Norwegian University
of Life Sciences

Master's Thesis 2020 30 ECTS
Faculty of Science and Technology
Professor Cecilia Marie Futsæther

The Effects of Methylphenidate on Brain Structures of ADHD-Diagnosed Children

Explorative Analyses using Radiomic Features

Inger Annett Grünbeck
Environmental Physics and Renewable Energy

Acknowledgements

I am very grateful for all the help I have received while working on my thesis over the last seven months. First and foremost, I would like to thank my main supervisor Prof. Cecilia Marie Futsæther, for valuable support and encouragement through the entire writing process. I could not have wished for a better supervisor.

I would also like to thank Ass. Prof. Oliver Tomic, Dr Inge Groote, Prof. Atle Bjørnerud and MSc Aurora Rosvoll Grøndahl for answering countless questions, and helping me on the way. My thanks also go to Prof. Liesbeth Reneman and PhD Anouk Schrantee for providing the dataset analysed in this thesis.

Vilde and Maylinn, thank you for supporting me, cheering me on and of course proofreading. Last but not least, I have to thank my family for supporting me through my entire school time. I could not have done it without you.

Ås, 06.08.2020

Inger Annett Grünbeck

Abstract

The main objective of this thesis was to examine whether the treatment of ADHD-diagnosed male children using methylphenidate-based medication (MPH) caused changes to selected subcortical brain structures. The effects of MPH were examined by comparing MPH-treated trial subjects to a placebo cohort. Radiomics was utilised to extract high-dimensional datasets from T1-weighted MR images, which were obtained as part of The effects of Psychotropic drugs On Developing brain study (ePOD). The analyses performed in this thesis were limited to the trial subjects' caudate, hippocampus, pallidum, putamen and thalamus.

Radiomic shape and texture features were extracted from the left and right side of the selected brain structures and analysed. The analyses included a comparison of the subjects' structure surface area and volume and a principal component analysis of the extracted shape and texture features. Further, classification experiments were performed, predicting the treatment method by cross-validating six classification algorithms and four feature selectors. The performance of the classification experiments was evaluated using the area under the receiver operating curve (AUC).

According to the analyses performed, the subjects' right pallidum's suface area and volume were enlarged after methylphenidate-treatment. Furthermore, the MPH-treated trial subjects showed signs of an increased right hippocampal volume and right thalamus surface area. None of the principal component analyses performed based on the shape features or the texture features appeared to be able to distinguish the MPH-treated subjects from the placebo group, indicating that no changes due to the medication were detected by the principal component analyses.

The classification experiments of the pallidum and the putamen associated features yielded AUC scores close to 80. The best performing models in the caudate's, hippocampus' and the thalamus' classification experiments resulted in AUC scores close to 65, 75 and 70, respectively. The highest AUC scores were achieved by combining the Variance Threshold feature selector with either the Decision Tree classifier or the Extremely Randomised Tree classifier, or by combining the Light Gradient Boosting Machine classifier with either the Fisher Score or ReliefF feature selectors. The AUC scores achieved by the classification experiments combined with the identified changes in the structures' surface area and volume, indicated that MPH-medication may cause detectable, significant changes in male childrens' brain structures which should be examined closer.

Contents

List of Abbreviations

ADHD	Attention-Deficit/Hyperactive Disorder
AUC	Area Under the Receiver Operating Curve
DA	Dopamine
DT	Decision Tree Classifier
ePOD	The effects of Psychotropic drugs On Developing brain study
ET	Extremely Randomised Tree Classification
FN	False Negative
FP	False Positive
FPR	False-Positive Rate
GLCM	Gray Level Co-occurence Matrix
GLDM	Gray Level Distance Matrix
GLRLM	Gray Level Run Length Matrix
GLSZM	Gray Level Size Zone Matrix
ID	Identification Number
IQR	Interquartile Range
LGBM	Light Gradient Boosting Machine
LR	Logistic Regression
MPH	Methylphenidate
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
NGTDM	Neighbouring Grey Tone Difference Matrix
PC	Principal Component
PCA	Principal Component Analysis
Q - T plot	Q - residual and Hotelling's T^2 plot
Q1	First Quantile
Q3	Third Quantile

Ridge	Ridge Regression
RMSE	Root Mean Square Error
RMSEC	Root Mean Square Error of Calibration
RMSECV	Root Mean Square Error Cross Validation
ROI	Region of Interest (left or right side of brain structure)
SD	Standard Deviation
SVC	Support Vector Classifier
TN	True Negative
TP	True Positive
TPR	True-Positive Rate

Chapter 1

Introduction and Motivation

Attention-Deficit/Hyperactivity Disorder (ADHD) is ranked as the most commonly diagnosed neurodevelopmental disorder, with a worldwide occurrence of approximately 5 - 8 % amongst children and adolescents [1–5]. Behavioural symptoms such as hyperactivity, impulsivity and inattention arise during childhood due to the disorder and may persist into adulthood [4, 6]. In addition, ADHD has been associated with low performance in school, family problems and an increase in social challenges [3, 5]. The number of children diagnosed with ADHD has increased drastically during the last decades, accompanied by an increase in prescriptions of methylphenidate-based medication (MPH) [2].

MPH is one of the most common prescribed psychostimulants reducing symptoms of ADHD [3]. The medication has been proven to have an efficiency up to 80 % [1, 3] and safe to use according to literature [7]. However, due to the still-developing brain of children, MPH-medication may result in undesirable, yet unknown, long term effects [8, 9]. The combination of lack of knowledge regarding the effects of MPH on the brain structure and function [7], and the heightened number of children exposed, has led Schrantee et al. (2020) to emphasize the need for closer examinations of the medications' effects on patients' brains [9].

Papers linked to the effects of Psychotropic drugs On the Developing brain (ePOD) study reported on short-term age-dependency effect(s) of MPH such as changes in the dopamine (DA) transport system [9, 11]. In addition, Schrantee et al. (2020) assessed possible biomarkers related to nonresponse to MPH as part of the study [10] and Bouziane et al. (2019) examined the medication's effect on human brain white matter [1]. However, none of the papers related to the ePOD study have assessed the structural changes in the human brain grey matter due to MPH, including textural changes.

Extraction of quantitative features from medical images, known as radiomics [12], enabled this thesis to analyse mentioned changes. The concept of radiomics is based on the idea that medical images contain information about underlying illnesses and disorders not visible to the human eye, but detectable through quantitative image

analyses [13]. Applying radiomics to medical images converts them to minable, high-dimensional datasets, which may be further analysed [12].

The primary objective of this thesis was to assess the effects of methylphenidate on the brain structure of ADHD-diagnosed patients aged ten-to-twelve, based on T1-weighted magnetic resonance images (MRI). The images analysed in this thesis were collected as part of the ePOD-MPH study [11]. Effects of the treatment were examined by extracting features from selected brain structures using radiomics. Characteristics discriminating MPH-treated patients from non-treated subjects were identified by comparing the groups' radiomic features using statistical significance tests and searching for distinct patterns separating the groups through principal component analyses.

Furthermore, machine learning was used to predict the subjects' treatment group based on the extracted features in order to determine the presence of distinct changes due to MPH-treatment. Multiple feature selectors and classifier algorithms were cross-validated for this purpose. An initial screening of the resulting models was performed, in order to provide recommendations for future work related to the topic of this thesis.

As no explicit guidelines on how to best pre-process MR images before extracting radiomic features from regions of interest located in the brain was found, a procedure based on recommendations and previous studies, such as Collewet et al. (2004), Isensee et al. (2018) and Duron et al. (2019), was implemented in this thesis [14–16]. The procedure was also designed to correct for possible inaccuracies due to segmentation.

The thesis is divided into several chapters, and structured according to the IMRaD model: Introduction, Method, Results and Discussion [17]. Chapter 2 covers the theory this thesis was based on, explaining ADHD, and its treatment using MPH, magnetic resonance imaging (MRI) and radiomics. Methods and materials applied in this thesis are presented in Chapter 3. This includes a description of the dataset, the segmentation and pre-processing of the MR images as well as the analyses performed. The thesis' findings and their assessment are placed in Chapter 4 and 5, respectively. Chapter 6 covers the conclusion, and Chapter 7 the appendices.

Chapter 2

Theory

2.1 Attention-Deficit/Hyperactivity Disorder

Attention-Deficit/Hyperactive Disorder (ADHD) is a prevalent world-wide, heterogeneous neuropsychiatric disorder, affecting around 5-8% of the world's children and adolescent population [1, 3, 18]. According to Wilens and Spencer (2010), the disorder may be caused by multiple factors such as environmental influences and a combination of small effects from numerous genes [18], and may persist into adulthood [19, 20].

In order to be diagnosed with ADHD, certain criteria have to be fulfilled. The symptoms used to diagnose patients were derived from behaviour patterns of children and adolescents, and include inattention difficulties, forgetfulness, distractibility, hyperactive-fidgeting, excessive talking and restlessness [18]. The symptoms can cause academic difficulties and low school achievements, and affect the patients interpersonal contact and family relationships negatively [3].

2.1.1 The Impact of ADHD on the Brain

In addition to affecting the behaviour of patients, ADHD has been proven to affect chemical brain processes such as the dopamine system [18]. The neurotransmitter dopamine (DA) plays an important role in our brains' reward system, and reacts to pleasant stimuli such as food and sex. Drugs, such as cocaine and amphetamine, increase the dopamine production in the reward centre, which can cause addictions [21]. Children not affected by ADHD will experience a release of dopamine by the reward center when exposed to positive stimuli, associating the stimuli to a similar reward in the future. This will reinforce behaviour connected to the stimuli, as illustrated in Figure 2.1 [20]. Tripp and Wickens (2009) suggest a delay in the expected dopamine response associated with the positive stimuli in children with ADHD [20]. The delivered dopamine may even be delayed to the point where it is ineffective, hindering the positive associations between the reward and the behaviour causing the stimulus, or the stimulus itself [20].

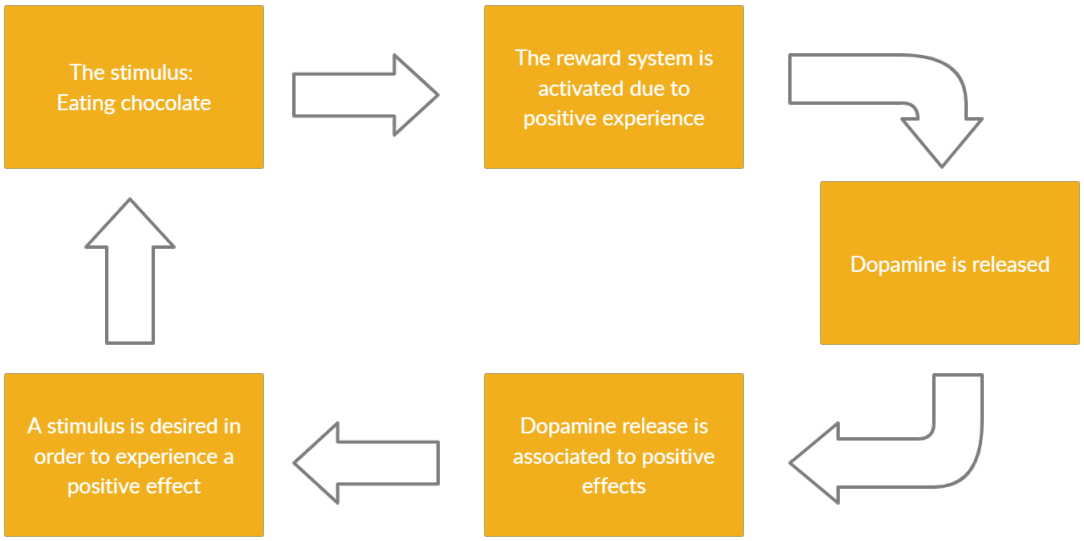


Figure 2.1: The figure illustrates the effect of dopamine. A positive stimulus activates the brain's reward center, releasing dopamine. The dopamine released simulates a positive effect, creating a desire for more stimulation. This desire reinforces the associated behaviour.

Imaging studies have also reported anatomical changes in ADHD-diagnosed children when compared to control cohorts [22]. According to Tripp and Wickens (2009), an overall reduction in brain size has been reflected in multiple studies [20]. Furthermore, reduced volumes of brain regions such as the caudate nucleus, the cerebellum, the pallidum and the corpus callosum have been identified [20, 22]. A mega-analysis performed by Hoogman et al. (2017) indicated significantly smaller volumes in ADHD-diagnosed trial subjects' caudate, amygdala, accumbens, hippocampus and putamen [22]. One more structure to be highlighted as affected by ADHD was the thalamus, according to Ivanov et al. (2010) [23]. All of the mentioned brain regions, except for the cerebellum and the corpus callosum, have been illustrated in Figure 2.2.

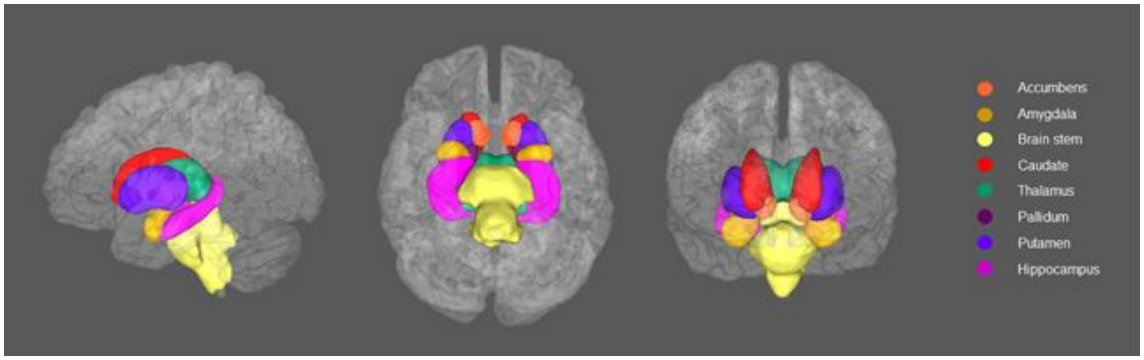


Figure 2.2: Subcortical structures related to ADHD such as the accumbens, the amygdala, the caudate, the hippocampus, the thalamus, the pallidum and the putamen were illustrated by Shen et al. [24].

2.1.2 Treating ADHD

When treating children for ADHD, educational remediation, parent training, psychotherapy and psychopharmacological therapy are being considered [4, 18]. Medication for ADHD may reduce hyperactivity, impulsivity and task-irrelevant activity in school, in addition to improving academic performances and parent-child interaction [3, 19]. Methylphenidate-based medications (MPH) such as Ritalin and Concerta are amongst the most commonly used compounds to treat ADHD in children effectively and safely, and considered the golden standard [7, 18, 25]. Methylphenidate-based medication increases extracellular dopamine (DA), and has an efficiency of 70 % in affected children [9, 25].

Even though the safety of the medication has been documented in literature, the long-term effects on the developing brain of children are still not fully understood [7, 9]. The brains of children and adolescents are highly plastic, and its neural structures and functions are easily influenced by medications [7]. Due to this and the limited safety controls carried out with regard to the effect of the medicine on the DAergic system, methylphenidate-based medication is still the objective of multiple studies, such as The effects of Psychotropic drugs On Developing brain study (ePOD) [9, 11].

Methylphenidate-caused changes in the dopaminergic system were confirmed based on changes in the cerebral blood flow in the thalamus and the striatum, which consists of the caudate and the putamen region [9, 21]. According to Wilens and Spencer (2010), medication may normalise brain functions and structures caused by ADHD [18]. An example of possible normalisation is the normalised activation in the accumbens of MPH-treated patients, compared to a placebo cohort [18]. Frodl and Skokauskas (2012) also reported findings indicating a decrease in ADHD-associated brain abnormalities due to treatment [26].

2.2 Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is a widespread structural medical imaging technique, producing two- and three-dimensional images [27, 28]. The imaging modality is suited for acquiring detailed, anatomical images, and especially for neurological examinations. MRI provides information based on the hydrogen atom density of the examined area, and is therefore mainly linked to water and fat tissue, making it one of the most flexible medical-imaging modalities available [28].

The imaging technique does not involve ionizing radiation and is non-invasive [28, 29]. Because of its ability to be used as often as necessary and detecting a variety of neural disorders, magnetic resonance imaging is used in multiple studies [27, 28].

Three-dimensional MR images will be referred to as image stacks or volumes in this thesis. An image stack, or volume, of size $i \times j \times k$ consists of k slices, as illustrated in Figure 2.3. The k slices each represent an $i \times j$ - sized matrix, $\mathbf{M}_k$, containing intensity values. Each of the slices consist of $i \times j$ voxels, and each of the voxels are coupled to a corresponding intensity value.

