TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technologies

Kaspar Haavajõe, 183827 IVSM

Depression discrimination based on visually evoked potentials of electroencephalography

Master's thesis

Supervisor: Laura Päeske
PhD

Tallinn 2021

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Kaspar Haavajõe, 183827 IVSM

Depressiooni tuvastamine elektroentsefalograafiast visuaalselt esilekutsutud potentsiaalide alusel

magistritöö

Juhendaja: Laura Päeske
PhD

Tallinn 2021

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author: Kaspar Haavajõe

10.05.2021

Abstract

As more people are today living with depression, a detection of depression using technological solutions is of increasing importance. The aim of current thesis is to build a machine learning model that would distinguish depression between subjects with depression and healthy control group using visually evoked potentials extracted from EEG.

Previously recorder EEG sessions together with evoked potentials of 40 subjects are used as an input. Evoked potentials are the response to the 3 seconds of external image stimuli presented to the subjects.

Subcomponents of extracted visually evoked potentials extracted from the EEG will be used as features. Significance of feature will be analysed, selection is carried out and top-performing features will be used for training the model.

The main outcome of the current thesis is a classifier model that can differentiate the subjects with depression from the healthy controls providing a prediction performance approximately 86% using multiple features selected by genetic algorithm and validated with multiple cross-validation methods.

This thesis is written in English and is 33 pages long, including 8 chapters, 10 figures and 3 tables.

Annotatsioon

Depressiooni tuvastamine elektroentsefalograafiast visuaalselt esilekutsutud potentsiaalide alusel

Erinevate tehnoloogiliste lahenduste väljatöötamine depressiooni tuvastamiseks on muutunud väga oluliseks kuna depressioonis inimeste arv on pidevas tõusutrendis. Antud töö eesmärk on treenida selline masinõppe mudel, mis suudaks eristada deppressioonis patsiente kontrollrühmast kasutades selleks visuaalselt esilekutsutud potentsiaale.

Eelnevalt salvestatud EEG sessioonide andmed 40 uuritavalt koos nendega seotud visuaalselt esilekutsutud potentsiaalidega on antud töö sisendiks. Andmetes on potentsiaalid esile kutsutud näidates uuritavatele pilte stiimulitena 3 sekundilise intervalliga.

EEG andmetest leitakse esilekutsutud potentsiaalide alamkomponendid, mida kasutatakse tunnustena. Nende tunnuste statistilist olulisust hinnatakse, seejärel viiakse läbi tunnuste selektsioon. Parimaid tunnuseid kasutatakse masinõppe mudeli treenimiseks.

Antud töö tulemuseks on masinõppe mudel, mis suudab eristada depressioonis patsiente kontrollrühmast täpsusega 86% kasutades selleks mitud tunnust, mis on varasemalt leitud geneetilise algoritmi abil. Tulemusi valideeritakse mitme erineva ristvalideerimise meetodi abil.

Lõputöö on kirjutatud inglise keeles ning sisaldab teksti 33 leheküljel, 8 peatükki, 10 joonist, 3 tabelit.

List of abbreviations and terms

ISCEV	International Society for Clinical Electrophysiology of Vision
ICA	Independent component analysis
VEP	Visually evoked potential
LR	Logistic regression
SVM	Support vector machine
EEG	Electroencephalogram
WHO	World Health Organization
DFA	Detrended fluctuation analysis
MDD	Major depressive disorder
LZC	Lempel-Ziv complexity
MLZC	Multiscale Lempel-Ziv complexity
SASI	Spectral asymmetry index
AEP	Auditory evoked potential
ANOVA	Analysis of variance
IAPS	International Affective Picture System
LPP	Late positive potential
GA	Genetic algorithm
EP	Evoked potential
LOOCV	Leave-one-out cross-validation
FD	Fractal dimension
KFD	Katz's fractal dimension
HFD	Higuchi's fractal dimension
PNN	Probabilistic neural network

Table of contents

List of figures

List of tables

1 Introduction

Depression is characterized by sadness, loss of interest or pleasure, feeling of guilt or low self-worth, disturbed sleep or appetite, poor concentration and many other symptoms. According to WHO (World Health Organization), the total number of people living with depression in the world is approximately 322 million. South-East Asia Region and Western Pacific Region are the regions where around half of these people live. The number of people who are estimated to live in a depression increased by 18.4% between 2005 and 2015, which on the one hand is caused by the overall growth of the population. On the other hand, there has been an increase of the age groups in which the depression is prevalent. Depression is ranked by WHO as the single largest contributor to global disability (7.5%) which can in extreme cases lead to suicide. [1]

Depression as a mental disorder is always accompanied by abnormal brain activity and obvious emotional alternation. Therefore, as a method tracking the brain function, EEG may be suitable for detecting these abnormal activities.

With further advancements in electroencephalography (EEG), scientists have used it as a tool to investigate the brain. Today the clinical methods used to detect depression are very subjective. As there is a lack of characteristics that could be used to objectively measure and evaluate normal biological processes using questionnaires and similar subjective methods, different biomarkers are of interest with the objective to predict, help with the diagnosis and treat specific disorders in the future. [2]

Primary goal of present thesis is to build a machine learning model that would be capable of distinguishing depression between healthy control group and group of patients who have been diagnosed with depression. To achieve that goal, an analysis of visually evoked potentials and its' subcomponents will be conducted. Components will be measured during the EEG session, features will be extracted. Feature selection is performed and eventually machine learning model will be trained using top-performing features.

1.1 Problem statement

Depression is one of the diseases belonging to the global burden of diseases. Worldwide, approximately 322 million persons are living with depression [1]. Currently, depression is diagnosed with methods like interviews, physical examination, lab tests and structured symptom severity scales. These methods are highly subjective as they assume that the patient makes an accurate assessment about oneself [3]. Additionally, the doctors' assessments and decisions are subjective. For that reason, there is an increasing need for technology that would automatically detect and could be used to early detect depression and anxiety disorders. One common application of EEG in medicine is the diagnosis of pathologies like epilepsy. There is an evidence that EEG could be used to predict depression with machine learning approach [3]. Using such digital technology would simplify alerting the patients on time. For example, different brain-computer interface (BCI) devices are already available for commercial use. The EEG being easily accessible in the future enables possibilities of self-care and early detection of depression before further discussion with doctor. To accomplish that, smart machine learning models should be trained.

There are several methods used in research to build machine learning models for classification of depression. Classification of depression directly using frequency band power or measures of complexity are often used in research. Regarding visually evoked potentials, there are for example studies focusing on the detection of Attention deficit hyperactivity disorder (ADHD) or Bipolar mood disorder (BMD) [4]. Present study is focused specifically on classification of depression using visually evoked potential.

The objective of current study is to build a machine learning model that would distinguish depression between subjects with depression and healthy control group. Here, the subcomponents of extracted visually evoked potentials extracted from the EEG will be used as features. Significance of feature will be analysed, selection is carried out and top-performing features will be used for training the model.

1.2 Related work

In studies classifying between subjects with depression and healthy control subjects, there are mainly 5 different types of biomarkers and methods used for depression discrimination from EEG data. Most widely used are band power (spectral analysis), alpha asymmetry, functional connectivity, evoked potentials, and other time-domain features. These biomarkers are contributing up to 96% of all biomarkers found in research. [2]

As the principal EEG frequency bands are well known, it is possible to analyse the power amplitudes for each band, namely delta, theta, alpha, beta and gamma band. Each band is found to be related to some mechanisms in the brain. Using measurements from left hemisphere, the power of alpha band (8-12 Hz) and theta band (4-8 Hz) have been found to be good features to discriminate depressed patients from control group as these are responsible for emotional processing [2]. Čukić et al. [3] found that alpha frequency has indicated that those with high suicidal ideation experienced increased whole night absolute alpha activity compared to participants with low suicidal ideation. In beta frequency band (13-30 Hz) they it was found that suicide group was related to absolute beta power. No effect was observed between suicidal ideation and delta power. However, findings in another study are supportive of previous studies showing activity of alpha as well as delta (0.5-4 Hz) to distinguish depressed and healthy samples. In this study, the most accurate results were obtained by decision tree model for both alpha and delta band, that was widely distributed across regions in both hemispheres [5]. Gamma band (30-100 Hz) is less well characterized with respect to depression. In 2018, study reported that gamma band may be related to mood swings [6].

