuals." In: *PLoS biology* 3.2 (2005), e27.

[19] Isabelle Guyon and André Elisseeff. "An introduction to variable and feature selection." In: *The Journal of Machine Learning Research* 3 (2003), pp. 1157–1182.

[20] Trevor. Hastie, Robert. Tibshirani, and J Jerome H Friedman. *The elements of statistical learning*. Vol. 1. Springer New York, 2001.

[21] Pascal Held, Christian Moewes, Christian Braune, Rudolf Kruse, and Bernhard A Sabel. "Advanced Analysis of Dynamic Graphs in Social and Neural Networks." In: *Towards Advanced Data Analysis by Combining Soft Computing and Statistics*. Springer, 2013, pp. 205–222.

[22] Erich Kasten, Stefan Wüst, Wolfgang Behrens-Baumann, and Bernhard A Sabel. "Computer-based training for the treatment of partial blindness." In: *Nature medicine* 4.9 (1998), pp. 1083–1087.

[23] JH Kelsey. "Blindness and the Electrical Activity of the Brain. Electroencephalographic Studies of the Effects of Sensory Impairment." In: *The British Journal of Ophthalmology* 59.3 (1975), p. 172.

[24] Kristin Koch, Judith McLean, Ronen Segev, Michael A Freed, Michael J Berry II, Vijay Balasubramanian, and Peter Sterling. "How Much the Eye Tells the Brain." In: *Current Biology* 16.14 (2006), pp. 1428–1434.

[25] Markus Köppen. "Abhängigkeiten zwischen Phase Lag Index von elektroenzephalografischen Netzen und klinischen Variablen visuell geschädigter Probanden." MA thesis. Otto-von-Guericke-Universität Magdeburg, 2012.

[26] Anne Kriegseis, Erwin Hennighausen, Frank Rösler, and Brigitte Röder. "Reduced EEG alpha activity over parieto-occipital brain areas in congenitally blind adults." In: *Clinical neurophysiology* 117.7 (2006), pp. 1560–1573.

[27] Gerhard Klaus Lang. *Augenheilkunde.* Georg Thieme Verlag, 2008.

[28] Wei-Yin Loh. "Classification and regression tree methods." In: *Encyclopedia of statistics in quality and reliability* (2008).

[29] Haiping Lu, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos. "Regularized common spatial patterns with generic learning for EEG signal classification." In: *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE.* IEEE. 2009, pp. 6599–6602.

[30] Scott Makeig, Anthony J. Bell, Tzyy-Ping Jung, and Terrence J. Sejnowski. "Independent Component Analysis of Electroencephalographic Data." In: *Advances in Neural Information Processing Systems.* Ed. by David S. Touretzky, Michael C. Mozer, and Michael E. Hasselmo. Vol. 8. Cambridge, MA: MIT Press, 1996, pp. 145–151.

[31] Tom M Mitchell. *Machine learning. WCB.* 1997.

[32] Christian Moewes and Rudolf Kruse. "Unification of fuzzy SVMs and rule extraction methods through imprecise domain knowledge." In: *Proceedings of the International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU-08).* 2008, pp. 1527–1534.

[33] Christian Moewes and Rudolf Kruse. "The Effects of Edge Weights on Correlating Dynamical Networks: Comparing Unweighted and Weighted Brain Graphs of nervus opticus Patients." In: *Proceedings of the 5th International Joint Conference on Computational Intelligence.* Ed. by Agostinho Rosa, António Dourado, Kurosh Madani, Joaquium Filipe, and Janusz Kacprzyk. Villamoura, Algarve, Portugal: ScitePress, Sept. 2013. URL: http://www.fcta.ijcci.org/.

[34] Christian Moewes, Rudolf Kruse, and Bernhard A Sabel. "Analysis of Dynamic Brain Networks Using VAR Models." In: *Synergies of Soft Computing and Statistics for Intelligent Data Analysis* (2012), pp. 525–532.

[35] J. Ross Quinlan. "Induction of decision trees." In: *Machine learning* 1.1 (1986), pp. 81–106.

[36] DA Rosser, DAH Laidlaw, and IE Murdoch. "The development of a "reduced logMAR" visual acuity chart for use in routine clinical practice." In: *British Journal of Ophthalmology* 85.4 (2001), pp. 432–436.