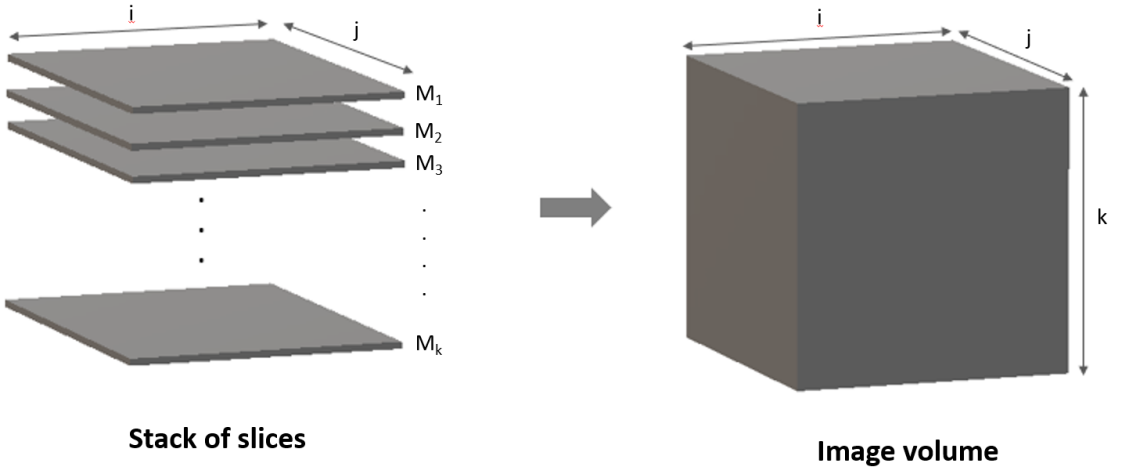


Figure 2.3: A three-dimensional MR image is referred to as an image volume or a stack of slices, were each slice corresponds to a two-dimensional MR image. Each of the k slice represents a $i \times j$ – sized matrix, $\boldsymbol{M}_k$. The bundled slices form an image volume of size $i \times j \times k$.

2.2.1 Intensity Representation in MRI

One of the disadvantages of using MRI are the arbitrary intensity values acquired [30]. Other imaging modalities use standardised units in order to achieve fixed representations of the tissue types across images. Computed tomography rescales its data relative to the intensity value of water before displaying the images. The rescaled values are referred to as Hounsfield Units. Positron emission tomography applies a semi-quantitative index, referred to as the standard uptake value (SUV), which would be equal to unity if the tracer used were to be distributed uniformly through the entire body [28].

In contrast to other modalities, the intensity value representation in MR images vary, despite of applying the same protocols, scanners or when imaging the same patient [27]. Due to the lack of standardised units for MRI, extra pre-processing is required when comparing images [29].

2.2.2 The Physics behind MRI

Clinical MRI is based on hydrogen nuclei, as they occur in the human body in abundance [27]. The hydrogen's proton possesses a spin angular momentum I and a charge $+e$, as illustrated by Figure 2.4. The charge can be considered as being distributed and rotating around the protons axis because of the angular momentum, resulting in a magnetic field [28].

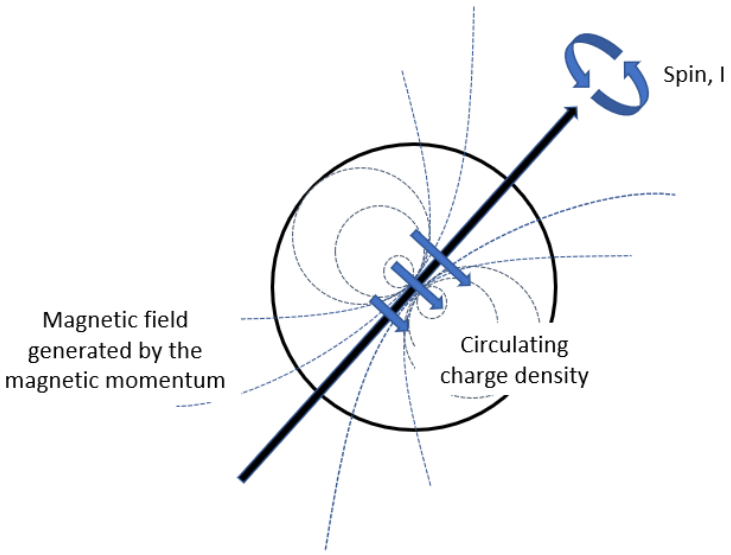


Figure 2.4: Representation of a nucleus with a spin angular momentum I. Due to the circulating charge density, a dipolar magnetic field is generated around the nucleus' axis. The figure was adapted from Fig 7.1 in Flower [28].

When considering an accumulation of nuclei, the axes are oriented randomly, as seen in Figure 2.6a [27, 28]. Introducing an external, static magnetic field B_0 causes the protons (see Figure 2.6b), and their net equilibrium magnetisation M_0 (see Figure 2.5a), to align with the field B_0 [27, 28]. By applying a pulse consisting of a weaker magnetic field B_1, oriented in the xy-plane, the net magnetisation M will experience a torque. The pulse causes an excitation of the nuclei, and rotates M by an angle α, as illustrated by Figure 2.5b. When the pulse stops, the protons will gradually re-align with the original field B_0. The realignment of the magnetic spin results in the emission of low-energy radio frequency photons, referred to as RF signals [28]. The RF signal is recorded by an RF-coil, and will be transformed into a medical image [27]. The realignment process illustrated by Figure 2.6d is referred to as relaxation, and the length of it depends on the nucleus under observation [28].

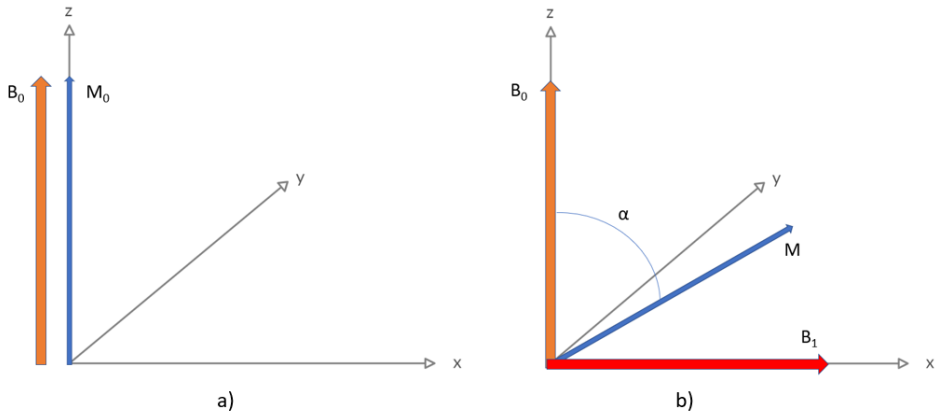


Figure 2.5: (a) If applying a static magnetic field B_0, as illustrated in Figure 2.6, the net eqilibrium magnetisation M_0 of the protons aligns with B_0. (b) When introducing a new, weaker magnetic field B_1 in the xy-plane, the net magnetisation is rotated by an angle α. The figure was adapted from Fig 7.3 in Flower [28].

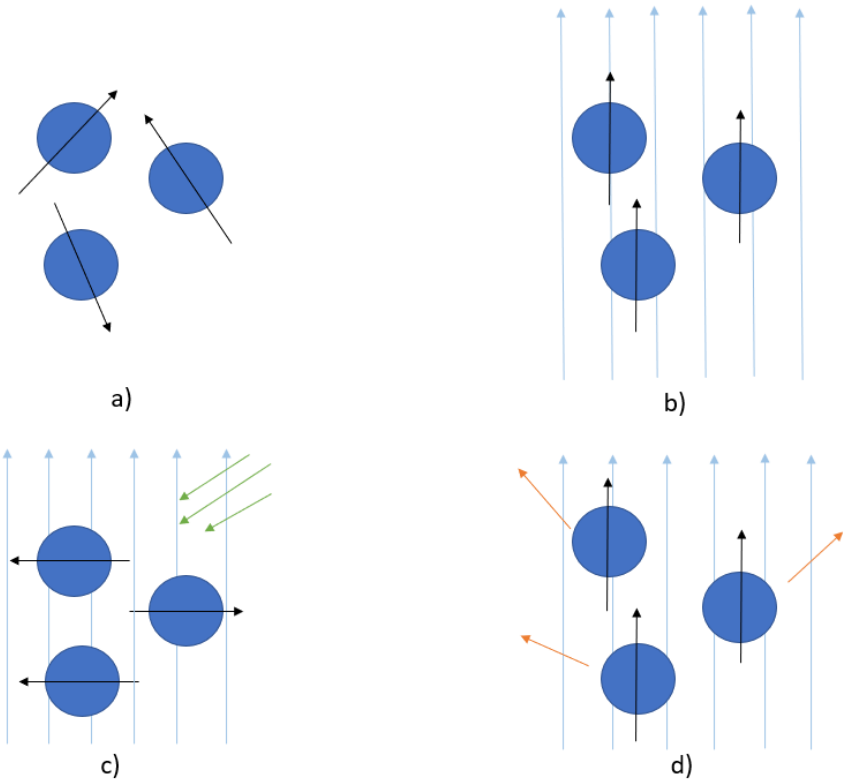


Figure 2.6: (a) The axes of nuclei are oriented independently of each other. (b) By applying a static magnetic field B_0, the axes align with the introduced field. (c) Introducing a pulse of a weaker, magnetic field B_1 (green arrows) causes the nuclei to excite and change the direction of the net magnetisation M. (d) When the pulse stops the protons re-align with B_0 and send out radio frequency photons (orange arrows), which are transformed into an image. The figure was inspired by Flower [28].

The RF signal's strength and duration depends on three main factors, each resulting in different image contrasts [27]. The images weighted based on the factors, are illustrated in Figure 2.7.

T1 - weighting

Spin-lattice relaxation reflects the time it takes for the net magnetisation M to reach its equilibrium M_0 [27]. T1-weighted images highlight fat tissue by imaging it white. In brain scans, the weighting colours the grey matter grey and the white matter white [28].

T2 - weighting

Transverse, or spin-spin relaxation, reflects the time it takes for the RF signal to decay in the transverse plane [27]. T2-weighted images highlight fat tissue as well as water. Similar to T1-weighted images, the brain's grey matter is coloured grey. However, the white matter is coloured dark, making it difficult to separate white from grey matter [28].

ρ - weighting

Proton density-weighting depicts fluids and fat tissue white. The weighting gives insight into cellularity, such as tumors, cell swelling and edema [28].

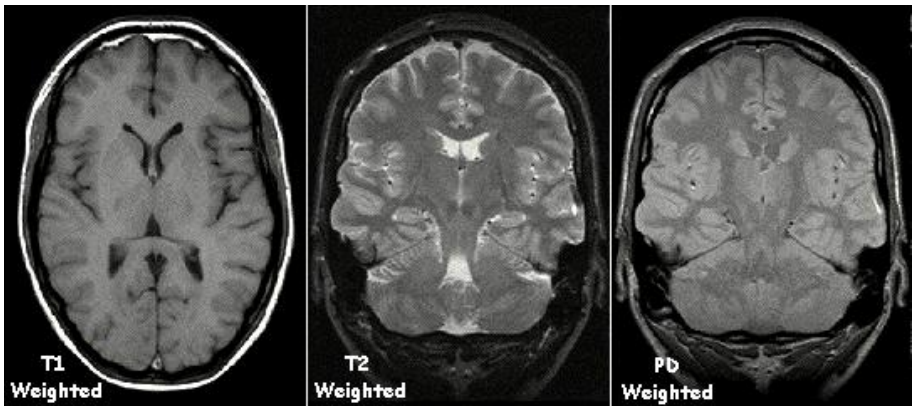


Figure 2.7: The contrast of magnetic resonance images can be varied by weighting the images, which will reveal different features. The figure illustrates a brain slice which was weighted with either T1, T2 or proton density weighting. The figure was collected from Maher [31].

2.3 Radiomics

Radiomics is an emerging field of research aiming to quantify information collected from biomedical images in the form of extracted features, in order to create mineable, high-dimensional datasets [12, 13, 32]. The development of radiomics was

motivated by the idea that medical images contain information reflecting underlying pathophysiology not detectable by the human eye [12, 13]. The information extracted by radiomics includes complex and possibly previous unknown patterns and markers which can be used to further research disease evolution, progression and treatment response, in addition to aiding in evidence-based decision-making and outcome prediction [13, 32].

The research surrounding radiomics has mainly been focused on and most developed in regard to oncology, but is in theory applicable to a wide range of diseases, disorders and other applications [12]. By identifying patterns and possible biomarkers associated with diagnoses, radiomics contributes to evolve and improve the related concept of personalised medicine [10, 12]. Personalised medicine, also called precision medicine, considers individual variations amongst patients, such as the patient's genes, environment and lifestyle, in order to customise their treatment [10, 12]. Biomarkers applied in personalised medicine can either be diagnostic and reflecting biological characteristics associated with a disease or of a predictive nature, describing processes tied to treatment response [10].

According to Schrantee et al. (2020), personalised medicine should also be applicable to psychoradiology, which uses medical images to analyse psychiatric disorders [10]. Biomarkers based on the human brain can provide molecular, anatomical and physiological characteristics, possibly enabling detection of Alzheimer disease and diagnosing autism spectrum disorder or ADHD [10, 33]. Biomarkers associated with radiomics are created by extracting features from a region of interest (ROI) from medical images. The radiomic features may include first order statistic, and features describing the ROI's shape and texture [34].

2.3.1 First Order Statistics

First order radiomic features describe the voxel intensity distribution of regions of interest by calculating commonly used statistical measures, such as the mean, maximum, minimum, skewness and kurtosis [13]. The features are based upon the images' grey-level histograms, which count the number of voxels, $P(i)$, of each intensity level, i, present in the region of interest as illustrated in Figure 2.8 [34].

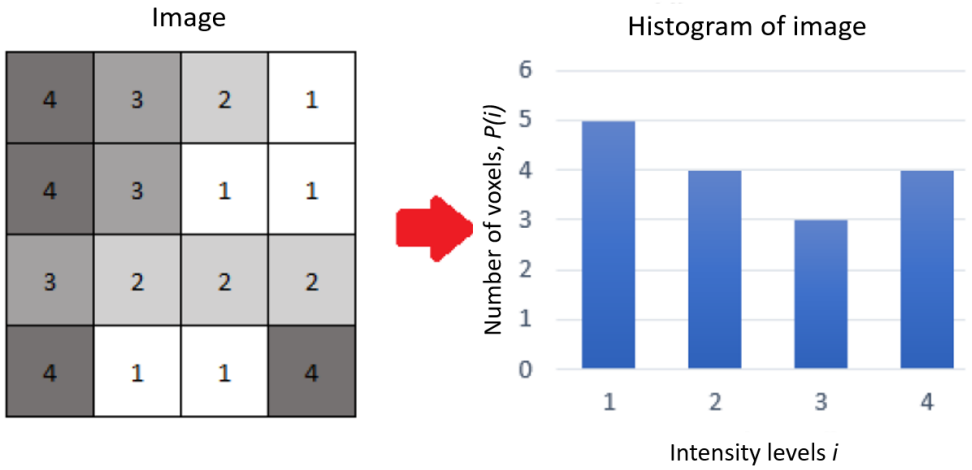


Figure 2.8: The figure illustrates how a histogram is calculated based on an image. The histogram counts the number of voxels, $P(i)$, of each intensity level, i, present in the image. First order statistics are then based upon the generated histogram. The figure was used with permission from Midtfjord (2018) [35].

2.3.2 Shape Features

Shape-based features describe the ROI's geometric properties such as the volume, surface area and maximum diameter [13, 32]. As the features are based solely on the shape of the ROI, they are independent of the voxels' grey-level value distribution in contrast to the first order and texture features [34].

2.3.3 Texture Features

Texture features describe the spatial information of voxel intensities found in images [32]. The features are measured by assessing the absolute gradient of the image, where the gradient represents the degree of grey-level intensity fluctuation across an image [13]. The gradient is at its highest if one of two neighbouring voxels is black and the other white, while the gradient measured between two voxels is zero if the intensity values are equal [13].

There are multiple texture feature subcategories, and five of them will be presented in this thesis. Each subcategory calculates a matrix, from which the texture features are calculated. As the matrices are based on the image's intensity values, the size of the matrices dependent on the number of grey-levels included in the image.

Grey Level Dependence Matrix

Analyses based upon the grey Level Dependence Matrix (GLDM) evaluate the dependency between neighbouring voxel intensities in an image [34]. Figure 2.9 illustrates how a two-dimensional GLD-matrix is calculated. Two voxels, a center voxel i and a neighbour voxel k with a distance δ, are considered dependent if $|i-k| \leq \alpha$, where α is a predefined constant. In the GLD-matrix $P(i,j)$, the $(i,j)^{th}$ element is

the number of times a center voxel i has j dependent neighbour voxels [34].