A biomarker called alpha asymmetry measures relative alpha band activity between the brain hemispheres. Usually, the alpha asymmetry measure is used at frontal electrodes. A study suggests that depressed patients and normal subjects differ in alpha band more significantly than other bands like delta, theta, and beta [7]. Moreover, it is concluded that the mean alpha power has significant difference in T3, F7, O1, P3, C3 in left hemisphere and O2 in right hemisphere [7]. Left alpha activity is found to be a good predictor for depression diagnosis. Mean alpha power were found to be higher in depressed patients compared to normal subjects [7]. However, it is not quite clear under which conditions it

holds true. A study demonstrated greater frontal alpha EEG asymmetry for non-depressed women but had no significant difference in alpha asymmetry between depressed and non-depressed men [8].

It has been demonstrated that the analysis of a single channel signal can also provide high accuracy for differentiating depression from EEG. The accuracy of linear, spectral asymmetry index (SASI) and nonlinear, detrended fluctuation analysis (DFA) were studied in combination. SASI, which reflects the relative difference between the higher and lower EEG frequency band power was found to have superior discrimination ability with classification accuracy of 76.5% over the DFA, which had classification accuracy of 70.6%. SASI is calculated by first finding the boundary frequencies of interest, calculating the powers of the lower and higher frequency bands, and finally finding the relative difference between the powers. DFA is calculated in time-domain giving the average fluctuations as a function of window length. The DFA values in channel were found to be significantly lower for depressive group compared to control group. Classification accuracy of 91.2% was obtained by the linear combination of SASI and DFA with the maximum difference in both measures occurring in the same channel, Pz. [9]

Instead of frequency-domain features, there have been studies concentrating on time-domain features which are the features extracted from pre-processed EEG time-series. Lempel-Ziv Complexity (LZC) and its' variation MLZC have been previously shown promising results in classification of neuronal or mental disorders on electroencephalographic signals. Lempel-Ziv complexity, which is well-known for the usage in zip-compression algorithm, has been used for detection of Parkinson disease, schizophrenia, Alzheimer disease, Major depressive disorder (MDD) etc. During the calculation of LZC, each signal segment is converted into binary sequence depending on some threshold value. Resulting sequence is scanned from left to right to investigate the patterns occurring. LZ is the total number of distinct structures or patterns found. During calculation of MLZC, multiple threshold values are used to obtain multiple binary sequences. Differences between depressive patients' and healthy control subjects' traditional LZC and MLZC values were tested using statistical analysis and highest classification accuracy (86.36%) was reached in the channel F3 by MLCZ. [10]

Best results regarding time-domain features have been achieved using fractality analysis [11]. That study concluded that using methods based on nonlinear features might give better results compared using only linear methods. For pre-processing, wavelet analysis was done. Wavelet analysis was more effective compared to common spectral decomposition method like Fourier transform. Next, fractal dimension (FD) was found from the EEG. Fractal dimension is a nonlinear measure of irregularity of the time series. The irregularity in the time series is considered to be an average variation of distances of successive points. Katz's (KFD) and Higuchi's (HFD) fractal dimensions are two different measures of complexity used in research. According to analysis of variance, the most discriminative FDs were found and used for classification with probabilistic neural network (PNN). Accuracy of 91.3% was achieved for classification based on HFDs of frontal brain beta sub-band. [11]

Another biomarker that is often investigated are network-based features. MDD is hypothesized to arise from dysfunction in the brain networks [12]. In one study, the power spectral density of the artifact-free EEG data in four frequency bands were calculated for each nearest neighbour. Each pair of nearest neighbour electrodes represented an edge of a brain network. Coherence was calculated between pair of nodes that represented a normalized measure between the signal at the nodes at given frequency. Coherence was calculated as a function of the power spectral outputs for signal form separate nodes for each frequency. MDD subjects showed statistically significantly greater median coherence (overall median coherence across all edges) than control group in each frequency band. Nearest centroid classification analysis identified six edges in the alpha band that best characterized the depressed state. On average, classifier accurately distinguished 81% of MDD from healthy control subjects [12]. One recent research indicated that adding graph theoretical measures (level of segregation or integration of nodes) to functional connectivity does not significantly increase classification accuracy between depressive and healthy subjects [13].

An evoked potential is an electrical potential in a specific pattern recorded from the brain following presentation of a stimulus such as a light flash or a pure tone. When individual is presented some external stimuli, the brain waves fluctuate in some specific manner. These fluctuations are usually named after their direction and latency [2]. Visual evoked potential (VEP) is signal response to external visual stimuli. According to the International Society for Clinical Electrophysiology of Vision (ISCEV) standard for

VEPs, the visual evoked potential in response to flash stimulation consists of a series of negative and positive components. These components will be further discussed as the present thesis will mostly be focused on the classification using visually evoked potentials.

1.2.1 VEP components

The P100 or P1 is the first positive component with the positive voltage deflection that peaks approximately 90-160 ms following stimulus onset. A reduction in P1 mean amplitude has been reported primarily in schizophrenia, but also in some subtypes of depression. It is suggested that abnormalities in P1 component may be initiated by the withdrawal from the environment or misinterpretation of the visual stimuli. One study found that higher severity delusional thinking is associated with reduced P1 amplitude, regardless of the diagnosis [14]. Specifically, recent research investigated P100 component. Amplitude and latency were extracted as features and applied to 1-knn classifier. According to leave-one-out method, 92.85% classification accuracy was obtained to distinguish the BMD, ADHD, and healthy subjects from each other [4].

In slightly different study, it was tested whether event-related brain potentials (ERPs) change when change in sound intensity is introduced and could it differentiate depressed subjects and healthy control group. Auditory response is supposed to be increasing when the intensity of an auditory stimulus increases. This reactivity was seen when measuring the auditory evoked responses (AEPs) such as N1. The N1 component in this case was an automatic response elicited in the auditory cortex at approximately 100 ms after the stimulus onset. Visual observation of the averaged waveforms indicated that N1 responses were elicited. Mean response amplitude values for N1 were calculated for latency 90-140 ms after the stimulus which showed that depression disorder subjects had significantly larger amplitudes in the N1 response compared to control group. Generally, it is found that the visual stimuli situated in visual fields with focused attention elicited components P1, N1 with larger amplitude than ignored or unnoticed stimuli. This amplitude enhancement is sensitive to the specific localization of the stimuli in the visual field. [15]

The N200 component or N2 is usually evoked around 180 to 325 ms following the specific visual stimulus. N2 may be evoked by exposing continuously two types of stimuli

to the subject, one presented regularly and other displayed rarely, randomly. When rare, random stimuli is presented, the N200 is observed. Goal of one study was to observe whether changes related to N2 exist in depression. Depressed and non-depressed participants completed a SSRT task. Analysis of variance (ANOVA) was conducted for the N200. No significant relationship was found between amplitude differences. That suggested there might not be associations between N200 and depressive symptoms. [16]

The Late Positive Potential (LPP) is an event-related potential (ERP) that indexes selective attention toward emotional stimuli. In adults, LPP arises around 200 ms after stimulus and continues if the stimulus is present. It suggests that LPP represents ongoing emotional processing which makes it well-suited for investigating psychiatric disorders. Previous research has conducted that emotional faces are good visual stimuli for evoking change in EEG signal. Moreover, it is well known that individuals with depression disorder have different reaction to some emotional stimuli compared to healthy subjects. Latest LPP research used stimuli from the International Affective Pictures System (IAPS) that compromises photos of emotional scenes. LPP mean activity was measured within a time frame of 400 to 1500 ms. According to scalp distribution of positive activity in this interval, a set of electrodes (Pz, P4, CP1, CP2, Cz, C4) were chosen for analysis. ANOVA on the mean LPP activity showed a significant effect for emotion. Activation was highest for fearful faces and lowest for happy faces. Depressed group presented lower LPP mean activity than the healthy control group. [17]

The P300 or P3 is a positive EEG signal voltage peak at 300 ms after the onset of a stimulus. As usual the P300 is computed after performing averaging of various stimuli. P300 component is often used for classification of mental activities. Moreover, the P300 amplitudes are considered as an indicator of symptom improvements in depressed condition after treatment with antidepressants. During one research, the ERP data were recorder during performance of visual oddball task. The features of P300 such as amplitudes and latencies were computed at central channels Fz, Cz, Pz. Features were ranked by t-test, wilcoxon and roc. Top features were selected. Logistic regression (LR) classifier was utilized. P300 amplitudes were observed significantly higher in the healthy control group as compared to MDD patients. In addition, larger P300 latencies were found in the MDD patients. The highest classification accuracy of 90.5% was achieved. [18]

2 Implementation

The implementation of machine learning model consists of classical machine learning pipeline as well as some EEG data related operations. The overview of the steps carried out in the thesis are specified in the following list:

1. EEG data pre-processing
    a. Data description
    b. Pre-processing configuration
    c. Filtering
    d. Artifact removal
    e. Re-referencing
2. VEP extraction and analysis
    a. Baseline correction
    b. Finding epochs
    c. Averaging
    d. Identification of initial components
3. Feature extraction
    a. Characteristics of different visual evoked potential's components
    b. Method for extracting components from subjects' VEPs
4. Feature selection
    a. T-test, wilcoxon-test statistical analysis / ordering features
    b. Selection with genetic algorithm (GA)
5. Classification
    a. Logistic regression, SVM, Adaboost classification model training.