[37] Bernhard A Sabel, Anton B Fedorov, Nicole Naue, Antonia Borrmann, Christoph Herrmann, and Carolin Gall. "Non-invasive alternating current stimulation improves vision in optic neuropathy." In: *Restorative neurology and neuroscience* 29.6 (2011), pp. 493–505.

[38] John Shawe-Taylor and Nello Cristianini. *Kernel methods for pattern analysis.* Cambridge university press, 2004.

[39] Stefan Silbernagl. *Taschenatlas Physiologie.* Thieme, 2012.

[40] Farmy Silva, Oscar Arias-Carrión, Silmar Teixeira, Bruna Velasques, Caroline Peressutti, Flávia Paes, Luis F Basile, Manuel Menéndez-González, Eric Murillo-Rodríguez, Mauricio Cagy, et al. "Functional coupling of sensorimotor and associative areas during a catching ball task: a qEEG coherence study." In: *International archives of medicine* 5.1 (2012), pp. 1–9.

[41] Alex J Smola and B Schölkopf. *Learning with kernels.* Citeseer, 1998.

[42] Christopher Torrence and Gilbert P Compo. "A practical guide to wavelet analysis." In: *Bulletin of the American Meteorological society* 79.1 (1998), pp. 61–78.

[43] Francisco Varela, Jean-Philippe Lachaux, Eugenio Rodriguez, and Jacques Martinerie. "The brainweb: phase synchronization and large-scale integration." In: *Nature reviews neuroscience* 2.4 (2001), pp. 229–239.

[44] Brani Vidakovic and Peter Müller. "Wavelets for kids." In: *Instituto de Estadística, Universidad de Duke* (1994).

[45] V Virsu. "An estimation and application of the human cortical magnification factor." In: *Experimental Brain Research* 37.3 (1979), pp. 495–510.

[46] Harry Zhang. "The optimality of naïve Bayes." In: *A A* 1.2 (2004), p. 3.

[47] Jing Zhou, Robert J Schalkoff, Brian C Dean, and Jonathan J Halford. "Morphology-based wavelet features and multiple mother wavelet strategy for spike classification in EEG signals." In: *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE.* IEEE. 2012, pp. 3959–3962.

Online References

[48] Crosstemplejay. *LogMAR chart*. Retrieved: 2013-07-04. 2011. URL: `http://en.wikipedia.org/wiki/File:LogMAR_chart.jpg`.

[49] Marius t Hart. *10 20 EEG electrode Schema*. Retrieved: 2013-07-28. 2008. URL: `http://www.mariusthart.net/downloads/eeg_electrodes_10-20.svg`.

[50] Jhguch. *Boxplot vs PDF*. Retrieved: 2013-07-04. Mar. 2011. URL: `http://en.wikipedia.org/wiki/File:Boxplot_vs_PDF.svg`.

[51] Der Lange. *EEG Plot with 3Hz spike discharges in a child with childhood absence epilepsy*. Retrieved: 2013-07-15. June 2006. URL: `http://en.wikipedia.org/wiki/File:Spike-waves.png`.

Appendix

Wavelet	first		last		random	
	mean	variance	mean	variance	mean	variance
Haar	0.55	0.25	0.62	0.24	0.50	0.25
DB2	0.57	0.25	0.50	0.25	0.53	0.25
DB4	0.52	0.25	0.50	0.25	0.53	0.25
DB5	0.57	0.25	0.53	0.25	0.47	0.25
DB20	0.59	0.25	0.48	0.25	0.52	0.25
bior1.3	0.60	0.24	0.60	0.24	0.53	0.25
bior1.5	0.59	0.25	0.62	0.24	0.52	0.25
DB4+DB20	0.57	0.25	0.50	0.25	0.47	0.25
DB4+bior1.3	0.57	0.25	0.57	0.25	0.53	0.25
DB4+bior1.5	0.53	0.25	0.57	0.25	0.59	0.25
bior1.3+DB20	0.59	0.25	0.53	0.25	0.53	0.25
DB4+DB20+bior1.3	0.57	0.25	0.52	0.25	0.53	0.25

Table 1: Naïve Bayes classification results for eyes open versus eyes closed for studies NO2+NO3, averaging raw channels. EEG-Length(seconds): 32.768, #datasets: 58, #open eyes: 25, #closed eyes: 33, trivial forecast: 0.57