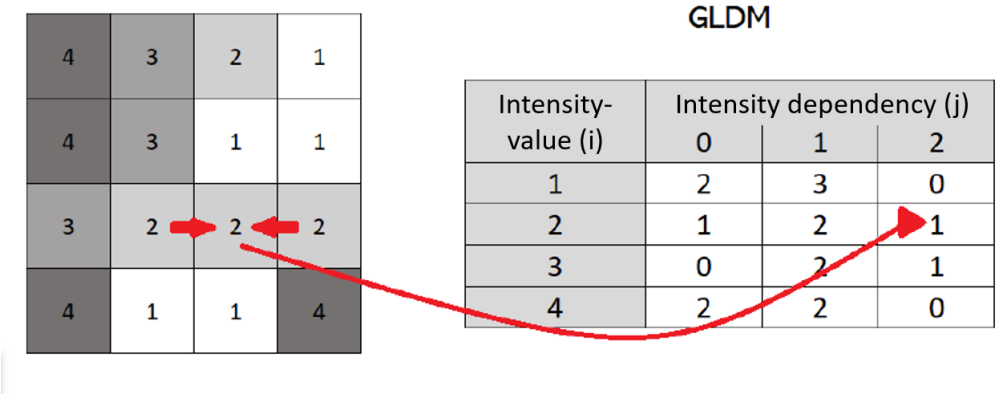
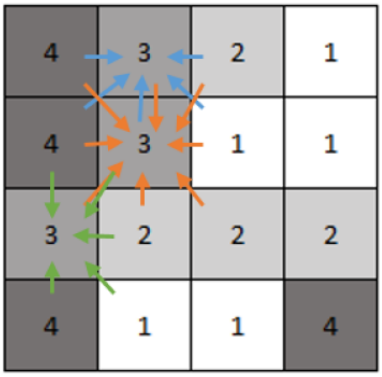


Figure 2.9: An example of how a GLDM may be calculated based on an image with four intensity levels. Here α and δ are set to 0 and 1, respectively. In matrix P, element $P(2,2) = 1$, as there only exists one center voxel with value two with two dependencies in the image. The figure was used with permission from Midtfjord (2018) [35].

Neighbouring Grey Tone Difference Matrix

Texture analyses using a Neighbouring Grey Tone Difference Matrix (NGTDM) assesses the difference between a voxel intensity i and the average intensity value of the voxels neighbours within a predefined distance δ [34]. Features extracted from an NGTD-matrix include coarseness, busyness, and complexity [13].

The NGTD-matrix contains the grey-level probability, p_i, and the sum of the absolute difference between the intensity value i and the neighbour voxels, s_i, as illustrated in Figure 2.10 [34].



NGTDM

i	n_i	p_i	s_i
1	5	0,3125	4,53
2	4	0,250	0,90
3	3	0,1875	0,78
4	4	0,250	6,00

Figure 2.10: An example of a two-dimensional NGTD-matrix. n_i corresponds to the number of voxels with intensity value i, p_i corresponds to the probability of intensity value i and s_i corresponds to the sum of the absolute difference between intensity value i and the neighbour voxels. The example figure illustrates the calculation of intensity value i=3 with a distance δ=1. As n_3=3 and the total number of voxels is 16, $p_3 = \frac{3}{16} = 0.1875$ and $s_3 = |3 - \frac{4+4+3+1+2}{5}| + |3 - \frac{4+4+3+2+2+1+2+3}{8}| + |3 - \frac{4+4+1+2+3}{5}| = 0.78$. The figure was used with permission from Midtfjord (2018) [35].

Grey Level Co-occurrence Matrix

Grey Level Co-occurrence Matrices (GLCM) capture spatial relationships between pairs of voxels in an image [13, 34]. The GLC-matrix $P(i, j)$ reflects how often a combination of the intensity values i and j occur within a predefined distance δ, and along angle θ in an image [34]. Figure 2.11 illustrates the calculation of a GLC-matrix in two dimensions with four intensity levels. The size of the matrix with n discrete intensity levels is $n \times n$, causing large, computational heavy matrices if the number of intensity levels included is too large [34].

GLCM

Intensity-value (i,j)	1	2	3	4
1	2	1	1	2
2	1	2	2	0
3	1	2	0	2
4	2	0	2	0

Figure 2.11: An example of how a GLCM is calculated with a distance = 1 between the voxels, and an angle $\theta = 0°$. Element (2,1) in the matrix equals one, as only one combination of the intensity values one and two occur horizontally in the image. The figure was used with permission from Midtfjord (2018) [35].

Grey Level Run Length Matrix

Grey Level Run Length Matrices (GLRLM) quantify intensity level runs, which are defined as a number of consecutive voxels with the same intensity values [13, 34]. In a GLRL-matrix $P(i, j|\theta)$, element (i, j) represents the number of intensity runs of intensity i and length j along angle θ [34], as illustrated in Figure 2.12.

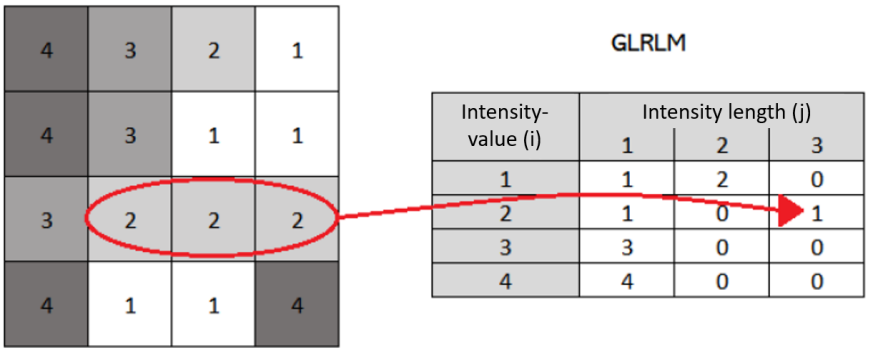


Figure 2.12: An example of how a GLRLM is calculated based on an image with four intensity values, and an angle $\theta = 0°$. Element (2,3) in the matrix equals one, as only one run with intensity level two and a run length of three occurs. The figure was used with permission from Midtfjord (2018) [35].

Grey Level Size Zone Matrix

The Grey Level Size Zone Matrix (GLSZM) resembles the GLRLM [13]. Instead of counting the number of runs, the GLSZM counts the number of zones, which is defines as the number of voxels connected with the same intensity value i [34]. Two voxels are considered as connected if the distance is one. As illustrated by Figure 2.13, the element (i, j) in a GLSZ-matrix $P(i, j)$ represents the number of zones with intensity value i and size j in an image [34].

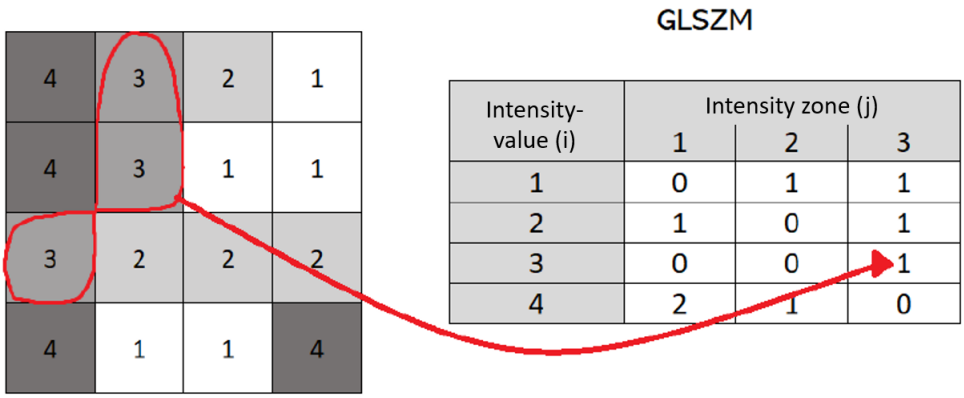


Figure 2.13: An example of how a GLSZM based on four intensity values is calculated. Element (3,3) equals one, as the image contains one zone with intensity value three and size three. The figure was used with permission from Midtfjord (2018) [35].

Chapter 3

Materials and Methods

The main objective of this thesis was to assess whether MPH-treatment causes traceable structural changes in brains of ADHD-diagnosed, ten to twelve-year-old trial subjects. Treated subjects were compared to a control group based on radiomic shape and texture features extracted from selected brain regions. The features were acquired through the Python package Biorad [36]. This chapter addresses how the data set was pre-processed before extracting radiomic shape and texture features, as no standard autoscaling process is established for MR images of brain structures.

Principal component analysis (PCA) was used to search for distinctive characteristics separating the trial groups. Moreover, machine learning was used to predict the group affiliations of the trial subjects based on the extracted features. Trial subjects standing out in the PCAs performed were identified. The feature selection methods and classifiers used were provided by the machine learning framework Biorad [36]. Biorad was developed as part of two earlier master theses, Albuni (2020) and Langberg (2019) [37, 38].

3.1 The ePOD-MPH Study

The effects of Psychotropic drugs On Developing brain study (ePOD) project was launched in March 2011, lasted four years and was monitored by the Clinical Research Unit of the Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands. The project was divided into three studies: ePOD-MPH, ePOD-SSRI and ePOD-Pharmo. In this thesis, only the ePOD-MPH study was considered. A full description of the study's method and design can be found in Bottelier et al. (2014), Schrantee et al. (2016) and the Netherlands National Trial Register (identifier NTR3103)(`https://www.trialregister.nl/trial/2955`) [9, 11].

ePOD-MPH was a multicentred randomised trial, designed to be double-blinded and placebo-controlled. The subjects were treated over a period of 16 weeks, followed by a washout period of 1 week. Participant MRI's were acquired before medication started (baseline session), after eight weeks of MPH and placebo treatment and at

the end of the washout period.

Subjects included in the ePOD-MPH study were distributed evenly between a group of children (age 10 to 12) and a group of adult (age 23 – 40) male subjects. Fifty children and fifty adults diagnosed with ADHD qualified to take part in the trial. Only males were included to limit subject variation [9, 11].

3.2 The Image Dataset

In accordance with the primary objective presented in Chapter 1, this thesis was limited to only analysing images acquired from ten to twelve-year-old trial subjects. Raw T1-weighted MRIs were used to analyse structural differences between placebo- and MPH-treated participants.

Participants were excluded from further analyses if either no baseline or 17-week image was acquired. Images disturbed by head motion also led to exclusion of images. Four participants were removed based on the exclusion criteria, leaving 22 MPH-treated and 24 placebo-treated children. Subjects placed in the placebo group will from here on be labelled as class 0 or placebo-treated. MPH treated subjects belong to class 1. Images acquired during the baseline session are further denoted as pre-treatment images, while the 17-week images are referred to as post-treatment images.

The subjects' trial IDs applied in the ePOD-MPH study were changed to consecutive numbers ranging from 0 to 45. The ID key converting between the subjects' original ID and their current ID can be found in Appendix A. Subjects nr. 0-21 were treated with MPH, while nr. 22-45 were part of the placebo cohort.

3.3 The Choice of ROI

Five subcortical brain structures were chosen to be analysed in this thesis. The structures were chosen based on findings presented by the ePOD-MPH study, articles explaining how ADHD affects the brain anatomy and recommendations received from Dr. I. R. Groote (personal communication, 29.01.2020).

Two of the structures chosen are the caudate and the putamen, which together are referred to as the striatum [21]. MPH caused a significant reduction of the dopamine transport system in young rodent striatum during animal trials, leading to behavioral abnormalities [11]. Related findings have been made in humans according to Bottelier et al. (2014), implying a structural change also may take place [11]. Schrantee et al. (2016) established a similar connection between MPH treatment of children and an increase in the cerebral blood flow response to a dopamine change in the striatum and thalamus [9]. Therefore, the thalamus was included as the third structure. The fourth and fifth selected structures were the hippocampus and

the pallidum, as multiple studies imply that both structures are affected by ADHD [22, 26, 39, 40]. This also applied for the selection of the caudate [20, 22, 26], the thalamus [23, 41] and the putamen [22, 26, 39].

In order to prevent confusion when referring to the selected structures, a distinction between the left/right side of a brain structure or the overall structure was defined in this thesis. When referring to both the left and right part of a brain structure, the term "brain structure" or the simplified "structure" are used. The term "ROI" denotes only one of the two, either the right or the left part of a specific brain structure.

3.4 Image Pre-processing

3.4.1 Segmentation

Before extracting radiomic features from a ROI, medical images were segmented, and a binary mask of each ROI created. Binary masks contain voxel labels which identify voxels belonging to the selected region of interest by setting their voxel label to one. Voxels not included in the region of interest were labeled as zero. Segmentation maps, or atlases, containing volumetric renderings of brain structures based on MRIs were generated by automated segmentation software. This thesis acquired the segmentation maps through the ePOD-MPH study, using the open-source software FreeSurfer(Martinos Center for Biomedical Imaging, Harvard-MIT, Boston USA). A binary mask of every ROI selected in chapter 3.3 was extracted from the maps using nordicICE (NordicNeuroLab, Bergen, Norway) (see Figure 3.1). Ten masks were generated for every trial subject, one mask for the left and one for the right part of each selected structure.

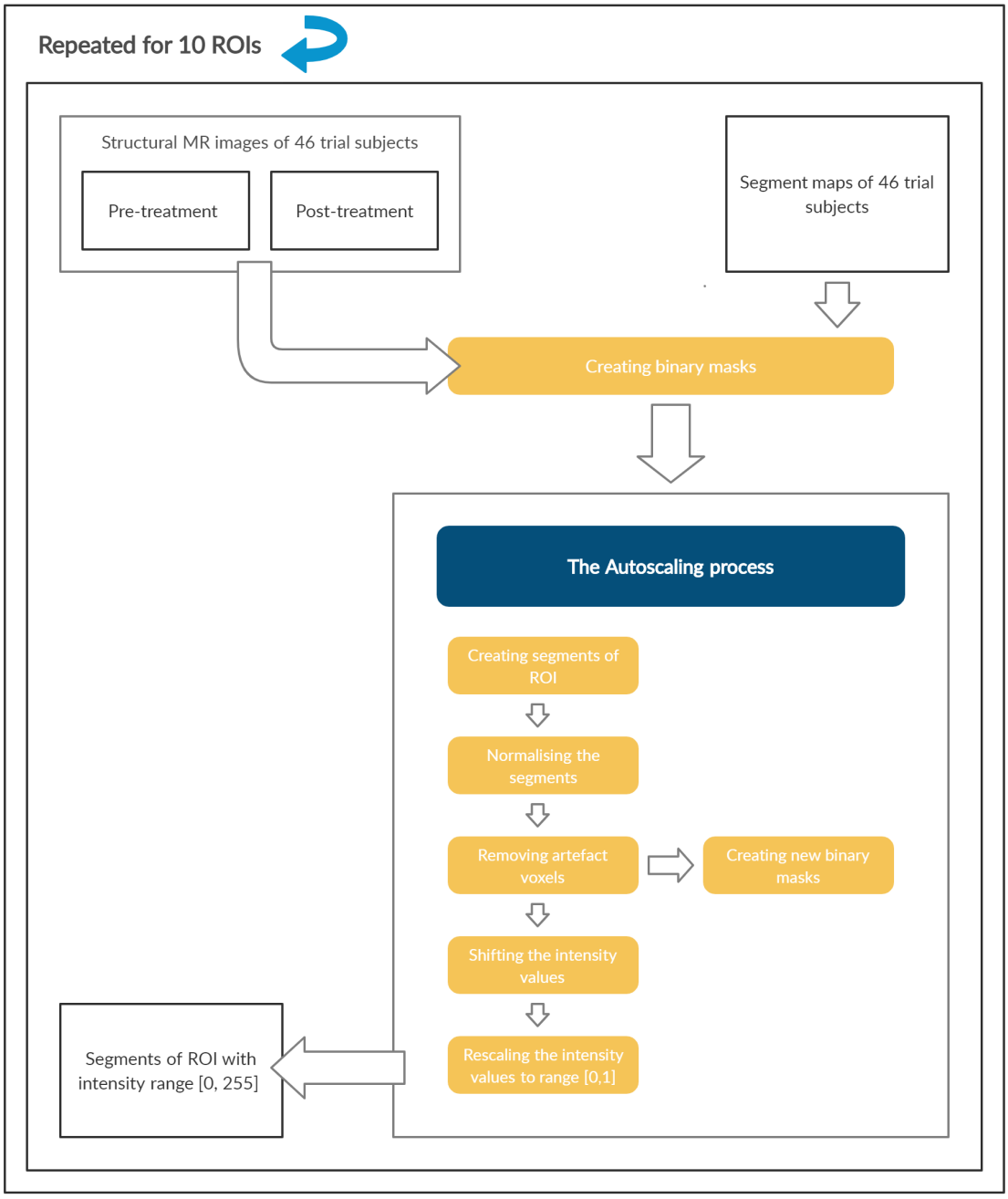


Figure 3.1: The workflow of the script creating binary masks for the ROIs and performing the autoscaling process before returning the segments of the ROIs corrected for artefact voxels and with normalised intensity levels. The corresponding script performong the autoscaling process can be found in Appendix B.