First, EEG data will be pre-processed according to the defined configuration. The data will be filtered into suitable frequency range. Then artifact removal will be done. Both manual as well as automatic artifact removal will be considered. As the last step of pre-processing, the re-referencing will be carried out. The processed EEG will be further used for VEP extraction and analysis starting with baseline correction and finding epochs form the EEG using the known timestamps of presented image stimuli. Time-locked epochs will be used to carry out averaging. Grand average VEP over all the subjects will be found and used to visually identify initial components of visually evoked potential. A specific configuration will be created based on the identified components. Each component will have a set of characteristics describing the component. Next, characteristics of the identified VEPs will be defined and these features will be extracted from each subjects' VEPs accordingly. Statistical analysis will be used to obtain significant features. A subset of best features will be selected and classification will be done for these individual features. Next, genetic algorithm will be used to select subset of features and build multivariate machine learning model using either logistic regression, SVM and Adaboost algorithm.

3 Data pre-processing

The first phase of EEG data ERP analysis is pre-processing. As a first step of pre-processing the raw EEG signal is filtered to specific frequency range. Next, artifacts are removed. That is done both manually and by using methods like ICA. Finally, re-referencing is performed. Parameters for these operations will be easily configurable.

Regarding pre-processing, different studies have conducted that there is a trade-off in every processing step. Artifact removal process itself introduces new artifacts into the signal. As ICA or similar methods are used, the dynamics of the signal can be changed. In general, to retain maximum amount of physiological signal information, it is advisable to do minimal pre-processing [19]. For that reason, present study applied different pre-processing methods, creates set of post-processed versions of the signals and eventually measures the quality of the data. Namely, during one version, artifacts were removed manually while the other version had artifacts removed by ICA. Next, in one version re-referencing was excluded while other version had re-referencing done with common-average method.

3.1 Data description

The data acquisition was not in the scope of current thesis. Data was previously collected in Tallinn University of Technology from 44 subjects, 22 subjects previously diagnosed with major depressive disorder (MDD) and control group of size 22 consisting healthy subjects. One EEG session lasted approximately 30 minutes. During that time subject was presented with set of colourful images with the objective to evoke response in EEG signal. Each image was on the screen for 3 seconds and with changing duration of pause between the images. For every subject, approximately 219 different images were presented during the session. Each image was indexed with specific number and was taken from the International Affective Pictures System (IAPS). For assessing the images over 3 dimensions (valence, arousal, dominance), a study was carried out to obtain standard

ratings for each of the images [20]. Ratings are scored such that 9 represents a high rating on each dimension and 1 represents a low rating on each dimension. In addition to standard ratings, the dataset also provides personal ratings of subjects that were collected after the session based on the same rating system used for standard ratings. In present thesis different ratings of the images were not considered. It can be assumed that images lower in valence and higher in arousal would have different response for healthy subject compared the one with depression. However, performing the analysis on a smaller subset of images have trade-offs like inability to average out enough noise.

3.2 Pre-processing configuration

Some of the parameters used during the pre-processing phase (like filtering frequency bands) of raw EEG signals are global to all of the subjects while some of the parameters (like bad artifact annotations) are specific to the subject. For that reason, a common configuration file was created that would capture all configurable parameters into one central file. An example of the configuration file is depicted on Figure 3.1.

```
{
  "verbose": false,
  "exclude_channels": [],
  "filtering": {
    "min_frequency": 0.01,
    "max_frequency": 30.0
  },
  "ica": {
    "random_state": 2021,
    "num_components": 12,
    "exclude_components": [0, 1, 2]
  },
  "epoch": {
    "time_before_event": -0.2,
    "time_after_event": 1.0
                                        },
  "data": [
    {
      "id": "MH004",
      "exclude": false,
      "label": "H",
      "annotations": [
        { "start": 270, "duration": 50, "description": "bad_eyeBlink" },
        { "start": 840, "duration": 50, "description": "bad_eyeBlink" }
      ]
    },
    ...
}
```

Figure 3.1. Example of pre-processing configuration file.

3.3 Filtering

To reduce the noise of raw EEG signal, the filtering was performed. Most relevant portion of the ERP waveform in a typical experiment consists of frequencies between 0.01 Hz and 30 Hz. Voltage changes in this frequency band is believed to be of neural origin. All non-neural voltage changes like contraction of the muscles (above 100 Hz) and slow voltage shifts (< 0.01 Hz) must be filtered out [21]. A finite impulse response (FIR) band-bass filter with cut-off frequencies of 0.01 Hz and 30 Hz were used to attenuate other frequencies out of that range. No notch filter was needed, because dataset already had line noise filtered out (~50 Hz, ~60 Hz). The power spectral density plot of all EEG channels after filtering is depicted on Figure 3.2.

The power spectral density (PSD) shows how the power is distributed over the frequency components present in the signal. As can be seen on Figure 3.2, the filtering process was effective, as the frequencies that are on the right side (greater) of the cut-off frequency have significantly less power compared to the ones on the left side of the cut-off while the left side has frequencies untouched.

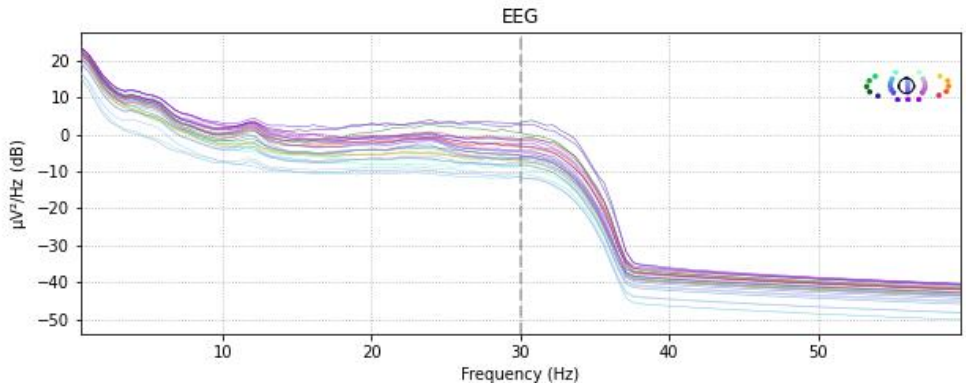

Figure 3.2. Power spectral density after filtering.

3.4 Artifact removal

The artifact removal was done both manually and automatically. For manual removal, each subject was annotated with "bad" artifacts in the configuration file. An annotation named "bad_eyeBlink" was assigned to regions where eye blink could be visually detected from the EEG signal. Eye blinks can be visually detected by looking for rapid voltage changes in most of the channels. In addition, a general annotation named "bad_artifact" was assigned to other very noisy segments found visually from the signal. Eventually ICA (automatic removal) was used instead of manually removing annotated regions. EEG data of 4 subjects were excluded completely due to extensive noise in the data. In the worst case ~ 75% of the data of a subject was corrupted. Further removal of segments based on peak-to-peak amplitude values will be done in next phase when analysing VEPs.