Wavelet	first		last		random	
	mean	variance	mean	variance	mean	variance
Haar	0.55	0.25	0.52	0.25	0.50	0.25
DB2	0.45	0.25	0.60	0.24	0.55	0.25
DB4	0.45	0.25	0.62	0.24	0.50	0.25
DB5	0.47	0.25	0.53	0.25	0.43	0.25
DB20	0.43	0.25	0.62	0.24	0.59	0.25
bior1.3	0.43	0.25	0.57	0.25	0.43	0.25
bior1.5	0.43	0.25	0.53	0.25	0.40	0.24
DB4+DB20	0.43	0.25	0.53	0.25	0.48	0.25
DB4+bior1.3	0.47	0.25	0.52	0.25	0.67	0.23
DB4+bior1.5	0.50	0.25	0.62	0.24	0.45	0.25
bior1.3+DB20	0.34	0.24	0.36	0.24	0.41	0.25
DB4+DB20+bior1.3	0.33	0.23	0.43	0.25	0.41	0.25

Table 2: Support Vector classification results for eyes open versus eyes closed for studies NO2+NO3, averaging raw channels. EEG-Length(seconds): 32.768, #datasets: 58, #open eyes: 25, #closed eyes: 33, trivial forecast: 0.57

wavelet	first		last		random	
	mean	variance	mean	variance	mean	variance
Haar	0.59	0.25	0.62	0.24	0.57	0.25
DB2	0.55	0.25	0.53	0.25	0.55	0.25
DB4	0.52	0.25	0.52	0.25	0.48	0.25
DB5	0.53	0.25	0.52	0.25	0.50	0.25
DB20	0.59	0.25	0.53	0.25	0.52	0.25
bior1.3	0.60	0.24	0.53	0.25	0.48	0.25
bior1.5	0.55	0.25	0.60	0.24	0.53	0.25
DB4+ DB20	0.53	0.25	0.55	0.25	0.50	0.25
DB4+ bior1.3	0.55	0.25	0.52	0.25	0.55	0.25
DB4+ bior1.5	0.52	0.25	0.53	0.25	0.53	0.25
bior1.3+ DB20	0.57	0.25	0.53	0.25	0.55	0.25
DB4+ DB20+ bior1.3	0.55	0.25	0.53	0.25	0.52	0.25

Table 3: Naïve Bayes classification results for eyes open versus eyes closed for studies NO2+NO3, averaging features. EEG-Length(seconds): 32.768, #datasets: 58, #open eyes: 25, #closed eyes: 33, trivial forecast: 0.57

wavelet	first		last		random	
	mean	variance	mean	variance	mean	variance
Haar	0.57	0.25	0.62	0.24	0.62	0.24
DB2	0.55	0.25	0.45	0.25	0.50	0.25
DB4	0.48	0.25	0.52	0.25	0.45	0.25
DB5	0.55	0.25	0.5	0.25	0.36	0.24
DB20	0.45	0.25	0.53	0.25	0.55	0.25
bior1.3	0.47	0.25	0.64	0.24	0.52	0.25
bior1.5	0.47	0.25	0.57	0.25	0.66	0.24
DB4+ DB20	0.47	0.25	0.45	0.25	0.62	0.24
DB4+ bior1.3	0.43	0.25	0.53	0.25	0.60	0.24
DB4+ bior1.5	0.43	0.25	0.48	0.25	0.48	0.25
bior1.3+ DB20	0.40	0.24	0.50	0.25	0.41	0.25
DB4+ DB20+ bior1.3	0.40	0.24	0.47	0.25	0.47	0.25

Table 4: Decision Tree classification results for eyes open versus eyes closed for studies NO2+NO3, averaging features. EEG-Length(seconds): 32.768, #datasets: 58, #open eyes: 25, #closed eyes: 33, trivial forecast: 0.57

wavelet	first		last		random	
	mean	variance	mean	variance	mean	variance
Haar	0.55	0.25	0.53	0.25	0.52	0.25
DB2	0.41	0.25	0.52	0.25	0.50	0.25
DB4	0.48	0.25	0.58	0.25	0.41	0.25
DB5	0.48	0.25	0.45	0.25	0.45	0.25
DB20	0.29	0.23	0.60	0.24	0.38	0.24
bior1.3	0.48	0.25	0.60	0.24	0.38	0.24
bior1.5	0.33	0.23	0.64	0.24	0.55	0.25
DB4+ DB20	0.47	0.25	0.55	0.25	0.38	0.24
DB4+ bior1.3	0.41	0.25	0.52	0.25	0.52	0.25
DB4+ bior1.5	0.38	0.24	0.59	0.25	0.59	0.25
bior1.3+ DB20	0.47	0.25	0.47	0.25	0.45	0.25
DB4+ DB20+ bior1.3	0.48	0.25	0.48	0.25	0.64	0.24