3.4.2 Artefact Voxels

Manual segmentation is considered the gold standard today, as automated segmentation processes carry the risk of inaccurate renderings of the ROI [32, 42]. In order to achieve reproducible segmentations Gillies et al. (2016) recommends improving the automated segmentation by performing manual corrections [12]. However, manual segmentation is time-consuming due to the large number of images and can be prone to intra- and interobserver variations [42, 43]. Therefore, to yield consistent

results, software such as FreeSurfer is often used to render brain structures, even though this may cause minor uncertainties when performing analyses with the segmentations.

Figure 3.2 illustrates an example of inaccurate segmentation by FreeSurfer. Incorrectly rendered voxels were marked with an ellipse. The left caudate is distinctly separated from the anterior horn of the lateral ventricle, as illustrated in Figure 3.3, emphasising the flawed segmentation provided by FreeSurfer in Figure 3.2. Voxels rendered incorrectly have been denoted as artefact voxels in this thesis. New binary masks were created during the autoscaling process illustrated in Figure 3.1, excluding artefact voxels from final segmented ROIs.

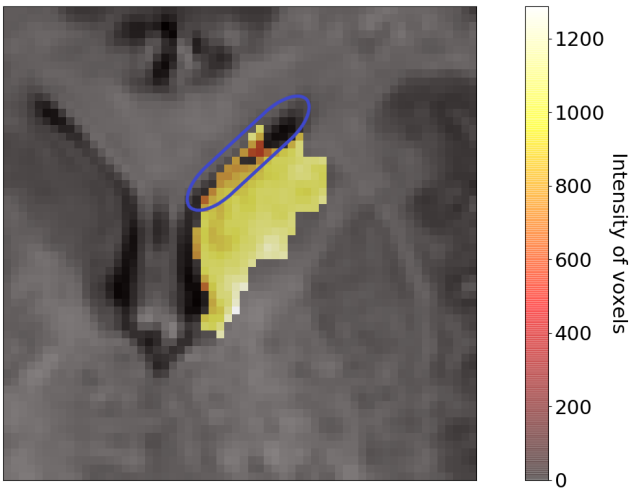


Figure 3.2: An example of inaccurate segmentation of a left caudate, rendered using FreeSurfer. The blue ellipse marks voxels possibly falsely included in the rendering, as they overlap with the anterior horn of the lateral ventricle if compared to Figure 3.3.

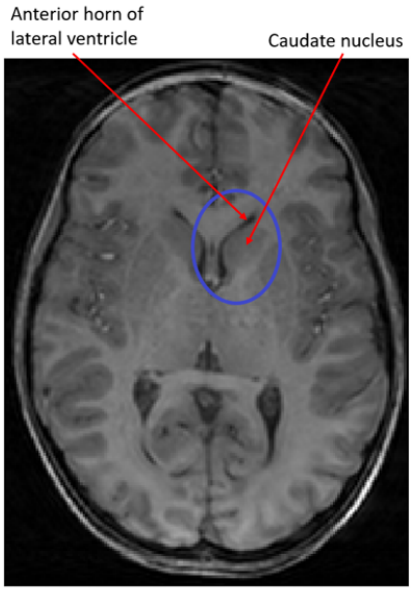


Figure 3.3: A brain slice with a blue ellipse marking the area illustrated in Figure 3.2. The red arrows mark the left caudate and the adjacent anterior horn of the lateral ventricle.

3.4.3 Intensity Normalisation

In order to compare medical images, similar intensity values must represent similar tissue types across patients. Computer tomography uses the Hounsfield scale, while positron emission tomography uses standardised uptake values [44, 45]. MRI, on the other hand, has no standard range for intensity values. The intensity range of MR images varies between scanners, protocols and patients [32, 46].

Different standardisation techniques can be applied to MR images to enable comparisons across patients and scanners. Matching of image histograms or scaling intensity values to a standard intensity range are methods suggested by several papers [14, 15, 29, 46, 47]. An autoscaling procedure based on Collewet et al. (2004), Isensee et al. (2018) and Duron et al. (2019) was applied in this thesis. Segmentations of the ROIs were created, normalised and rescaled to 8-bit images using the programming language Python [48], version 3.7.4. The workflow of the autoscaling process has been illustrated in Figure 3.1. The Python code can be found in Appendix B.

The Autoscaling Process

A list, L_{ROI}, of intensity values was created for each ROI. The intensity values included in L_{ROI} were extracted from the voxels of the pre- and post-treatment MR images of all trial subjects, using the corresponding masks generated by nordicICE. The autoscaling of L_{ROI} was initialised by normalising the list's content. Similar to Isensee et al. (2018), the z-scores of the intensity values, i, were calculated across L_{ROI} [15]:

$$z_i = \frac{i - \mu_{ROI}}{\sigma_{ROI}}. \tag{3.1}$$

μ_{ROI} and σ_{ROI} refer to the mean and the standard deviation of L_{ROI}, respectively. The normalised intensity values, z_i, were stored in another list, L_{norm}.

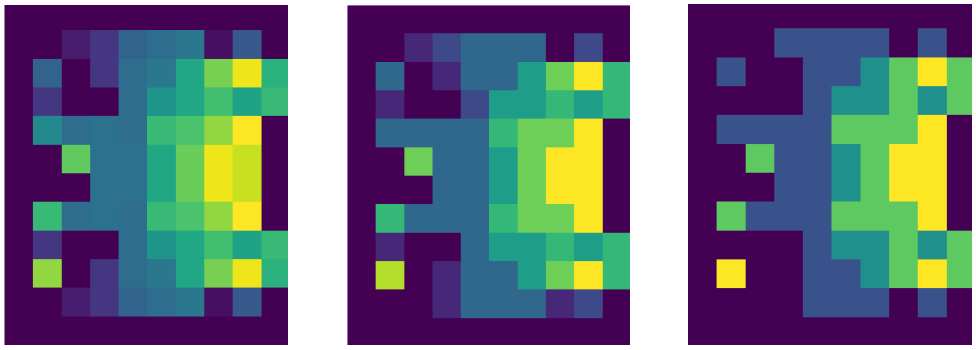
Similar to Hoogman et al. (2017), possible outlier voxels were removed based on their zscores to ensure no effects from FreeSurfer influenced the thesis' results, as described in Chapter 3.4.2 [22]. All voxels with intensity values located outside of the range $[\mu_{ROI} - 3\sigma_{ROI}, \mu_{ROI} + 3\sigma_{ROI}]$ were excluded from further analyses and removed from L_{norm} according to Collewet et al. (2004), Duron et al. (2019) and Zwanenburg et al. (2016) [14, 16, 30]. New binary masks were created based on the remaining voxels. Further, every remaining element, i, in L_{norm} was shifted by the global minimum to avoid negative values:

$$I_i = I_i - min(L_{norm}), \text{ for } i \text{ elements in } L_{norm}. \tag{3.2}$$

Finally, the intensity values were rescaled to the range [0,1], allowing the segmented ROIs to be extracted as 8-bit images. The intensity range of [0, 255] was chosen based on the size of the ROIs relative to the full brain size. A small bit-size is also favourable when extracting texture features. As explained in Chapter 2.3.3, the size of the texture features matrices depends on the number of intensity values included. A 12-bit image would generate a 4096 x 4096 - sized matrix. 8-bit images generate 256 x 256 - sized matrices, making them less computational heavy and therefore a preferable choice.

3.5 Image Intensity Discretisation

Before the radiomic features could be extracted, an image intensity discretisation was performed. By clustering voxels based on their grey-levels to a predefined number of clusters, discretisation facilitates the calculation of texture features and enables the construction of several distinctive feature sets [30, 49]. Further, discretisation should improve the robustness and reproducibility of the extracted features, and possess a noise-reducing effect [13, 30, 50]. Figure 3.4 demonstrates how an 8-bit image with 256 intensity levels can be discretised to images with ten and five intensity levels, referred to as 10-bin and 5-bin images respectively, from which unique texture features can be calculated.



(a) 8-bit image *(b) 10-bin image* *(c) 5-bin image*

Figure 3.4: The figure illustrates how intensity discretisation of an 8-bit image using 10 bins and 5 bins changes the number of intensity levels included in the images, and affects the images' texture.

Clustering of voxels can be performed by discretisation to a fixed number of bins or with a fixed bin size. The approach must be chosen based on the circumstances [30]. A general recommendation proposes choosing a bin size resulting in a bin number between 30 and 130, based on good performance and reproducibility in literature [34]. No clear recommendations or standards exist for choosing the number of bins or bin size when considering MR images [16].

In this thesis the images were discretised using two different bin sizes, creating two distinct datasets in addition to the original images. One dataset was discretised by clustering two and two intensity levels, reducing the images' intensity level range from 256 to 128 intensity levels. The second discretised dataset was generated by implementing a bin size of four, reducing the the intensity level range to 64 intensity levels. The discretised datasets were further reffered to as 128-bin dataset and 64-bin dataset.

3.6 The Radiomic Feature Extraction

Radiomic features were extracted from the autoscaled ROIs with code provided by the Biorad project [36]. The features generated by Biorad were calculated using the Python package Pyradiomics [51]. Fourteen shape features were extracted together with 75 texture features distributed between five subcategories, as shown in Table 3.1. The definitions of the extracted features applied in the thesis were defined by [34].

Table 3.1: Number of radiomic features extracted from the images in this thesis, distributed between shape features and five subcategories of texture features. The abbreviations of the texture feature subcategories can be found in Chapter 2.3.3.

	Texture				
Shape (3D)	GLCM	GLDM	GLRLM	GLSZM	NGTDM
14	24	14	16	16	5

In addition to extracting a set of shape features from the image dataset, two sets of texture features were acquired based on the discretised 128-bin and 64-bin datasets: a 128-bin and a 64-bin texture feature set. Altogether, six feature sets were extracted for every ROI, as the features of the pre- and post-treatment images were extracted separately. The extraction process is illustrated in the workflow-chart in Figure 3.5.

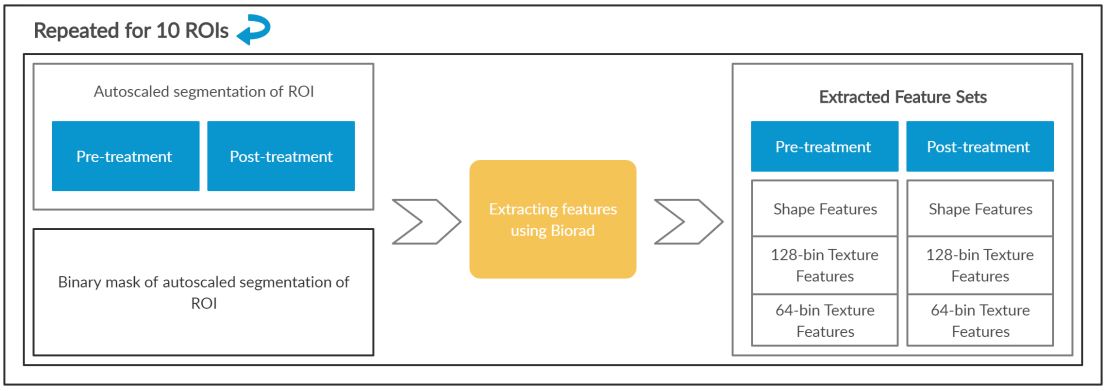


Figure 3.5: The workflow of the feature extraction process. Shape features and texture features from 128-bin images and 64-bin images were extracted from the pre-treatment and post-treatment images of the ROI, using the Python-package Biorad [36].

Each of the feature sets was extracted using the default parameters set by Pyradiomics [51]. These include the distance between voxels, δ=1, applied in GLDM, NGTDM and GLCM features, and the cutoff value of dependence in GLDM features, α=0.

First-order statistics were not included when extracting features, as the autoscaling process applied makes features calculated based on intensity-characteristics such as the minimum, maximum and mean, redundant. However, future analyses of the dataset should consider including features based on first-order statistics not negatively affected by intensity value normalisation.

3.6.1 The Feature Matrices

In order to assess changes in brain structures due to the assigned treatment, pre-treatment feature sets, Pre_{set}, where subtracted from the post-treatment feature sets, $Post_{set}$:

$$C_{(m,n)} = Post\_set_{(m,n)} - Pre\_set_{(m,n)}. \tag{3.3}$$

$C_{m,n}$ denotes the change of a single feature value belonging to feature, m, and trial subject, n, in an arbitrary feature set. By concatenating the feature change, $C_{m,n}$, of the left and right part of a structure, a shape feature matrix, a 128-bin texture feature matrix and a 64-bin texture feature matrix were constructed, as illustrated in Table 3.2. Combining the three feature matrices belonging to the same structure

resulted in five large matrices, one for each of the analysed structures, which were referred to as full feature matrices.

Based on the created full feature matrices, further matrices were generated by calculating the mean of the corresponding left- and right-sided features of a structure.

Table 3.2: An example of a feature matrix constructed based on the pre-treatment and post-treatment feature set of an arbitrary structure, using Eq. 3.3. Class 1 corresponds to MPH-treated, class 0 to the placebo group.

Participant ID	Class	Left segment feature 1	Left segment feature 2	Right segment feature 1	Right segment feature 2
0	1	$C_{l1,0}$	$C_{l2,0}$	$C_{r1,0}$	$C_{r2,0}$
1	1	$C_{l1,1}$	$C_{l2,1}$	$C_{r1,1}$	$C_{r2,1}$
...					
22	0	$C_{l1,22}$	$C_{l2,22}$	$C_{r1,22}$	$C_{r2,22}$
23	0	$C_{l1,23}$	$C_{l2,23}$	$C_{r1,23}$	$C_{r2,23}$
...					

3.7 Analysing the Extracted Features

The autoscaling process was examined to determine the effect of excluding artefact voxels on ROI surface area. The percentage of surface area change of the ROIs due to autoscaling was plotted. In addition, the effect of autoscaling on intensity values of trial subjects was assessed.

3.7.1 Preliminary Statistical Analysis

Structural changes such as the volume size of the caudate is thought to be associated with MPH treatment [10]. Therefore, the MPH and the placebo group were compared based on their ROIs surface area and volume. Shape feature's Mesh Volume was used as a volume measure, while the surface area corresponds to the same-named feature. The analysis was performed using boxplots. In addition to analysing the features extracted from the left and right brain structures, the mean of the left and right ROIs were calculated and analysed.

Boxplot

A boxplot is a standardised distribution representation of data based on a five-number summary: the first quartile (Q1), the third quantile (Q3), the "minimum", the "maximum" and the median. The first and third quartile were defined as the 25^{th} and the 75^{th} percentile, respectively. As illustrated in Figure 3.6, the interquartile range (IQR) spans from the first to the third quantile, covering approximately 50

% of the observations. The terms “minimum” and “maximum” refer to the outlier-limit. Observations outside of the range [Q1 - 1.5 IQR, Q3 + 1.5 IQR] were defined as outliers (see Figure 3.6) [52, 53].

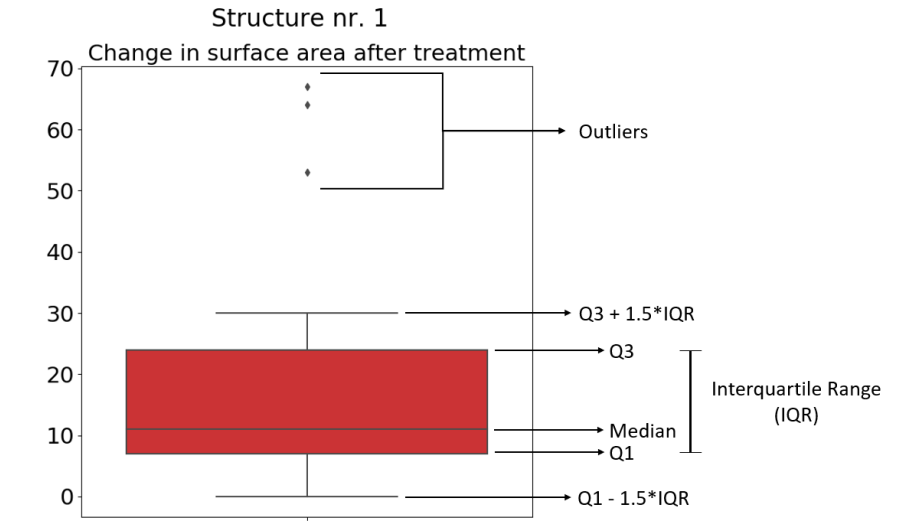


Figure 3.6: An example of a boxplot.