3.4.1 ICA – Independent Component Analysis

The ICA solves the blind source separation problem to recover independent source signals after they hare linearly mixed by an unknown matrix. Nothing is known about the sources except that there are N different sources. ICA creates statistically independent components, meaning that the joint probability of the components equals to the product of their probability. It is usually done by either finding independent sources that maximize the joint entropy of the rotated components (infomax) on non-gaussianity of the rotated components (fastICA). [22]

Each resulting component of ICA is a linear combination of channel signals and are thus often thought as a source signal from the brain capturing one specific aspect. Usually, ICA outputs the same number of components as number of sources. However, some implementations perform principal component analysis (PCA) as preliminary step, extract n components explaining most of the variance and then use these components form further ICA. Thus, 12 components were created in present thesis. The ocular component was found to be the first one by visual inspection. The ICA000 on Figure 3.3 represents eye movements because of its frontal location. Moreover, the corresponding time-graph shows 4 distinct changes in the voltage which is matching with the 4 changes

occurring in EEG data. On Figure 3.3 are the selected ocular component and first 10 seconds of that component. EEG data before and after component exclusion is also depicted.

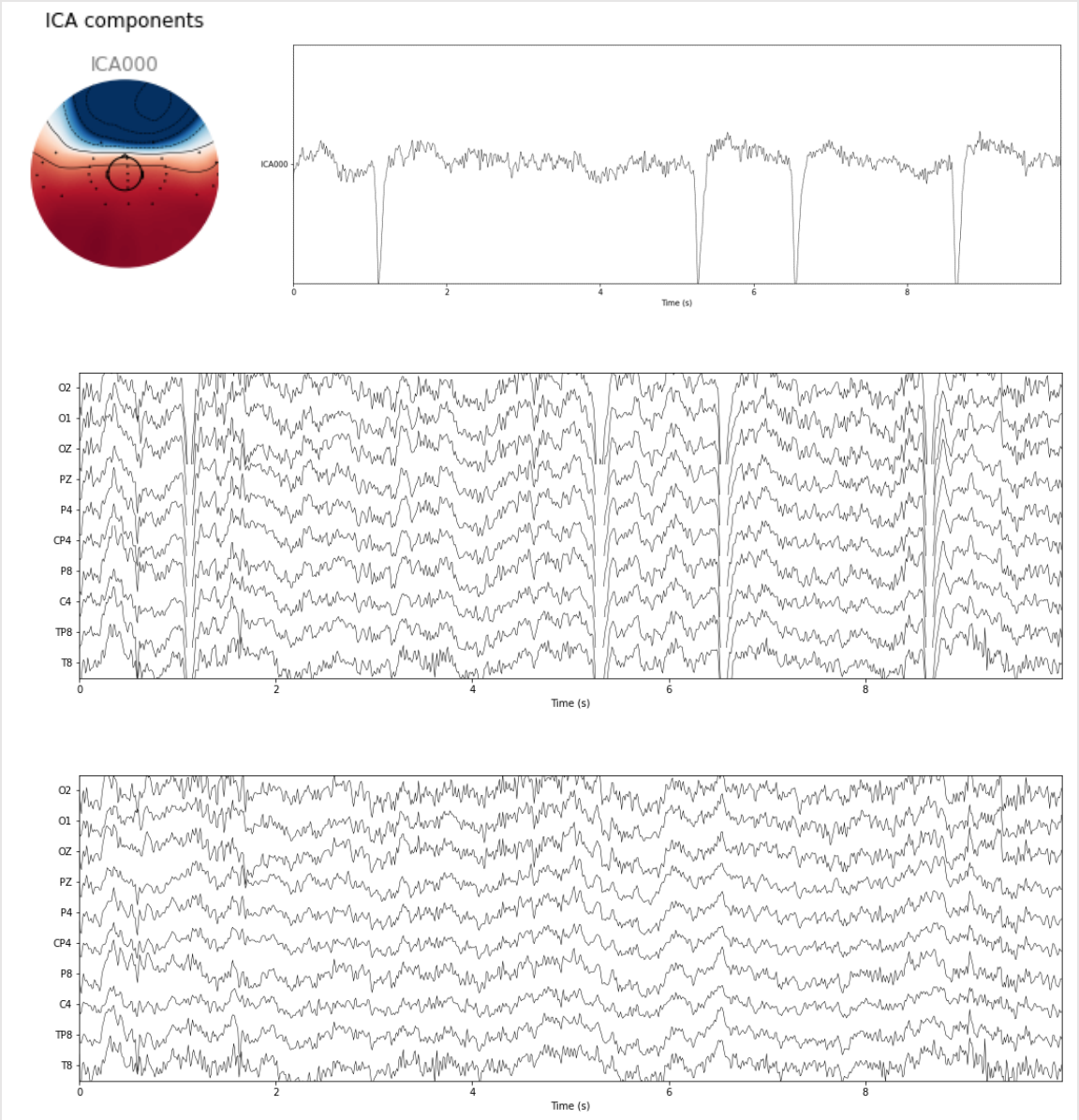

Figure 3.3. Elimination of ocular artifact using ICA.

3.5 Re-referencing

In EEG, voltage in each electrode is measured relative to the reference electrode. The reference needs to be carefully selected because any activity in the reference channel will evoke activity at other channels. Dedicated electrode is usually used as an online reference. Changing the reference offline after recording is called re-referencing. The re-referencing is done to express EEG signals with respect to new reference. That eliminates the online reference as well as noise that was caused by the usage of it. In present study, common-average re-referencing method was used. During that method, a common value of all channels will be subtracted from each channel. Thus, the total sum of channels' values will sum up to 0.

4 VEP extraction and analysis

The objective of VEP extraction and analysis phase is to extract visual evoked potentials form the pre-processed continuous EEG signal and analyse the outcome. The shape of the VEP might differ depending on the stimulus. However, there are always fluctuations and negative and positive peaks present in the signal. Typical VEP for standard flash stimulation (according to ISCEV standard for clinical visual evoked potentials) is depicted on Figure 4.1.

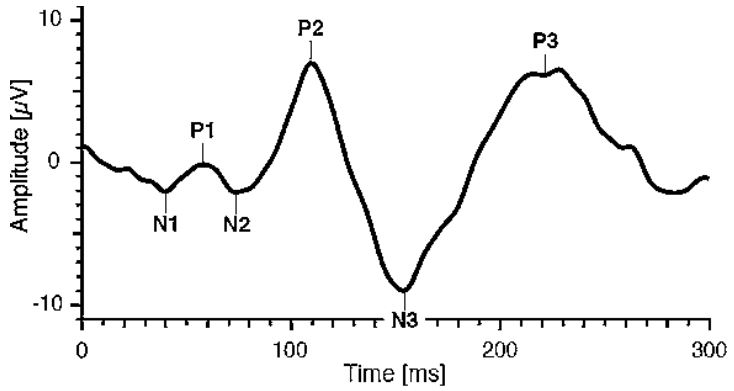

Figure 4.1. Typical VEP for standard flash stimulation (ISCEV).

The exact steps needed to extract VEPs from data are listed below.

- Baseline correction + generating epochs + epoch rejection.
- Averaging
- Component identification

During one EEG session approximately 200 events are presented to the subjects. By knowing the exact timing of these events, it is possible to extract the signal response caused by these events from the continuous EEG signal. For that, epochs are found based

on the timing of events presented to the subjects. Epochs are set of time-windows that are extracted from the continuous EEG signal. Some of these responses can however be outliers in terms of too large absolute amplitude values. For that reason, a rejection-criteria is used to exclude all epochs that doesn't fulfil the criteria. As common in literature, maximum peak-to-peak amplitude value thresholding is used as a rejection method. In present thesis, peak-to-peak amplitude value thresholding was used (with rejection threshold of 100e-6).

4.1 Epoch generation

After the external event occurs, the response in EEG signal appears after some time interval. It is important that the produced response of EEG signal would be as invariant across the trials as possible. Extracted epochs might differ significantly due to drifts in EEG signal caused be external factors. To make the epochs more comparable, a baseline correction is done. Namely, a time slice of EEG activity before the stimulus is used to correct the post-stimulus time interval. The mean of the baseline segment will be subtracted from the baseline itself as well as from the post-stimulus segment. That way, all the epochs can further be compared on the same basis. For baseline segment usually 0.2 s is used as pre-stimulus time-window. For post-stimulus response time-window, 0.6-1.0 s is used. That period is enough to cover all necessary VEP components (e.g. P1, N1, P2, N2, P3). Same parameters were used for baseline correction in present thesis.