Table 5: Support Vector classification results for eyes open versus eyes closed for studies NO2+NO3, averaging features. EEG-Length(seconds): 32.768, #datasets: 58, #open eyes: 25, #closed eyes: 33, trivial forecast: 0.57
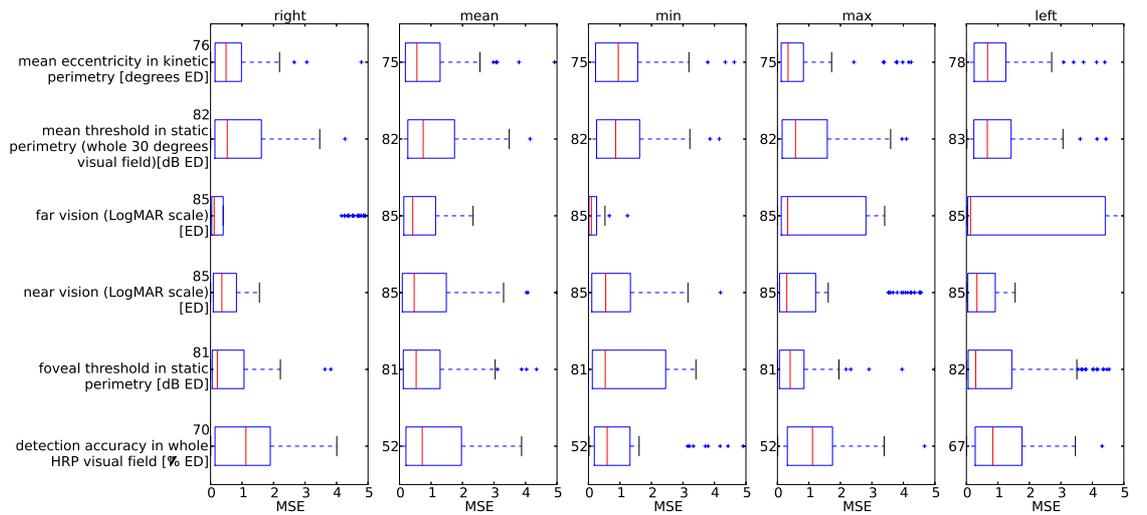