Using a boxplot enables the comparison of value range and distribution between datasets. Combining the plot with a statistical significance test confirms whether the datasets significantly differ from each other. In this thesis, Welch’s t-test was applied.

Welch’s t-test

A significance t-test evaluates whether two sample group’s mean value (μ_1, μ_2) differ from each other. Given the groups variance, s, and the sample size, n, the t statistics of Welch’s t-test is calculated as

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}. \tag{3.4}$$

Compared to other t-tests, Welch’s t-test does not assume equal variance or sample size if the assumption of a normal sample distribution is covered [54–56]. The test is also robust against a non-normal sample distribution, given an approximately homogeneous variance and a sample size difference close to zero [54, 57]. As the sample sizes were sligjtly dissimilar, Welch’s t-test was applied.

The significance test and annotation in the boxplot were performed using the Python package Statannot [58]. A 95% confidence interval was applied, alongside the annotation listed in Table 3.3. The difference in mean was accepted as significant given a p-value ≤ 0.05.

Table 3.3: The table displays the annotation used by the Python package statannot [58] to indicate the level of significance detected by the Welch's t-test.

Annotation	p-value
Not significant (ns)	$5.00 \times 10^{-2} < p \leq 1.00$
$*$	$1.00\times 10^{-2} < p \leq 5.00 \times 10^{-2}$
$**$	$1.00 \times 10^{-3} < p \leq 1.00 \times 10^{-2}$
$***$	$1.00 \times 10^{-4} < p \leq 1.00 \times 10^{-3}$
$****$	$p \leq 1.00 \times 10^{-4}$

3.7.2 Principal Component Analysis

Principal Component Analysis (PCA) was used to explore the feature matrices and detect patterns in this thesis. It is a dimensionality reducing linear transformation technique transforming datasets into orthogonal components, referred to as principal components. The goal of PCA is to identify the directions of highest variance and projecting the original d-dimensional dataset $\mathbf{X}$ onto a new k-dimensional subspace spanning the principal components, as illustrated in Figure 3.7. Typically, $k << d$ [59].

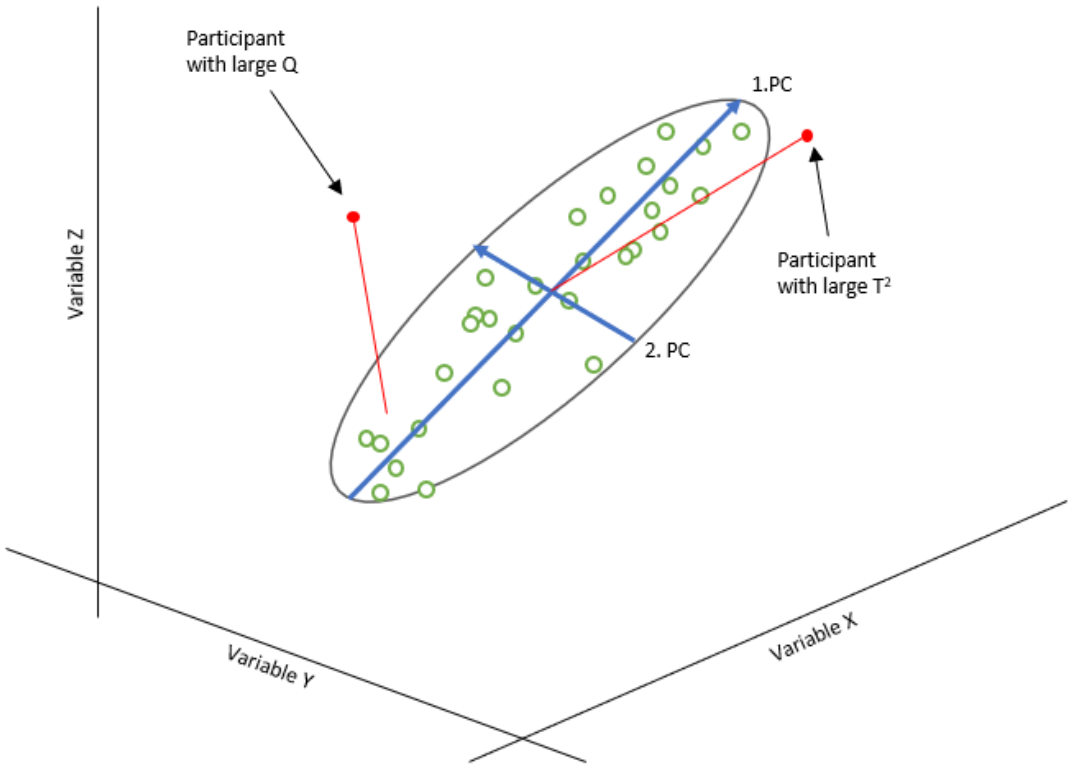


Figure 3.7: An illustration of how principal components span across the original subspace of the dataset. The first principal component (PC) captures the largest amount of variance possible. The second PC is orthogonal to first PC and captures the second-highest amount of variance. Two of the samples deviate from the PCA model due to a large Q residual or Hotelling's T^2. The illustration was created and modified based on Wise et al. [60].

By constructing the projection matrix $\mathbf{W}$ based on the k largest eigenvalues and the corresponding eigenvectors, the sample vector $\mathbf{x}$ can be mapped to a new feature subspace $\mathbf{z}$:

$$\begin{aligned} \mathbf{x} =& [x_1, x_2, ..., x_d], \quad \mathbf{x} \in \mathbb{R}^b \\ & \downarrow \mathbf{xW}, \quad \mathbf{W} \in \mathbb{R}^{d \times k} \\ \mathbf{z} =& [z_1, z_2, ..., z_k]. \quad \mathbf{z} \in \mathbb{R}^k \end{aligned} \tag{3.5}$$

Similarly, the dataset $\mathbf{X}$ can be projected unto the new subspace using $\mathbf{W}$, creating the matrix $\mathbf{T}$:

$$\mathbf{T} = \mathbf{XW}. \tag{3.6}$$

Equation 3.6 can be modified to include the variance of dataset $\mathbf{X}$ not explained by the PCA model as residual matrix $\mathbf{E}$:

$$\mathbf{X} = \mathbf{TW}^T + \mathbf{E}. \tag{3.7}$$

Elements of matrix $\mathbf{T}$ are often referred to as scores, and elements of $\mathbf{W}$ as loadings [59, 61].

The first principal component calculated when constructing a PCA model will contain the highest amount of variance possible, due to the projection of dataset $\mathbf{X}$ onto the new feature subspace. The following principal components PCs will contain the largest remaining variance of the dataset, given that they are orthogonal to the previous components. The maximum amount of variance extracted from a dataset corresponds to a 100 % cumulative variance. Variance captured by a single PC was denoted as the component's individual explained variance [59].

Principal component analyses are capable of de-noising datasets by excluding the last calculated principal components when calibrating the PCA model. The variance captured by these components can be classified as either noise or as non-deterministic [59, 60]. An inclusion of PCs explaining noise may result in an overfitted model [62]. How many components to retain can be decided based on PCA statistics such as the eigenvalues, the explained variance and the root mean square error of calibration and cross-validation.

Choosing the Number of Principal Components to Include

The manual of the chemometric tool PLS-Toolbox [60] suggests looking at the eigenvalues from smallest to largest, searching for sudden jumps between the values. The principal components corresponding to the eigenvalues below the jump should be considered excluded.

Further, the eigenvalues determine how much information from the original features each principal component covers, provided the dataset was autoscaled before calibrating the model. Therefore, excluding eigenvalues smaller than one reduces the

amount of noise captured by the model, as the corresponding PC covers less information than one original feature [60, 61]. A further rule of thumb recommendation is to include enough components to ensure 80% cumulative explained variance, but no more than 90-95% [61, 63].

The number of PCs to include can also be chosen based on the accuracy and the precision of the PCA model using the root mean square error (RMSE). RMSE refers in general to the squared sums of squared errors:

$$RSME = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}, \tag{3.8}$$

where y_i represents a sample and $\hat{y}_i$ the fitted or predicted corresponding value.

The RMSE of calibration (RMSEC) measures the goodness of fit between the dataset and the calibrated model. For every new included principal component in the model, a new RMSEC is calculated. y_i in equation 3.8 is replaced by a calibration sample and $\hat{y}_i$ by a prediction based on the entire calibration sample. As illustrated in Figure 3.8, the RMSEC decreases for each new PC in the model due to the information (variance) included. The RMSEC tends to flatten out near the optimal value but is not as well suited for deciding upon the number of PCs to include as the RMSE Cross-Validation (RMSECV) [62].

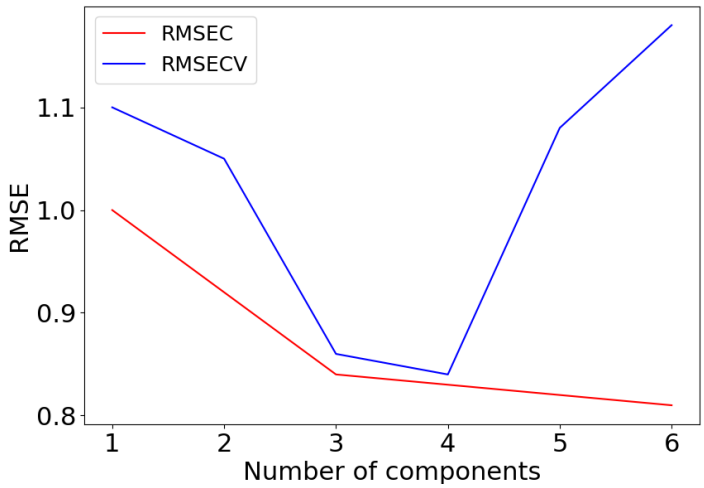


Figure 3.8: An example of how the RMSEC and the RMSECV of a PCA model might behave. While the curves decrease, information describing systematic variance is gathered by the model. An increasing RMSECV-curve indicates the model is collecting noise, or unusual information and overfits [61, 62].

RMSECV functions as model validation and estimates how well the calibrated model performs for unknown samples. Due to the lack of unknown data to perform the validation, RMSECV leaves out parts of the calibration data when constructing the

PCA model. $\hat{y}_i$ from equation 3.8 is predicted based on the remaining calibration data not used to build the model. Similar to the RMSEC, a new RMSECV is calculated for every new principal component included in the model. If an added PC describes large amounts of systematic variance, the RMSECV should decrease. The opposite will occur by including PCs describing noise, as illustrated in Figure 3.8. The PC corresponding to the minimum of the RMSECV should be the last PC included in the model [60–62].

Performing the PCA

PCA was performed using PLS-Toolbox [60], version 8.8.1, coupled to MATLAB [64], version 9.7.0.1296695 (R2019b). Fifteen PCA models were generated based on the shape, 128-bin texture, and the 64-bin texture feature matrices of the structures. PCA models were created using default settings. In addition to autoscaling the input data and using the singular value decomposition algorithm, cross-validation with Venetian Blinds was implemented.

PLS-Toolbox recommended the number of principal components to take into further consideration based on the eigenvalues and the RMSECV. If PLS-Toolbox suggested less than three PCs, a manual examination of the models' statistics was performed. A manual examination included assessing the eigenvalues, cumulative explained variance, RMSECV and RMSEC.

The principal component corresponding to the number of components included in a model was denoted as the model's "cutoff"-component. If, for example, eight components were included in a model, the eighth component would be the "cutoff".

Searching for Class Distinction

Plots of the model's scores, or score plots, were created using the statistical programming language R [65], version 3.6.2, together with RStudio [66], version 1.2.5033. The score plots were collected in pair plots, enabling a visual analysis of the principal components. As the number of components included in a pair plot was restricted to five, models exceeding this limit were split into multiple plots. Samples belonging to class 0 were coloured blue and class 1 samples' coloured red, making it possible to search for principal components distinguishing the classes.

Identifying Abnormal Trial Subjects

By visually examining the pair plots generated, individuals or clusters of subjects deviating from the norm were detected and identified, as illustrated in Figure 3.9. The extraction was performed component by component. To be able to identify the subjects, score plots created by PLS-toolbox were used.

Moreover, possible outliers were detected by examining the Q and Hotelling's T^2 statistics. The Q statistics, referred to as Q residuals in this thesis, measure unusual variation outside of the model based on the residual matrix **E** from equation 3.6 [60]. Figure 3.7 illustrates a sample with a large Q residual, corresponding to a considerable distance between the sample and the model's subspace. A sample with a large Hotelling's T^2 is also illustrated in the figure. Hotelling's T^2 measures the variation in each sample within the model [60]. The Q and Hotelling's T^2 statistics were plotted as scatter plots using PLS-Toolbox. Such plots are further referred to as Q - T plots. Participants with a Q residual or Hotelling's T^2 value outside of the 95% confidence interval limit were regarded as possible outliers in this thesis.

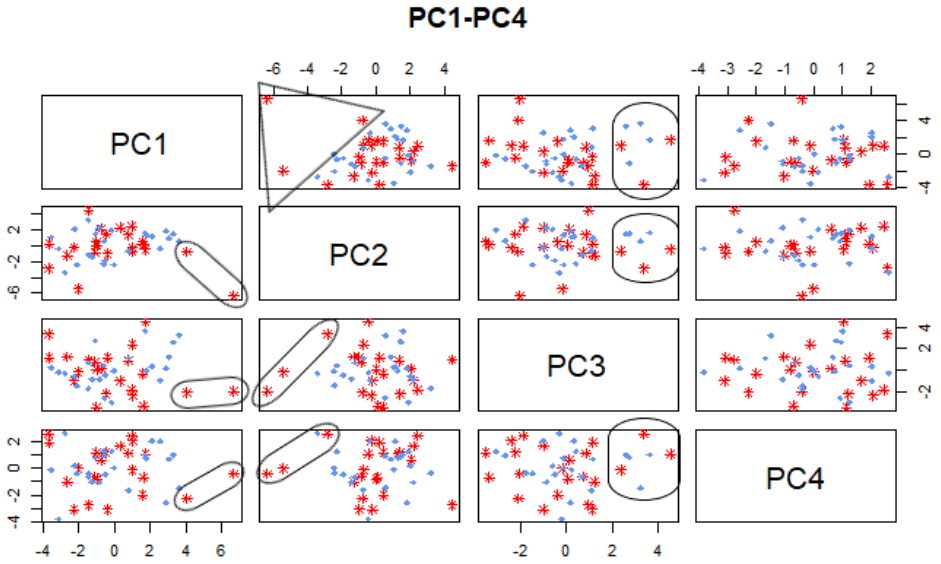


Figure 3.9: Individuals and clusters of subjects standing out in generated score plots were detected and identified, as shown in this example figure.

3.7.3 The Classification Experiment

In this thesis, various supervised classification algorithms were utilised to assess whether class 0 was distinguishable from class 1 based on the extracted radiomic features. In contrast to the PCA algorithm, supervised classification is driven by the samples' class labels. The goal of supervised classification is to predict the class labels of unseen or future samples, based on past observations [59]. Two classification experiment was performed for every brain structure, predicting the subject's treatment. The first experiment was performed based on the full feature matrices, analysing the radiomic features extracted from the left and right brain structures. In the second experiment, the mean values of the left and right brain structures' features were calculated and used to perform the classifications.

The classification experiments were executed using Biorad [36], based on the full feature matrices. Every experiment implemented cross-validation of six classification algorithms, combined with four feature selectors. In addition, classifications

were performed without feature selection, resulting in thirty classifications for every experiment. Biorad implemented Scikit-Learn's RandomizedSearchCV-package [67] in order to perform hyperparameter optimalisation and cross-validation of the classification experiments.

The classification algorithms applied by Biorad were the linear classifiers logistic regression [59], ridge regression [59] and the support vector classification [68]. Further, the tree based learning algorithm decision tree classification [59], light gradient boosting machine [69] and extremely randomised tree classification [70] were included. The abbreviations of the classifiers used in this thesis were listed in Table 3.4.

Removing irrelevant and redundant information by excluding features from the datasets before classification reduces the complexity of a model and increases the performance [59, 71]. The feature selectors implemented by Biorad were ReliefF [72], Mutual Information Classifier [73] combined with Scikit-Learn's SelectKBest-package [67], Fisher Score [74] and Low Variance Threshold [74].

Table 3.4: Abbreviations of the classifier algorithms implemented by Biorad [36] and applied in this thesis.

ET	Extremely Randomised Tree Classification
DT	Decision Tree Classifier
LGBM	Light Gradient Boosting Machine
LR	Logistic Regression
Ridge	Ridge Regression
SVC	Support Vector Classifier

Evaluating the classification experiments

The area under the receiver operating curve (AUC) was used to measure the performance of the classification experiments. The AUC is determined by plotting the true-positive rate (TPR) and the false-positive rate (FPR), as illustrated in Figure 3.10, and calculating the area under the curve. When classifying samples, the prediction result can either be true positive (TP), false positive (FP), true negative (TN) or false negative (FN). If a prediction is true positive or true negative, the sample has been classified correctly [71, 75]. In this thesis, a positive classification corresponds to predicting class 1 and negative to class 0.