4.2 Averaging

All the epochs capture the evoked potential (EP) caused by stimuli and some noise from other brain activity. As the EP is assumed to be stationary (latency and morphology are invariant) over multiple trials and subjects while the noise is assumed to be with zero mean, the repetition and averaging will eliminate the noise while retaining the signal of interest. There are ~200 epochs corresponding to one session. As there are 40 subjects (4 were excluded previously), the total number of epochs over trials and subjects is approximately 8000. Considering all epochs, the VEP is a result of finding an average signal of those epochs. Such VEP is also called grand average visual evoked potential. Resulting evoked potentials in each of the channels are depicted on Figure 4.2.

The absolute amplitude values of the channels have been significantly reduced after the grand averaging. This is caused by different values of latencies and amplitudes of the peaks and thus the fluctuation will get wider and lose their power. The stimuli presented at time 0.0 has caused fluctuations in the signal already after 50 ms. Both, the pre-stimulus baseline segment as well as ending of the evoked potential are converging to zero voltage. This is expected as before the stimulus there was just other brain activity and after the stimulus the signal should converge back to the point where there is no impact of the stimuli and just other brain activity.

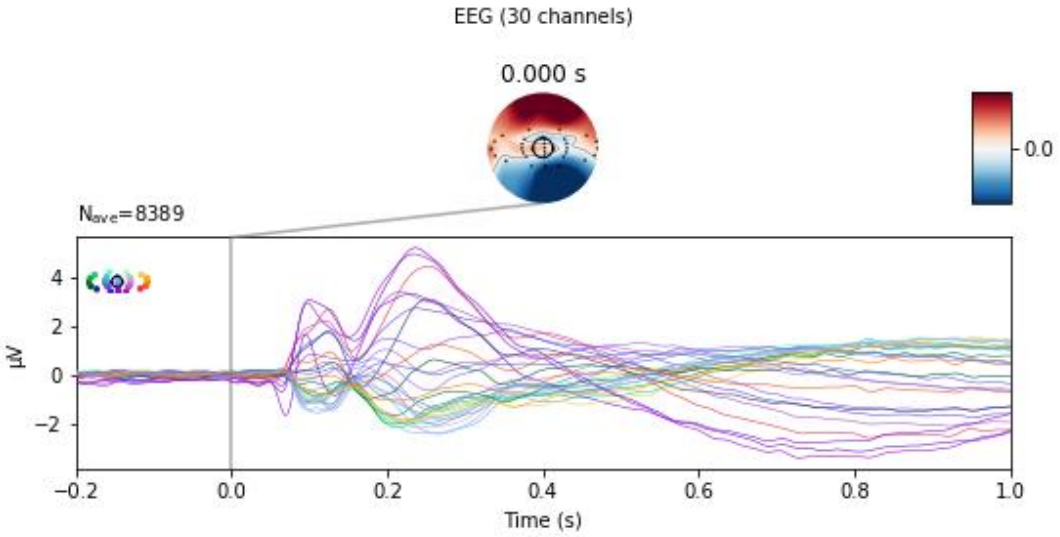

Figure 4.2. Grand average visual evoked potential in all channels.

4.3 Component identification

There are groups of channels behaving similarly with respect to the stimuli. Looking at the heatmap on Figure 4.3, 10 channels (O2, O1, OZ, PZ, P4, CP4, P8, TP8, P7, P3) have stronger positive fluctuation around 200 ms. This group contains mainly channels from the back of the head and are thus reflecting some common response from that area of the brain.

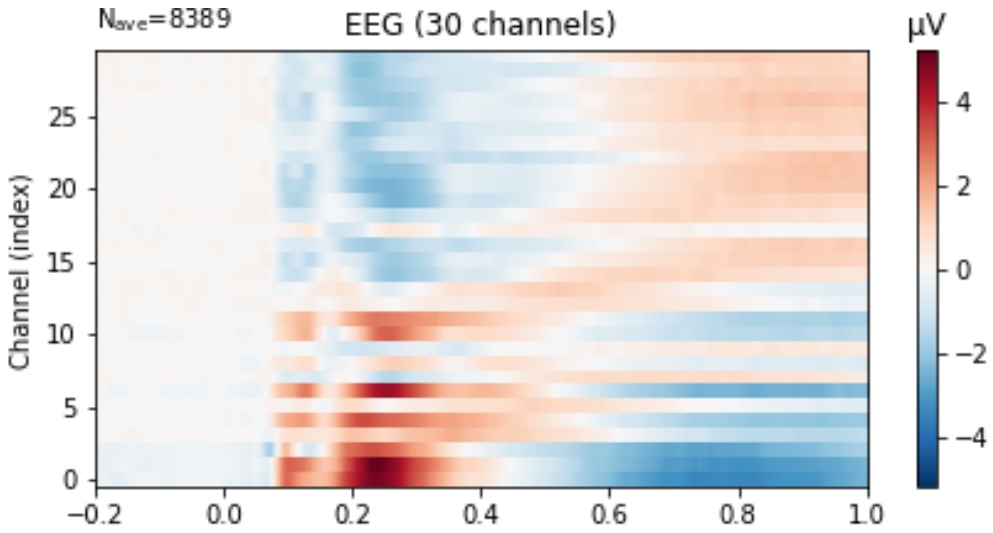

Figure 4.3. Evoked potentials in each channel.

On Figure 4.4, a group of back channels is depicted. The first negative peak around 65 ms after the stimuli is estimated to be component C1. C1 is highly sensitive to any stimulus parameter, such as contrast. As around 100 ms the channels have first positive fluctuation, the component resolving there is estimated to be P1. Similarly to C1, the P1 is also sensitive to the visual stimulus like the contrast of an image. As the peaks of P1 are not as aligned to each other as C1, a segment around 100 ms will be defined to capture all these peaks. Most of the following negative peaks were around 160 ms and were estimated to be component N1. The most powerful positive fluctuation is considered to be component P2. No further components could be identified visually. However, the mean

amplitude of additional components will used. For example, the segment at 275-400 ms and the Late Positive Potential (LPP) is of interest.

Another group of channels does not have such strong fluctuations. However, they are behaving similarly. Looking at the heatmap on Figure 4.5, 8 channels (FT8, FCZ, FZ, F4, F8, T7, FT7, FC3) have similar shapes. This group contains channels mainly from the frontal part of the head. The peaks of these channels are aligned with the previous group of channels but are reversed.

The grand average VEP was mainly created for analysis purpose to identify the VEP components. The grand average VEP was created using filtering, performing ICA, without manual artifact removal, performing averaging and epoch rejection. However, it is not clear to what extent each of these steps lose important information about the signal. Thus, multiple versions of VEPs using different configurations will be created for each subject and eventually the best configuration is selected.

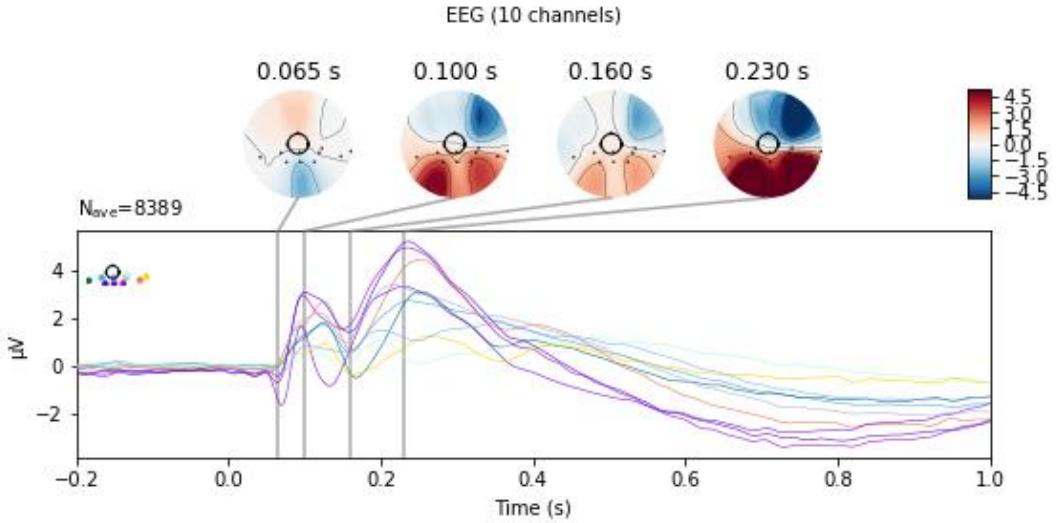

Figure 4.4. VEPs of back channels.