Figure 1: Experimental results containing data from all studies, first parts of the EEGs, using Support Vector regression with rbf kernel for prediction, wavelets used: DB4+DB20
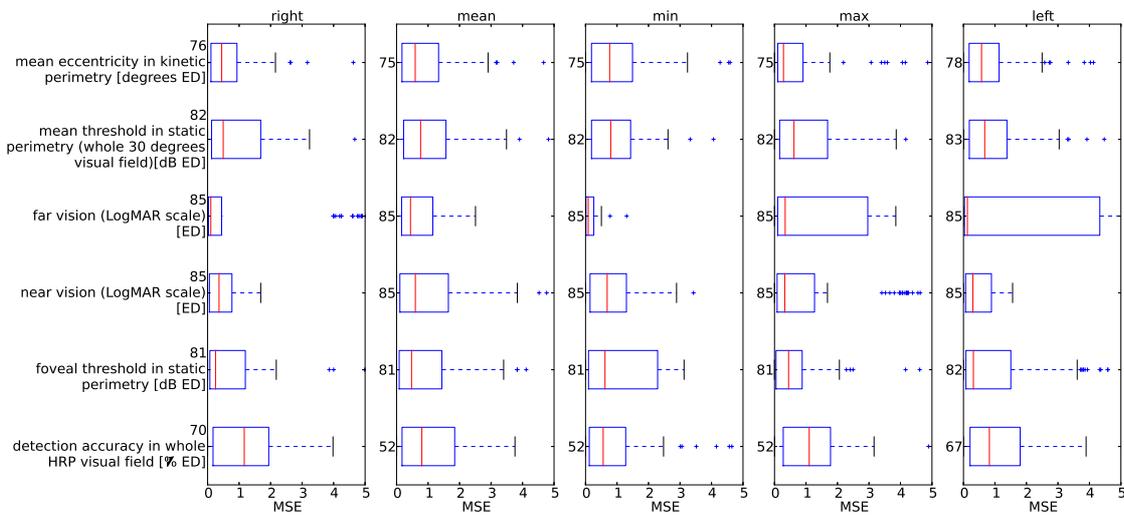
Figure 2: Experimental results containing data from all studies, random parts of the EEGs, using Support Vector regression with rbf kernel for prediction, wavelets used: DB4+DB20
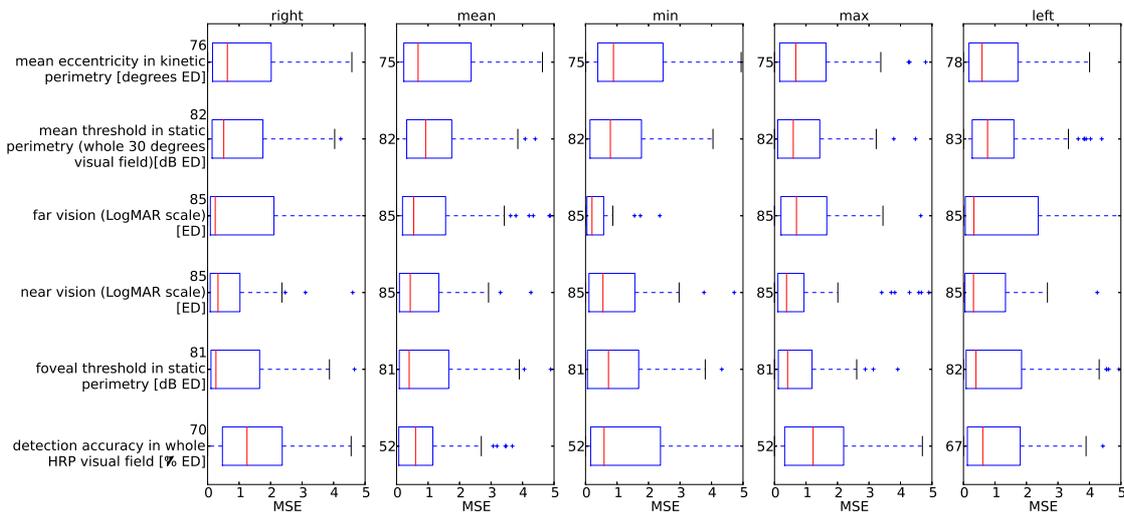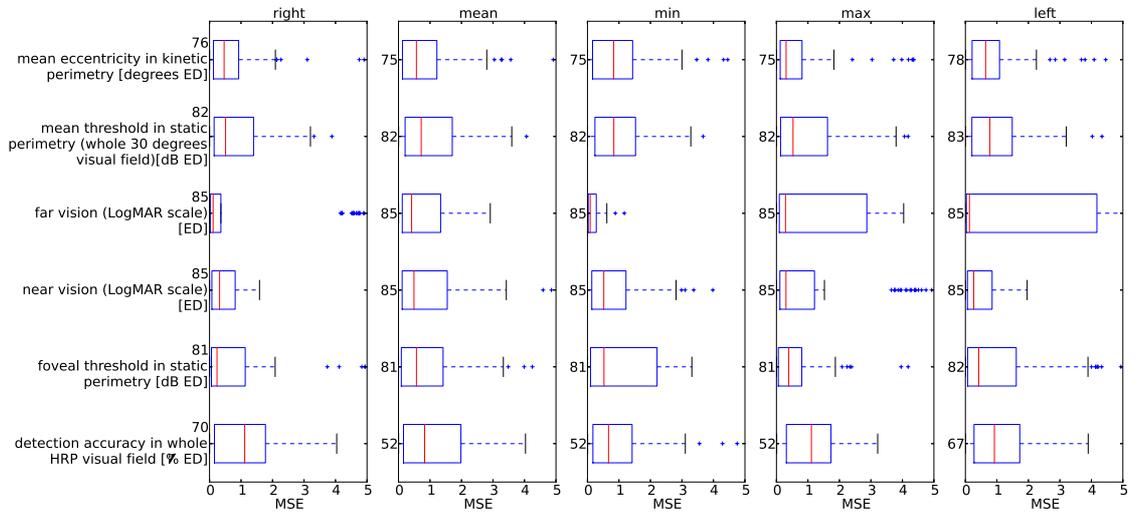

Figure 3: Experimental results containing data from all studies, first parts of the EEGs, using Support Vector regression with linear kernel for prediction, wavelets used: DB4+DB20