The true-positive rate was defined as [75]

$$TPR = \frac{TP}{TP + FN} , \tag{3.9}$$

and the false-positive rate as

$$FPR = \frac{FP}{TN + FP} . \tag{3.10}$$

As illustrated by the receiver operating curve in Figure 3.10, the AUC can range from 0.0 to 1.0, where 0.0 indicates no correct and 1.0 no incorrect classifications. An AUC of 0.5 suggests the classification essentially was performed at random [71]. The Biorad-package [36] illustrated the prediciton models' AUC-scores of the classification experiments as heatmaps, and scaled the scores by 100.

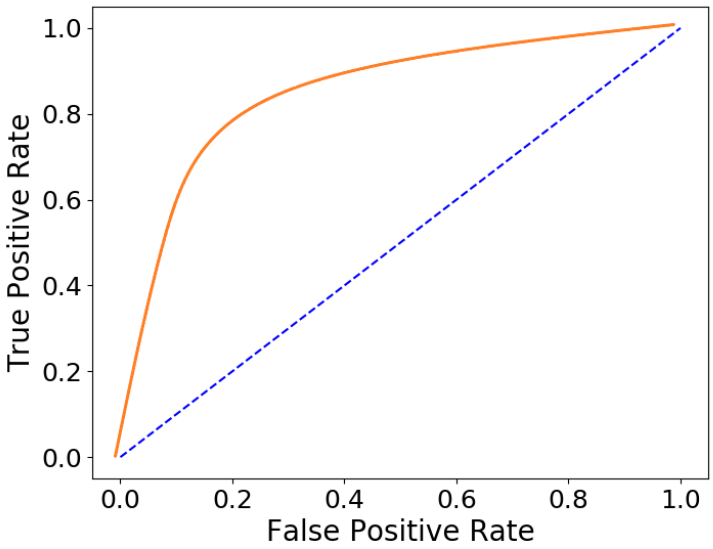


Figure 3.10: An example of a receiver operating curve (in orange). The area under the curve (AUC) was used to measure the prediction models' performance in this thesis. The blue dotted line marks were the classification essentially would have been at random.

Chapter 4

Results

The main aim of this thesis was to determine whether ADHD treatment-induced changes in brain structure geometry and texture. Computation of texture features required intensity rescaling of the images. The effects of the autoscaling process were analysed in this chapter, followed by an assessment of the statistical analysis results. Further, the generated score plots were examined together with the corresponding Q-T plots. The performances of the classification experiments were evaluated based on the calculated AUC. In addition, the selection rate of the features and the feature category distribution were examined.

All definitions and defined terms from Chapter 3 were adopted when presenting the results.

4.1 Evaluating the Autoscaling Process

Intensity normalisation was performed as described in Chapter 3.4.3. Figure 4.1 displays a slice of subject nr. 0's left caudate before and after performing the autoscaling. Due to the normalisation of intensity values and scaling to 8-bits the ROI displays sharper contrasts, highlighting texture variations. Moreover, voxels with corresponding intensity values outside of range $[\mu_{ROI} - 3\sigma_{ROI}, \mu_{ROI} + 3\sigma_{ROI}]$ were removed before extracting the radiomic features. This effect could be observed in the upper left of the caudate in Figure 4.1.

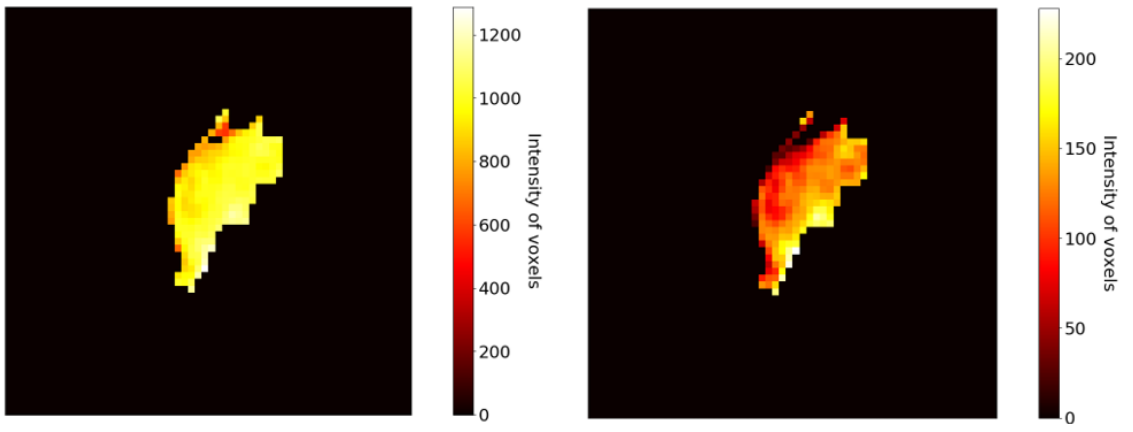


Figure 4.1: A slice of subject nr. 0's left caudate before and after performing the autoscaling process. Due to intensity level normalisation and exclusion of possible outlier voxels, the intensity distribution has changed and voxels disappeared in the caudate's upper left corner.

4.1.1 Examining the Change in Surface Area Due to Voxel Exclusion

The effects of the artefact voxel removal were assessed by calculating and plotting the percentage change of the ROIs' surface area due to the removal. Four plots were created for every structure, illustrating the change of the left and right structure side before and after treatment, as shown for the putamen in Figure 4.2. The figures of the remaining structures can be found in Appendix C. The mean value of the percentage change and three standard deviations (SD) were marked in the plots, enabling a comparison of the structures.

On average, the surface area of the trial subjects' caudate and thalamus decreased after voxel correction, according to Figures C.1 and C.4. However, an increase of the subjects putamen, hippocampus and pallidum surface area can be observed in Figure 4.2, C.2 and C.3. The average percentage change in the structures ranged from approximately -1.5% to 0.8%. Some of the subjects' putamen, hippocampus, pallidum, and thalamus changed more than three standard deviations from the mean, suggesting these subjects should have been examined closer before feature extraction. No such examination was performed, as none of the structures of the subjects in question, nr. 1, 2, 4, 13, 17, 20 and 24, changed more than 2% on average.

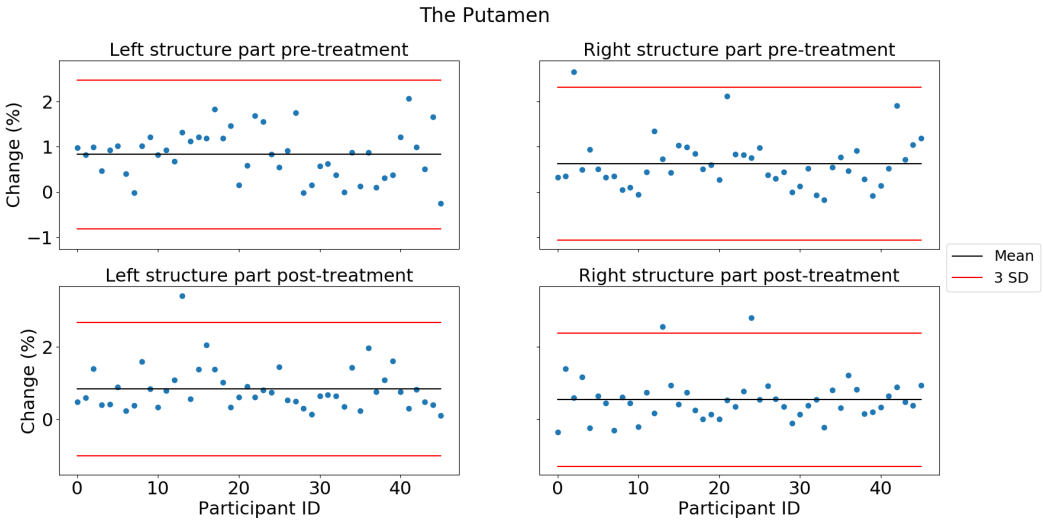


Figure 4.2: The figure illustrates the change in surface area of the left and right putamen due to artefact voxel exclusion. The mean of the percentage change of the structures (black line) and three standard deviations (red lines) were included in the plots. On average, the subjects' putamen changed by approximately 1 %. Subject nr. 13 deviated from the mean by more than three standard deviations in both post-treatment images. The change of the right putamen of subject nr. 2 pre-treatment, and the right putamen of subject nr. 24 post-treatment also both exceeded three standard deviations.

4.1.2 The Effect of Intensity Value Normalisation

To be able to compare the MR images across the patients, the intensity values were normalised as described in Chapter 3.4.3. The intensity distributions of the image stacks were plotted to assess the effect of the normalisation. Figure 4.3 displays the minimum, maximum and mean intensity values of the subjects' left pallidum before and after autoscaling. Due to the autoscaling process, the intensity statistics aligned across the image stacks. The same change was observed in the intensity distribution of the other ROIs, found in Appendix D. The image normalisation applied proved to have the desired effect.

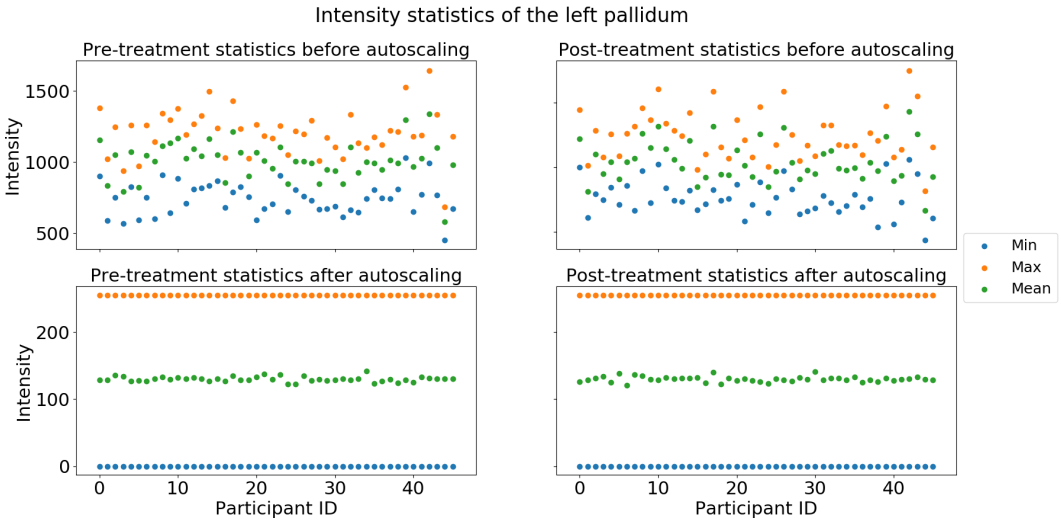
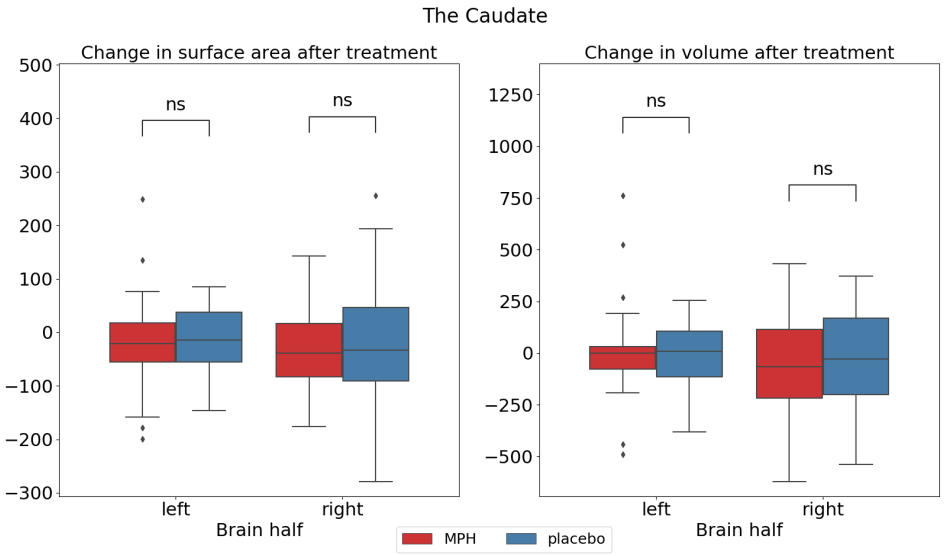


Figure 4.3: The change in intensity distribution of the left pallidum due to normalisation. The top and bottom rows display the minimum, maximum and mean intensity values of the ROI across the subjects before and after normalisation, respectively. The intensity levels of the subjects were normalised and rescaled to a common intensity level range, [0, 255]. After normalisation the maximum and minimum intensity values were equal across all subjects, and the mean approximately equal.

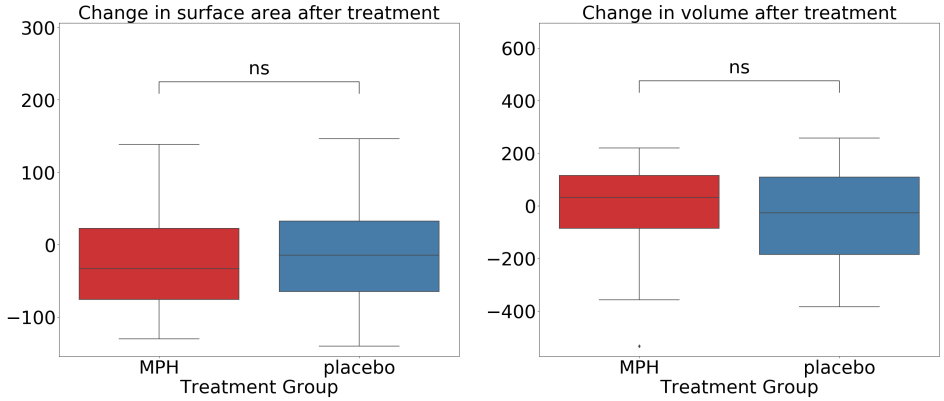
4.2 Statistical Boxplot Analysis

Literature connects treating ADHD-diagnosed patients using MPH with the volume size of the caudate [10]. Therefore, the MPH - and placebo-treated subjects' surface area and volume across the ROIs were compared using boxplots. A Welch's t-test evaluated the significance of the differences between the group's mean values. A 95% confidence interval was applied, together with the annotation explained in Table 3.3. The boxplot analysis was performed for the radiomic feature values extracted from the structures' left and right side, and for the mean values of the structures' left and right radiomic features.

No significant differences were detected in the caudates and the putamens due to treatment, as illustrated by Figure 4.5 and 4.7. The right pallidum of MPH-treated participants, however, showed signs of a significant larger surface area and volume compared to the placebo group in Figure 4.6a. The difference between the groups' surface area was also detected in Figure 4.6b. Moreover, the MPH-treated participants seem to have a significantly smaller right hippocampus volume and right thalamus surface area (see Fig. 4.5a and Fig. 4.8a).

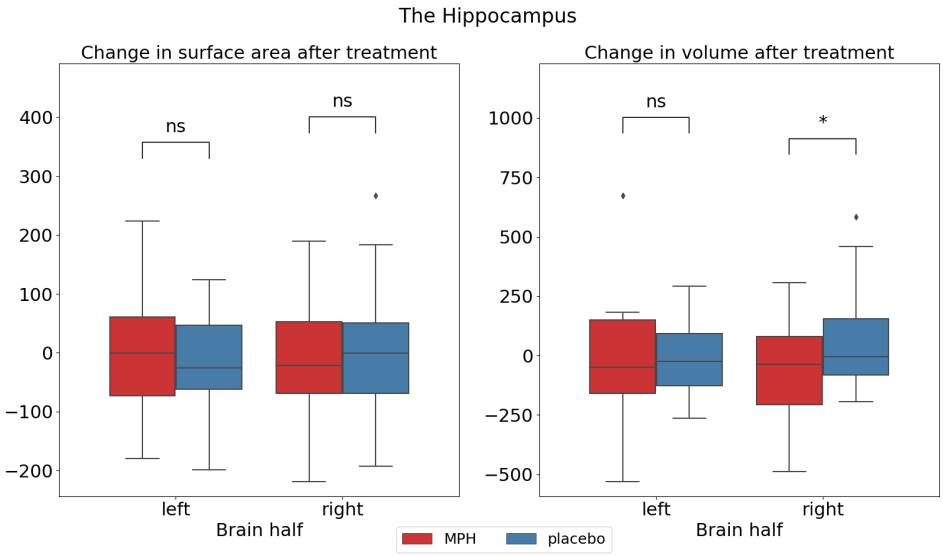


(a) Left and right brain structure separately.