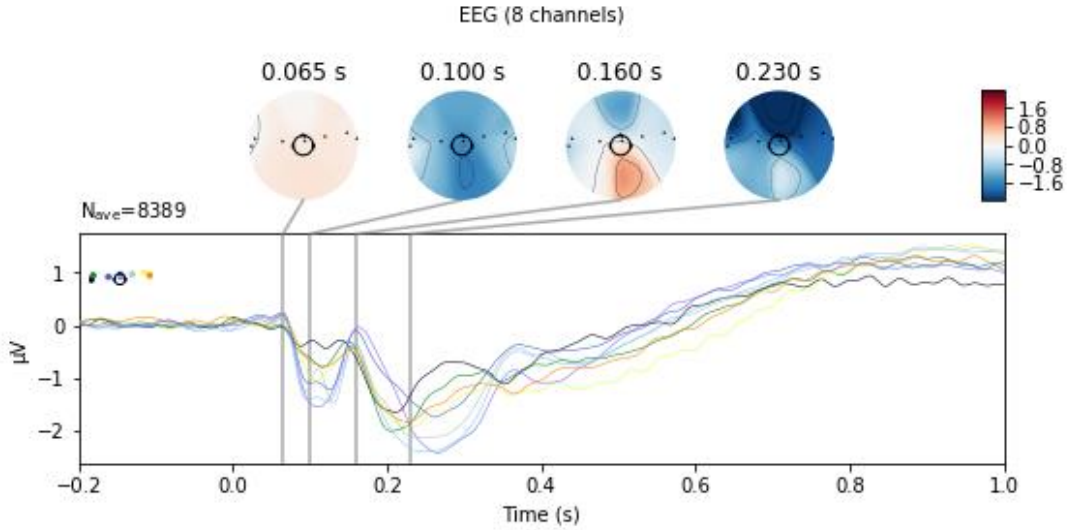

Figure 4.5 VEPs of frontal channels.

5 Features

The objective of feature extraction and analysis is to consider the visual evoked potentials of each subject and extract characteristics that would describe that VEP. During the VEP analysis phase a set of VEP components (C1, P1, N1, P2, N2, LPP) was identified. For each of these components, an amplitude of the peak, latency of the peak and the mean amplitude will be founded. The outcome of feature extraction and analysis is a dataset that contains values of the features for each subject as well as labelling for each subject.

5.1 Feature extraction

A specified configuration was used to extract features from VEP-s. The configuration is depicted on Figure 5.1. Each channel of grand average VEP was analysed. Two groups of channels were created. First group of channels share the property that the first fluctuation has positive peak around 100 ms (P1). The second group of channels have in contrary, negative peak around 100 ms (N1). Such division of channels was done, because in order to find the latency, the type of the peak must be known. For each group, a set of components are defined. The id refers to the name of the VEP component. The centre refers to the estimated time the component is present. The left and right parameters define an area around the centre that is still considered to be a part of the component. For example, for the component C1 on Figure 5.1, the length of the segment under consideration is 60 ms in total. Next, the colour parameter is used to visually differentiate between components. Finally, features parameter contains the set of calculations that will be performed with that segment. Min refers to finding minimum peak amplitude and latency, max refers to finding maximum peak amplitude and latency within specified segment. Avg refers to finding average amplitude value of the segment.

Example of the configuration applied to the VEP belonging to the first group is depicted on Figure 5.2. Example of configuration applied to the VEP from the second group is depicted on Figure 5.3. The lines of different colour represent segments corresponding to

each component in configuration. For example, the N1 component is centred around 100 with left radius of 20 ms and right radius of 30 ms.

```
{
  "group-1": {
    "channels": ["O2", "O1", "OZ", "PZ", "P4", …],
    "components": [
      {
        "id": "C1",
        "centre": "65",
        "left": "30",
        "right": "30",
        "colour": "C1",
        "features": ["min", "max", "avg"]
      }, …
    ]
  },
  "group-2": {
    "channels": ["C4", "T8", "CZ", "FC4", "FT8", …],
    "components": [
      {
        "id": "N1",
        "centre": "100",
        "left": "20",
        "right": "30",
        "colour": "C2",
        "features": ["min", "avg"]
      }, …
    ]
  }
}
```

Figure 5.1. Configuration used for extracting VEPs.

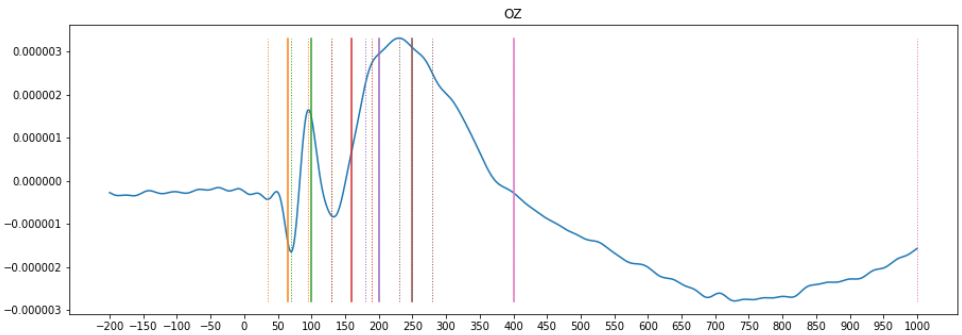

Figure 5.2. Example EEG VEP from channel (OZ), group 1.

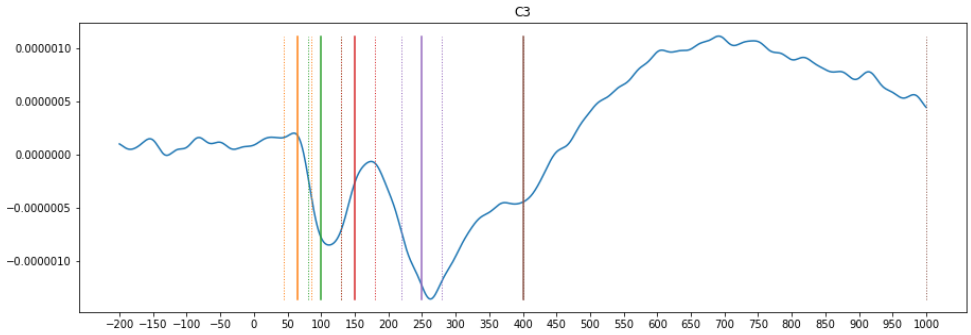

Figure 5.3. Example EEG VEP from channel (C3), group 2.

5.2 Statistical analysis

By applying the configuration to each subject, in total 429 features are found. All features are labelled with channel, component, and feature type codes (e.g. O1-N2-min-l). Each subject is marked as D (depression) or H (healthy). Based on the labelling each feature can be divided into two sub-groups and the mean of the groups can be compared by carrying out significance test. If the data is normally distributed, the difference between two groups of data can be found as a difference between the mean value as most of the occurrences are around the mean. For that t-test is usually used. On the other hand, if the data does not follow normal distribution, the mean value can not be used as a measure of discrimination. In this case nonparametric test like Wilcoxon rank-sum test can be used.

The features are then sorted by the p-value of the features. However, as 429 tests are performed, the obtained p-values must be corrected to reflect the real significance of the features. Namely, the probability of type I error increases by the number of tests performed. When the p-value is selected as 0.05, the probability of type I error after 429 tests is significantly high. For that reason, a correction of a used p-value must be done. The Bonferroni correction will be used for that purpose.

5.2.1 Bonferroni correction

The Bonferroni correction decreases the significance level by dividing the p-value by the number of tests performed. That way the p-value is small enough such that even after performing 429 tests, the probability of a false positive is less than 0.05. Thus, the new p-value used for analysis is approximately 0.0002.