Figure 4: Experimental results containing data from all studies, last parts of the EEGs, using Support Vector regression with rbf kernel for prediction, wavelet used: Haar
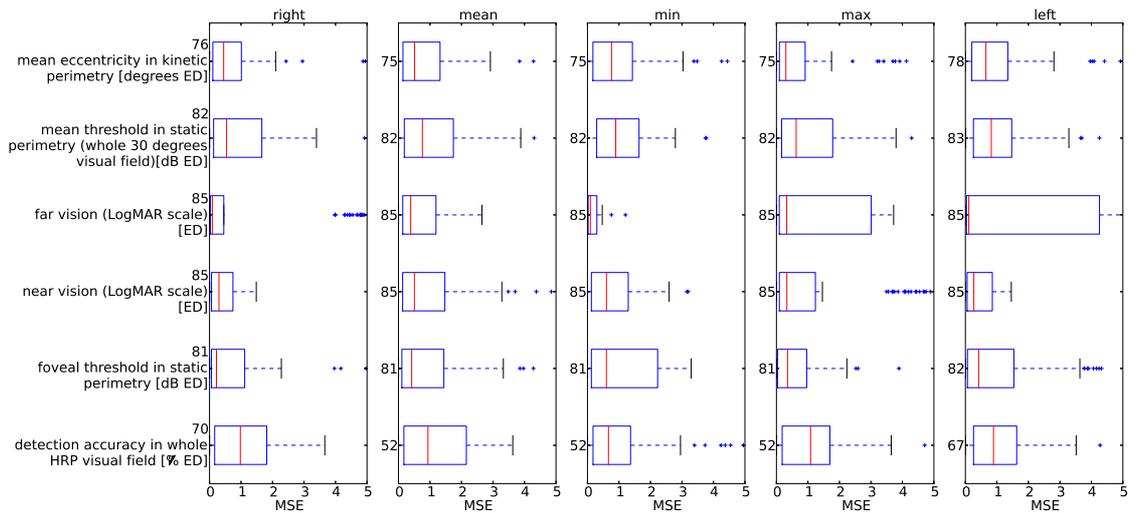

Figure 5: Experimental results containing data from all studies, last parts of the EEGs, using Support Vector regression with rbf kernel for prediction, wavelet used: DB20
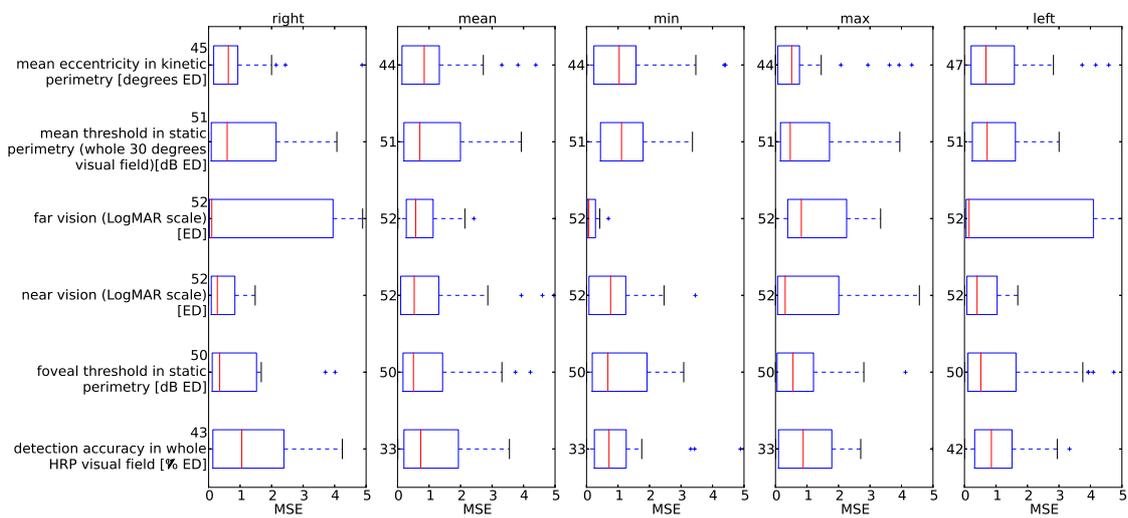
Figure 6: Experimental results containing data from all studies, filtering patients with noisy EEG, last parts of the EEGs, using Support Vector regression with rbf kernel for prediction, wavelets used: DB4+DB20

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die vorliegende Masterarbeit selbstständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Stellen sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch nicht veröffentlicht.

Magdeburg, den 26. August 2013

Christoph Doell