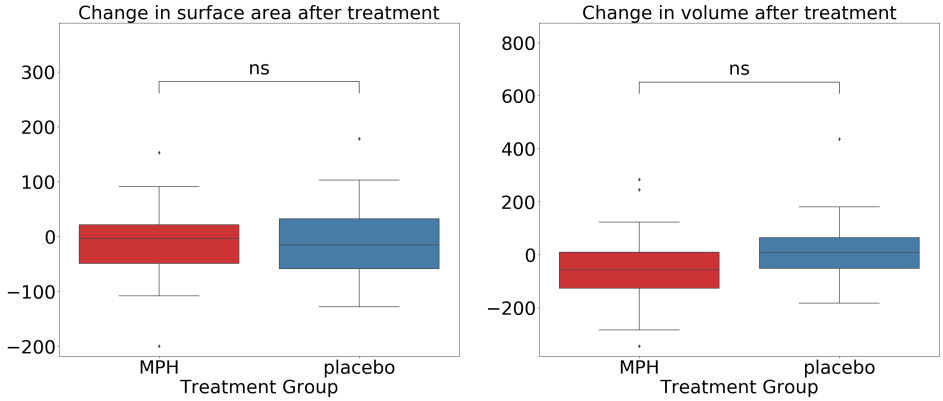


(b) Mean of the left and right side of the brain structure.

Figure 4.4: The change in surface area and volume due to treatment was compared between the treatment groups. The comparison was performed for the left and right part of the structure separately (a) and for the mean feature values of the right-sided and left-sided parts (b). No significant differences were detected between the MPH and placebo treated groups in the caudate.

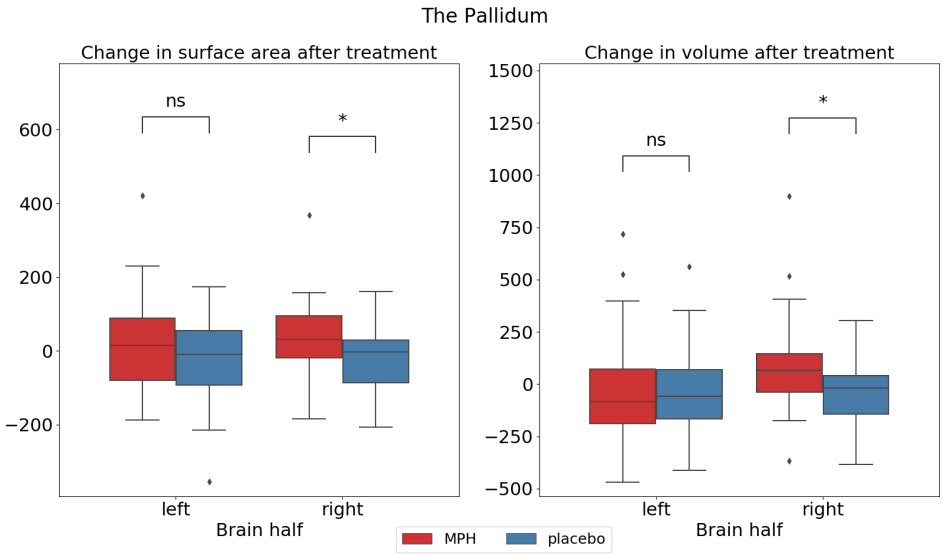


(a) Left and right brain structure separately.

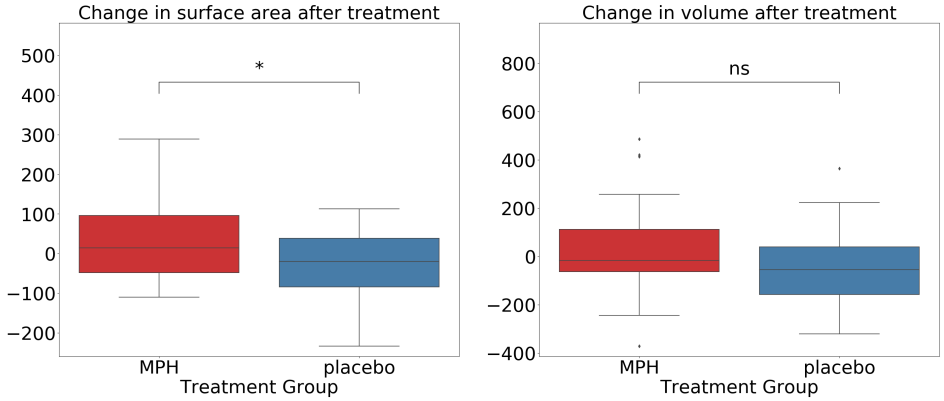


(b) Mean of the left and right side of the brain structure.

Figure 4.5: The change in surface area and volume due to treatment was compared between the treatment groups. The comparison was performed for the left and right part of the structure separately (a) and for the mean feature values of the right-sided and left-sided parts (b). A significant difference was detected in the right hippocampus' volume in subfigure (a), differentiating the treatment groups. No such difference was detected in subfigure (b).

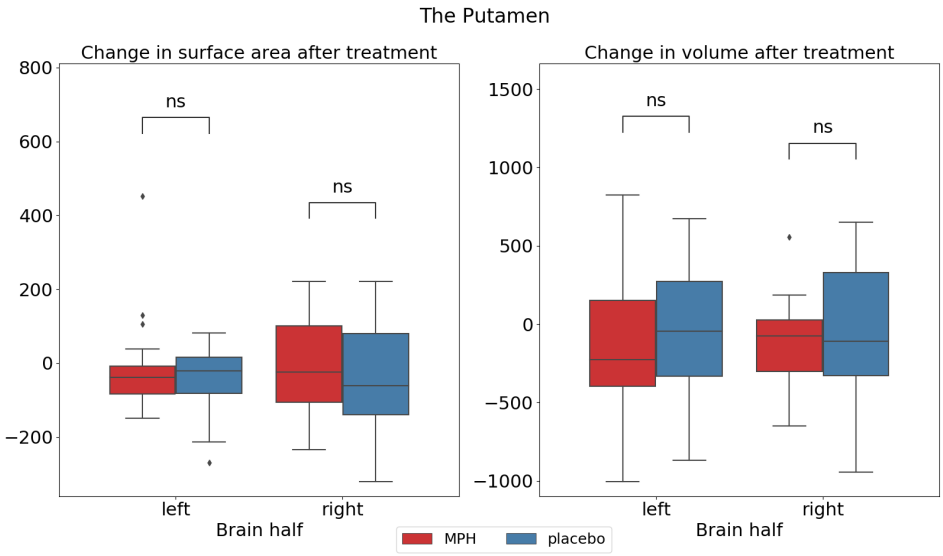


(a) Left and right brain structure separately.

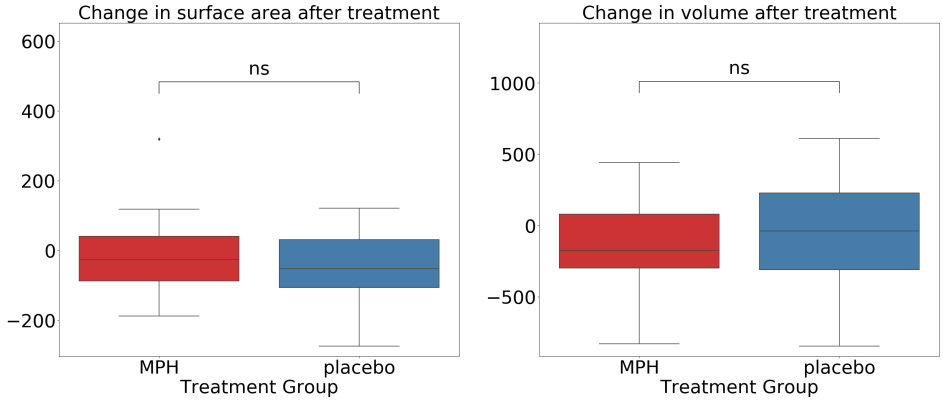


(b) Mean of the left and right side of the brain structure.

Figure 4.6: The change in surface area and volume due to treatment was compared between the treatment groups. The comparison was performed for the left and right part of the structure separately (a) and for the mean feature values of the right-sided and left-sided parts (b). Significant differences were detected in the right pallidum's surface area and volume in (a). A significant difference in the surface area was detected in (b), distinguishing the treatment groups.

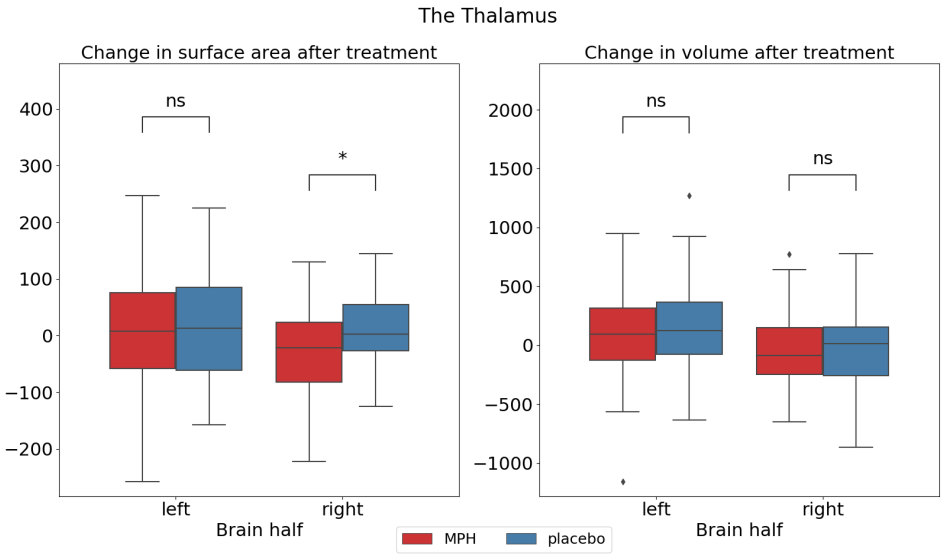


(a) Left and right brain structure separately.

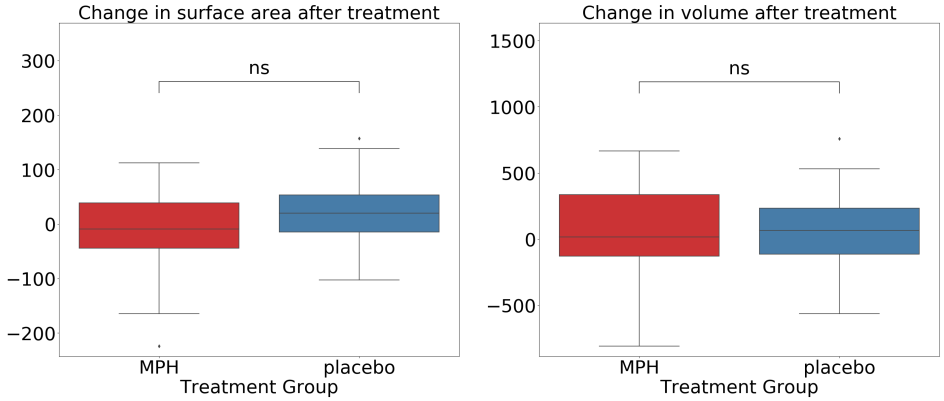


(b) Mean of the left and right side of the brain structure

Figure 4.7: The change in surface area and volume due to treatment was compared between the treatment groups. The comparison was performed for the left and right part of the structure separately (a) and for the mean feature values of the right-sided and left-sided parts (b). No significant differences were detected between the MPH and placebo treated groups in either of the boxplots.



(a) Left and right brain structure separately.



(b) Mean of the left and right side of the brain structure.

Figure 4.8: The change in surface area and volume due to treatment was compared between the treatment groups. The comparison was performed for the left and right part of the structure separately (a) and for the mean feature values of the right-sided and left-sided parts (b). A significant difference was detected in the right thalamus' surface area in subplot (a), differentiating the treatment groups. No such difference was detected in subplot (b).

4.3 The Principal Component Analyses

Principal component models were constructed based on the feature matrices of the structures, searching for patterns in the extracted radiomic features. Generating three PCA models for every bain structure enabled a search for characteristics distinguishing the treatment classes from each other. The creation of the models, together with the choice of PCs to include was performed according to Chapter 3.7.2. PLS-Toolbox recommended including three or more components for the texture-based models, with three exceptions. One component was suggested for all shape-based models. The number of principal component suggested by PLS-Toolbox and chosen for further analyses are listed in Appendix G. The model statistics inspected can be found in Appendix F.

Score plots were created based on the PCA models and visually inspected. If a model had captured distinct patterns differentiating class 1 from class 0, a separation of the classes would be visible. Following the inspection, individuals and clusters of subjects standing out in the score plots were identified, as explained in Chapter 3.7.2 and listed in Appendix H.

By comparing the subjects standing out in the score plots and listed in Table H.1 to the models' corresponding Q - T plots, possible outliers were detected. Multiple subjects could be categorised as outliers according to the plots. However, only subjects standing out in all three PCA models of a brain structure were considered when compared to the Q - T plots. A confidence limit of 95% was set, marking samples with high Q-residuals or Hotelling's T^2. According to the Q - T plots, the score plots captured between 89.31% and 68.14% of the existing variance. The remaining variance were captured by the residuals.

The PCA models based on the caudate were analysed and the findings presented in this Chapter, including a general summary of the remaining structures. A closer examination of the remaining structures, their RMSE curves, score plots and Q - T plots can be found in Appendix E.

4.3.1 Analysing the PCA-models of the Caudate

PLS-Toolbox recommended using one principal component when creating the shape feature PCA model. The same recommendation was given for the remaining structures' shape feature-based PCA model. According to the model statistics given in Table F.1, one PC would cover approximately 18% of the cumulative explained variance. Ensuring that a cumulative explained variance of 80% is secured, the eighth component was chosen as "cutoff"-component. The eighth component covered approximately 1.4 features, according to the eigenvalues. In order to ensure that the remaining shape feature-based models also covered close to 80% explained variance, more than one principal component was included in further analyses by each model. The number of principal components included in further analyses of the shape feature-based models can be found in Table G.1 in Appendix G.

The RMSECV and RMSEC of the shape feature model are illustrated in Figure 4.9. As the RMSEC does not flatten out, it just keeps collecting more information and noise from the dataset. The lack of decrease in the RMSECV curve in addition to the steady decreasing RMSEC suggests the model overfits strongly. Therefore, PLS-Toolbox recommended including as few PCs as possible. Still, eight components were chosen, as patterns distinguishing the group could be detected in PCs explaining less variance. The same assessment was made when deciding upon PCs for the remaining structures' shape feature models. The RMSE curves of the shape feature-based models of the remaining structures behaved similar to the caudate's curves, and indicated strong overfitting in these models as well, as seen in Figures E.1 and E.12, E.25, E.37.

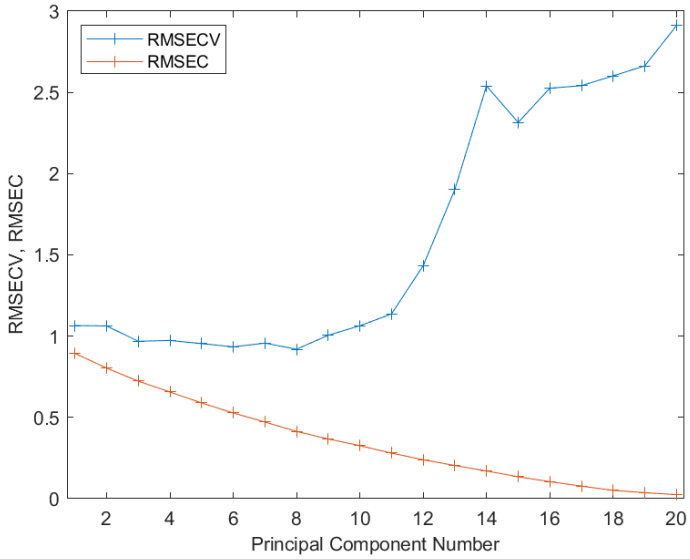


Figure 4.9: The RMSEC and RMSECV curve of the caudate's shape feature model. While the RMSEC curve decreased constantly, the RMSECV increased after eight components, indicating the model overfited stronlgy.

In contrast to the shape feature model's RMSECV, the texture feature models' RMSECV decreased continually, as displayed in Figures 4.10 and 4.11. The RMSE curves of the remaining structures' texture feature-based models behaved similar. While the RMSEC of the model steadily decreased, the RMSECV flattened out. The behaviour of the curves suggests that overfitting also occured in the texture feature models. Following the recommendation of PLS-Toolbox, nine components were included in the 128-bin texture feature model. Seven components were included in the 64-bin texture feature model, based on the cumulative explained variance.