5.2.2 Feature selection

Top 15 features are ordered in Table 5.1 by the p-value obtained by using the Wilcoxon rank-sum test. The best feature was evaluated to be latency of the minimum peak of N2 component in PZ channel. It can be conducted that none of the features are actually significant as the p-values of features are greater than the modified benchmark after Bonferroni correction (0.0002). Still, these top features are selected and further used in classification phase. Individual accuracy for each feature will be found as well as the combination of different features will be tested.

Table 5.1. P-values of features ordered by wilcoxon-test.

Feature	p-value (wilcoxon-test)	p-value (t-test)
PZ-N2-min-l	0.0014	0.0092
FT7-N250-min-l	0.0035	0.0016
CPZ-C1-mean-a	0.0045	0.0044
O1-P2-max-a	0.0064	0.0206
C4-N250-min-l	0.0074	0.0030
FC4-LPP-mean-a	0.0080	0.0065
FT8-N250-min-a	0.0080	0.0056
FC4-N250-min-a	0.0090	0.0246
O2-P2-max-a	0.0100	0.0389
FT8-P150-mean-a	0.0124	0.0256
O2-P2-mean-a	0.0124	0.0283
O1-N2-min-l	0.0130	0.0135
P3-N2-mean-a	0.0137	0.0402
F4-N250-min-a	0.0152	0.1475
F7-C1-max-l	0.0165	0.0963

6 Classification

A subset of best features was determined from the feature extraction and selection phase. The objective of a research was to construct a model for differentiating between subjects with depression and the control group. Subjects have been labelled correspondingly (D and H). As the number of labels in the dataset is 2, the classifier under consideration will be binary classification. A set of different classifiers will be trained for both – individual features as well as combination of features. Used binary classification machine learning algorithms are logistic regression, support vector machine (SVM) and AdaBoost.

The dataset contains multiple features that are significantly different in scale. For example, the latency values compared to the amplitude and mean amplitude values differ in more than 10e6 times. Thus, standardization of all features was done. The mean value of each feature was subtracted from the feature value and finally divided by the standard deviation of the feature values. That way all the values are close to 0 and with unit variance. Additionally, labels (D, H) were binarized (1, 0).

6.1 Genetic Algorithm

In current research genetic algorithm (GA) is used for feature selection and model construction. Genetic algorithm is a stochastic optimization method based on the processes present in biology. The genetic algorithm has a concept called population. Population is made of set of individuals. In the context of features, individual represent one possible selection of features. Individual is represented as binary vector of size N, where N is the total number of different features. Inclusion of a feature is represented with 1, while excluded features are represented with 0. During each iteration of genetic algorithm, the fitness of each individual is assessed by using fitness function. In the context of classification problem, the fitness function is the negative accuracy of the model trained with selected features. Minimization of fitness function is used. Next, a subset of individuals is selected as a parent for new populations. The better the fitness,

the greater is the likelihood of being selected as a parent. The parents are used to create new population. New individuals are created using the concepts of crossover and mutation. During crossover new individual is created by using parts from parent vectors. The crossover is probabilistic, meaning that the parents might not produce child at all. Moreover, mutations can occur, which mean random probabilistic changes in child vectors. The invariant of the genetic algorithm is that the average fitness of the population should be increasing each step. After iterations, the fittest individual is selected as the solution for the optimization problem.

6.2 LOOCV

Leave-one-out cross-validation (LOOCV) is a cross-validation algorithm often used in EEG research for model validation. LOOCV is a special case of the leave-p-out cross-validation with p=1. During each step of leave-one-out cross-validation, one observation is left out from the dataset for testing. Namely, for dataset of size N, the size of training set is N-1, while the size of testing set is always 1. A total of N models will be trained and tested to ensure that each sample is used as the test sample once. Finally, the average accuracy of the models will be computed.

6.3 K-fold cross-validation

K-fold cross validation is a cross validation algorithm method during which the dataset is partitioned into k subsets. During each step of the algorithm, one partition is used for testing while other dataset is used for training. The size of training set is (k-1)/k * N and the size of the testing set is 1/k * N. K models will be trained such that all the data will be used both for training and testing. Finally, the average accuracy of the models will be computed.

6.4 Adaboost

Boosting refers to a machine learning ensemble algorithm where models are added sequentially and every added model in the sequence corrects the predictions made by earlier models. Adaptive Boosting or Adaboost is one implementation of boosting algorithms. The main idea of Adaboost is to combine set of "weak" classifiers (accuracy close to 0.5) and output a boosted classifier. Boosted classifier is weighted sum of the weak classifiers. Usually, decision tree classifiers are used as base "weak" classifiers.

6.5 Individual models

Logistic regression model was trained on each individual feature. The accuracy and f-score were calculated for each feature. The top 20 features sorted based on the avg. accuracy are listed in Table 6.1. Features that had good average accuracy (marked with *) were added to the existing ones obtained in the statistical analysis phase.

There might be a trade-off of between using only one cross-validation method and using multiple methods. As the size of the dataset is relatively small (n=40), methods like LOOCV are often recommended compared to other methods as the testing set is always of size 1 and the size of the training set is thus maximized and allows to train the model with almost all the data. However, because of the exact same reason the LOOCV is overfitting due to the smallness of the testing set. K-fold has larger testing set compared to the LOOCV. The trade-off however is that the training set is consequently smaller. Thus, the obtained accuracy is usually smaller compared to the model validated with LOOCV. To get the best from both methods, another metric is introduced that is the average accuracy of both methods. The performance of classification models will be assessed based on that metric.

6.6 Multivariate models

20 base features listed in Table 6.1 were further used for building different kind of models, possibly multivariate models. First, the feature selection was done using genetic algorithm with fixed set of parameters (n_iterations=20, n_population=16, crossover_probability=0.9, mutation_probability=0.05). Average value of LOOCV and k-fold (k=8) was used as a fitness function. The genetic algorithm was run multiple times and the best feature sets were remembered. Three models were trained: logistic regression (LR), support vector machine (SVM) and boosting method (Adaboost). In table 6.2, best feature sets with accuracy are listed for different classification algorithms.

The best accuracy received for all the algorithms was 86.3%. However, Adaboost tends to use more features to receive same accuracy. Also, compared to the logistic regression, SVM best model used fewer features (7). Thus, SVM is considered superior over other trained methods.

Table 6.1. Performance of logistic regression models trained on individual features.

Feature	LOOCV	k-fold (k=8)	avg. accuracy
FT7-N250-min-l	0.725	0.700	0.713
*FP2-N250-min-l**	0.700	0.700	0.700
O1-N2-min-l	0.700	0.625	0.663
C4-N250-min-l	0.700	0.625	0.663
FT8-N250-min-a	0.650	0.600	0.625
CPZ-C1-mean-a	0.700	0.525	0.613
*FT8-N250-min-l**	0.650	0.575	0.613
*F3-N1-min-l**	0.650	0.550	0.600
*TP7-P1-max-a**	0.650	0.525	0.588
*CPZ-P1-mean-a**	0.650	0.525	0.588
*CPZ-N1-mean-a**	0.650	0.525	0.588
*CPZ-C1-min-l**	0.625	0.525	0.575
*F8-N250-min-l**	0.625	0.525	0.575
PZ-N2-min-l	0.600	0.550	0.575
*FT8-P150-max-a**	0.675	0.450	0.563
FC4-LPP-mean-a	0.675	0.500	0.588
O1-P2-max-a	0.600	0.525	0.563
O2-P2-mean-a	0.650	0.450	0.550
FC4-N250-min-a	0.625	0.475	0.550
*P8-C1-mean-a**	0.625	0.475	0.550

Table 6.2. Performance of multivariate models.

Method	Selected features	avg. accuracy
LR	[1, 4, 6, 9, 10, 11, 16, 18, 19]	**0.863**
LR	[1, 4, 5, 6, 9, 10, 11, 16, 18, 19]	0.850
LR	[10, 12, 14, 15, 16, 17, 18, 19, 20]	0.850
LR	[2, 3, 6, 9, 11, 14, 15, 16, 19]	0.850
LR	[2, 3, 5, 9, 11, 14, 16, 19]	0.838
SVM	[1, 2, 4, 7, 11, 12, 16]	**0.863**
SVM	[1, 4, 6, 10, 11, 14, 16]	0.862
SVM	[1, 4, 7, 9, 10, 11, 14, 16]	0.838
SVM	[1, 2, 4, 10, 11, 14, 15, 16]	0.838
SVM	[1, 2, 4, 7, 10, 11, 12, 13, 15, 16]	0.825
Adaboost	[2, 5, 6, 8, 10, 11, 14, 16, 17, 18, 19, 20]	**0.863**
Adaboost	[2, 3, 5, 6, 8, 9, 10, 11, 13, 14, 16, 17, 19]	0.850
Adaboost	[2, 3, 5, 6, 8, 10, 11, 13, 14, 16, 17, 19]	0.838
Adaboost	[1, 6, 11, 12, 13, 14, 15, 16, 17, 18]	0.837
Adaboost	[2, 3, 5, 6, 8, 10, 11, 13, 14, 16, 17, 18, 19, 20]	0.825

7 Results

Pre-processing of subjects' EEG signals was done considering additionally individual pre-processing. Pre-processing included filtering and artifact removal by either manual or automatic manner using ICA. Visual evoked potentials were extracted based on which sub-components were extracted and eventually set of features were generated. Features were analysed from the statistical analysis point of view as well as from the classification accuracy perspective based on which top-performing features were selected using genetic algorithm. Finally, individual models as well as multivariate models were trained for top-performing features using logistic regression, SVM and Adaboost classification algorithms. Performance of models were evaluated. As a best result, a multivariate SVM model with 7 features and with accuracy of 86.3% was obtained.

8 Conclusion

The aim of the thesis was to investigate a possibility to use visually evoked potentials for depression discrimination. Namely, the aim was to develop machine learning model that would correctly distinguish subjects diagnosed with depression from healthy control group. A dataset of 20 healthy and 20 subjects with depression was used.

Based on the obtained results it can be concluded that the primary goal of the thesis was achieved. As the dataset is relatively small, the main concern when training a classifier was overfitting. Cross-validation was used as a preventative measure against overfitting. To minimize the risk of overfitting several validation methods were used to change the training and testing set ratio accordingly and maximize the amount of new data when testing. However, the best way to ensure generalization of a model would be to introduce novel data for testing.

The results of the classification model could be further improved by introducing new data, performing more specific filtering as one of challenges regarding EEG classification task is noise, artifact removal. Moreover, in present thesis all images presented to the subjects were used to extract visual evoked potentials. Using only subset of images that are ranked as high in arousal or valence (attractiveness) might produce better results. This however assumes more data for the averaged evoked potentials to be noise-free. In general, finding good features appeared to be the main problem regarding successful classification. The statistical significance of the features tended to be small. To obtain features that would be more significant would be a good aim for next studies. In addition to the VEP features used in present thesis, some extra features could be introduced. For example, the disappearance of the images after 3 seconds might also have interesting evoked potentials in addition to just the appearance of the image. Thus, the significance of that aspect could also be further investigated in the future.

References

[1] World Health Organization, "Depression and other common mental disorders: global health estimates," WHO IRIS, 2017.

[2] F. S. d. A. Neto and J. L. G. Rosa, "Depression biomarkers using non-invasive EEG: A review," *Neuroscience & Biobehavioral Reviews,* vol. 105, pp. 83-93, 2019.

[3] M. Čukić, M. Stokić and S. Simić, "The successful discrimination of depression from EEG could be attributed to proper feature extraction and not to a particular classification," *Cogn Neurodyn,* vol. 14, pp. 443-455, 2020.

[4] A. D. Nazhvani, R. Boostani, S. Afrasiabi and K. Sadatnezhad, "Classification of ADHD and BMD patients using visual evoked potential," *Clinical Neurology and Neurosurgery,* vol. 115, no. 11, pp. 2329-2335, 2013.

[5] M. Mohammadi, F. Al-Azab and B. Raahemi, "Data mining EEG signals in depression for their diagnostic value," *BMC Med Inform Decis Mak,* vol. 15, p. 108, 2015.

[6] P. Fitzgerald and B. Watson, "Gamma oscillations as a biomarker for major depression: an emerging topic," *Transl Psychiatry,* vol. 8, p. 177, 2018.

[7] B. Hosseinifard, M. H. Moradi and R. Rostami, "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal," *Computer Methods and Programs in Biomedicine,* vol. 109, no. 3, pp. 339-345, 2013.

[8] E. Jesulola, C. F. Sharpley and L. L. Agnew, "The effects of gender and depression severity on the association between alpha asymmetry and depression across four brain regions," *Behavioural Brain Research,* vol. 321, pp. 232-239, 2017.

[9] M. Bachmann, J. Lass and H. Hinrikus, "Single channel EEG analysis for detection of depression," *Biomedical Signal Processing and Control,* vol. 31, pp. 391-397, 2017.

[10] K. Kalev, M. Bachmann, L. Orgo, J. Lass and H. Hinrikus, "Lempel-Ziv and multiscale Lempel-Ziv complexity in depression," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, 2015.

[11] A. Mehran, A. Hojjat and A. Amir, "Fractality analysis of frontal brain in major depressive disorder," *International Journal of Psychophysiology,* vol. 85, no. 2, pp. 206-211, 2012.

[12] F. L. A., A. C. I., M. H. A., C. Cai and S. Horvath, "Resting-State Quantitative Electroencephalography Reveals Increased Neurophysiologic Connectivity in Depression," *PLoS ONE.*

[13] O. L., B. M., K. K., J. M., R. J. and H. H., "Resting EEG functional connectivity and graph theoretical measures for discrimination of depression," in *IEEE EMBS*

International Conference on Biomedical & Health Informatics (BHI), Orlando, 2017.

[14] J. S. Bedwell, C. C. Spencer, C. C. Chan, P. D. Butler, P. Sehatpour and J. Schmidt, "The P1 visual-evoked potential, red light, and transdiagnostic psychiatric symptoms," *Brain Research,* vol. 1687, pp. 144-154, 2018.

[15] E. M. Ruohonen and P. Astikainen, "Brain responses to sound intensity changes dissociate depressed participants and healthy controls," *Biological Psychology*, vol. 127, pp. 74-81, 2017.

[16] E. N. Palmwood, J. W. Krompinger and R. F. Simons, "Electrophysiological indicators of inhibitory control deficits in depression," *Biological Psychology,* vol. 130, pp. 1-10, 2017.

[17] M. Grunewald, D. M. and D. Brandeis, "Attenuated LPP to Emotional Face Stimuli Associated with Parent- and Self-Reported Depression in Children and Adolescents," *Journal of Abnormal Child Psychology,* vol. 47, pp. 109-118, 2018.

[18] W. Mumtaz, A. S. Malik, S. S. A. Ali and M. A. M. Yasin, "P300 intensities and latencies for major depressive disorder detection," *IEEE International Conference on Signal and Image Processing Applications (ICSIPA),* pp. 542-545, 2015.

[19] A. L. Goldberger, C.-K. Peng and L. A. Lipsitz, "What is physiologic complexity and how does it change with aging and disease?," *Neurobiology of Aging*, vol. 23, no. 1, pp. 23-26, 2002.

[20] B. M.M. and L. P.J., "International Affective Picture System," *Encyclopedia of Personality and Individual Differences,* 2017.

[21] S. Luck, "An Introduction to the Event-Related Potential Technique," 2005.

[22] J. TP, M. S, H. C, L. TW, M. MJ, I. V and S. TJ, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology,* vol. 37, pp. 163-78, 2000.

[23] D. MR, C. P and A. JT, "Neurophysiological correlates of suicidal ideation in major depressive disorder: Hyperarousal during sleep," *J Affect Disord*, vol. 212, pp. 160-166, 2017.

Appendix 1 – Non-exclusive licence for reproduction and publication of a graduation thesis[1]

I Kaspar Haavajõe

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis

   "Depression discrimination based on visually evoked potentials of electroencephalography" , supervised by Laura Päeske.

   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;

   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.

2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

10.05.2021

1 The non-exclusive licence is not valid during the validity of access restriction indicated in the student's application for restriction on access to the graduation thesis that has been signed by the school's dean, except in case of the university's right to reproduce the thesis for preservation purposes only. If a graduation thesis is based on the joint creative activity of two or more persons and the co-author(s) has/have not granted, by the set deadline, the student defending his/her graduation thesis consent to reproduce and publish the graduation thesis in compliance with clauses 1.1 and 1.2 of the non-exclusive licence, the non-exclusive license shall not be valid for the period.