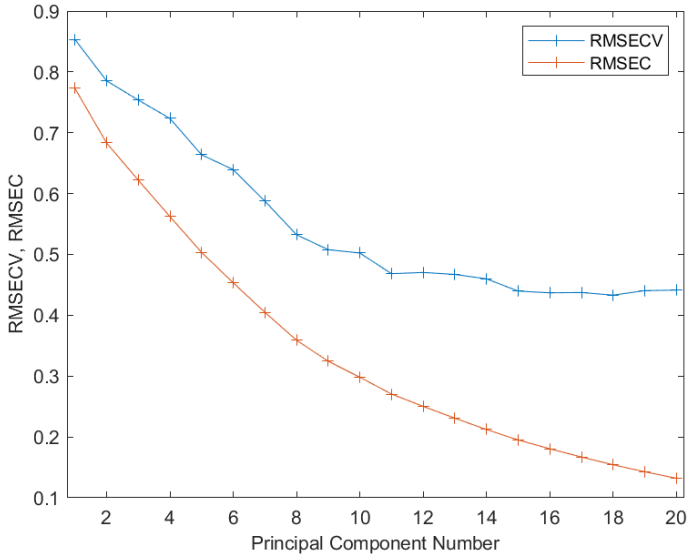


Figure 4.10: The RMSEC and RMSECV curve of the caudate's 128-bin texture feature model. Both curves decreased constantly, until the RMSECV curve flattened out slightly.

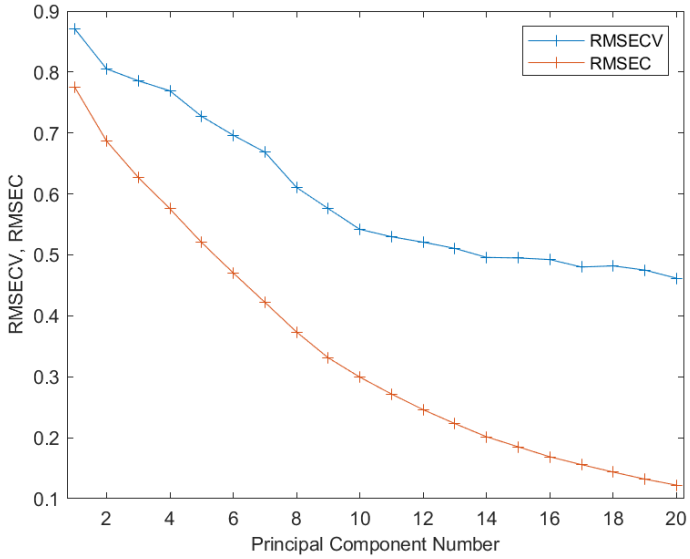


Figure 4.11: RMSEC and RMSECV curve of the caudate's 64-bin texture feature model.

Pattern Detection in the Score Plots

No visible patterns discriminating the MPH-treated subjects from the placebo group were detected in the caudates' score plots in Figures 4.12, 4.13 - 4.17. The same applied for the remaining structures, whose score plots were depicted in Appendix E.

The Shape Feature-based PCA model

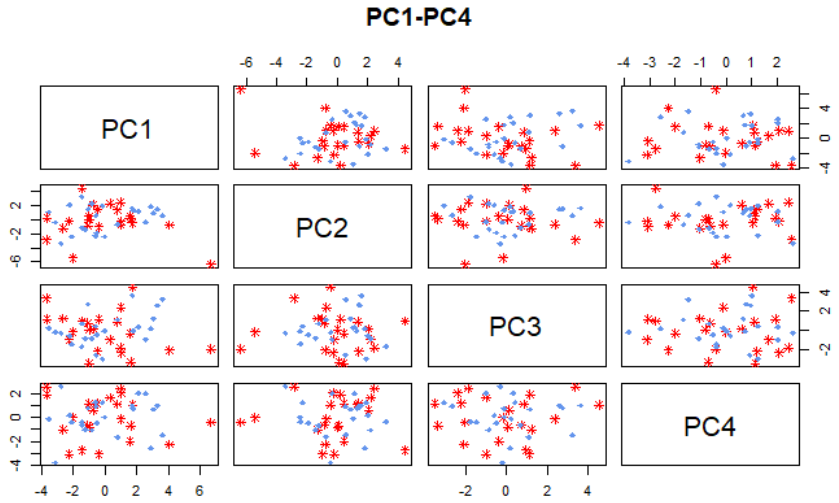


Figure 4.12: The caudate's pair plot of the shape feature principal components nr 1-4. No principal components discriminating the treatment groups were detected.

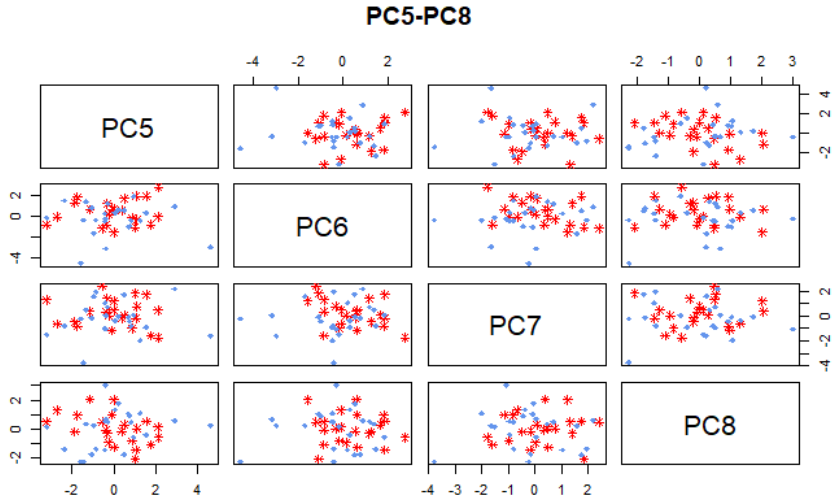


Figure 4.13: The caudate's pair plot of the shape feature principal components nr 5-8. No principal components discriminating the treatment groups were detected.

The 128-bin Texture Feature-based PCA model

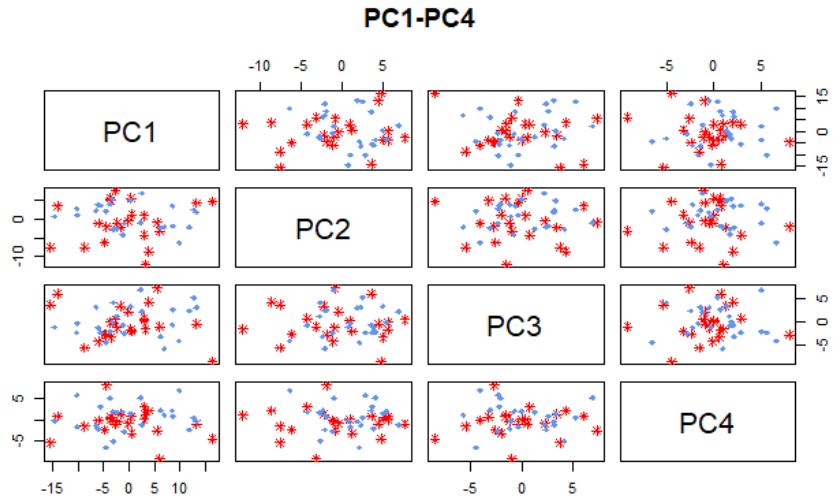


Figure 4.14: The caudate's pair plot of the 128-bin texture feature principal components nr 1-4. No principal components discriminating the treatment groups were detected.

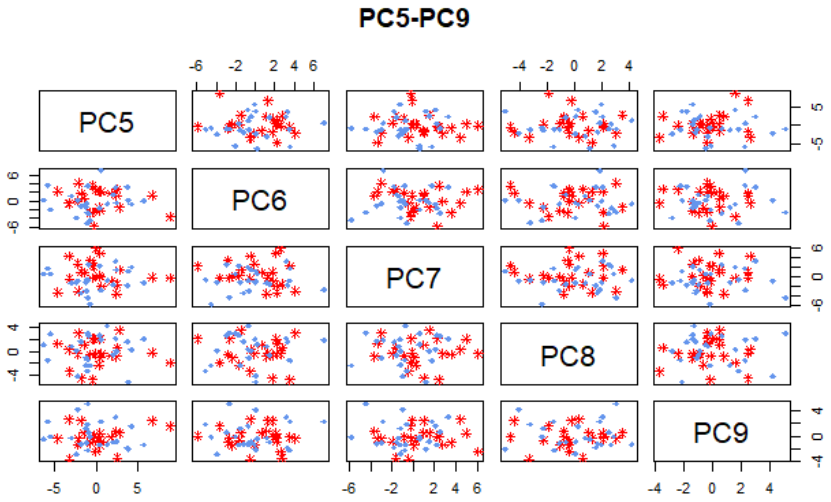


Figure 4.15: The caudate's pair plot of the 128-bin texture feature principal components nr 5-9. No principal components discriminating the treatment groups were detected.

The 64-bin Texture Feature-based PCA model

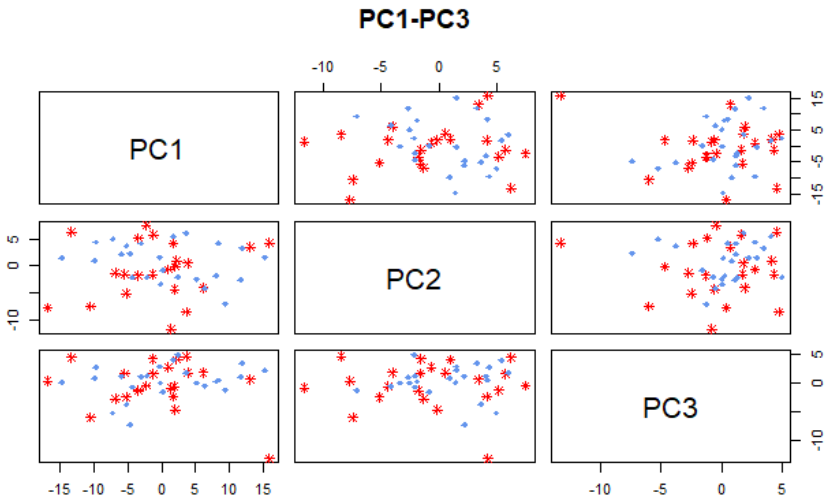


Figure 4.16: The caudate's pair plot of the 64-bin texture feature principal components nr 1-4. No principal components discriminating the treatment groups were detected.

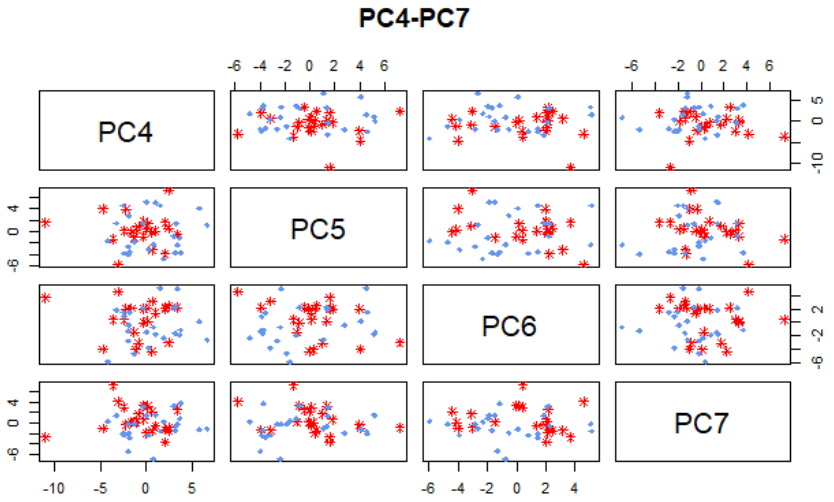


Figure 4.17: The caudate's pair plot of the 64-bin texture feature principal components nr 5-9. No principal components discriminating the treatment groups were detected.

Identifying Possible Outliers in the Score and Q-T Plots

Seven MPH-treated and nine placebo-treated subjects were identified as possible outliers in the shape feature model. Further, ten MPH-treated subjects were singled out in each texture feature. In both texture feature models six placebo-treated participants stood out. Participants nr. 2, 4, 13, 31 and 43 were detected in all three caudate models. Subjects nr. 2 and 13 are also characterised as possible outliers by the Q - T plots of the 128-bin and 64-bin texture feature model in Figure 4.19 and 4.20, respectively.

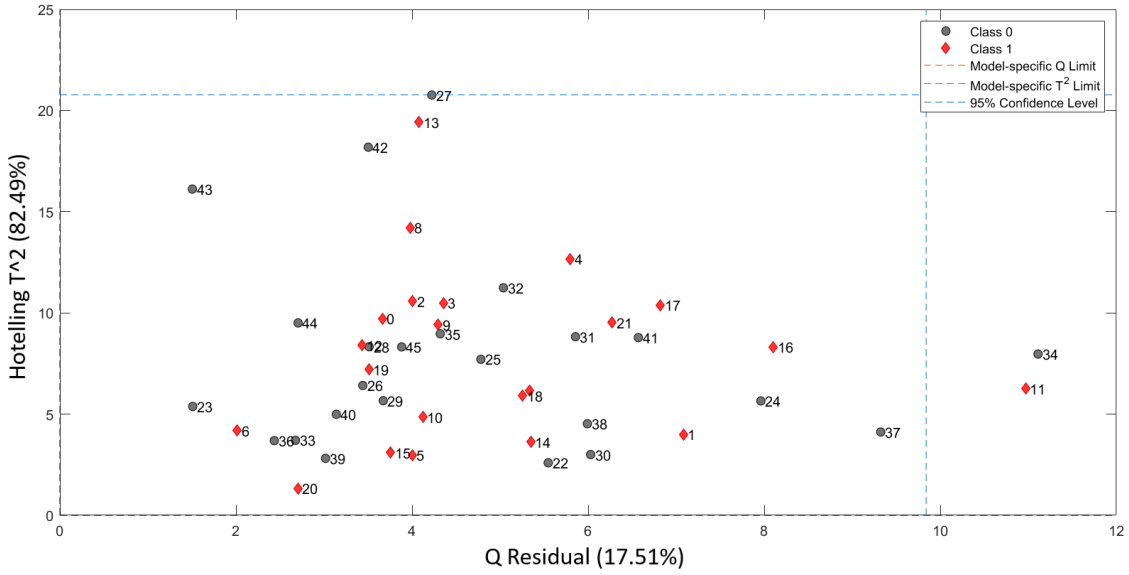


Figure 4.18: The Q-T plot of the caudate's shape feature model. The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

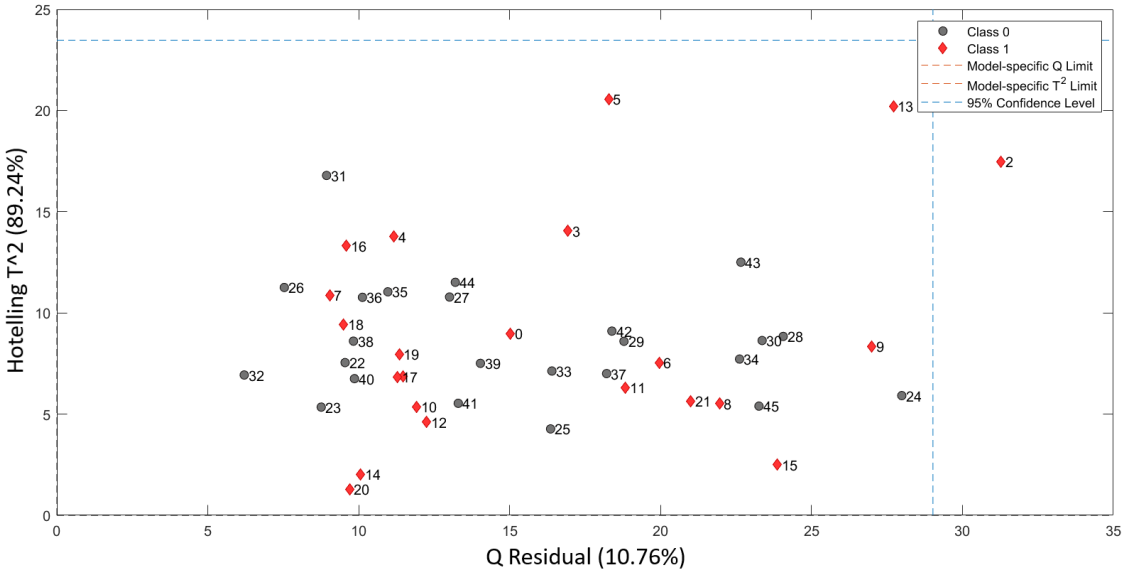


Figure 4.19: The Q-T plot of the caudate's 128-bin texture feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

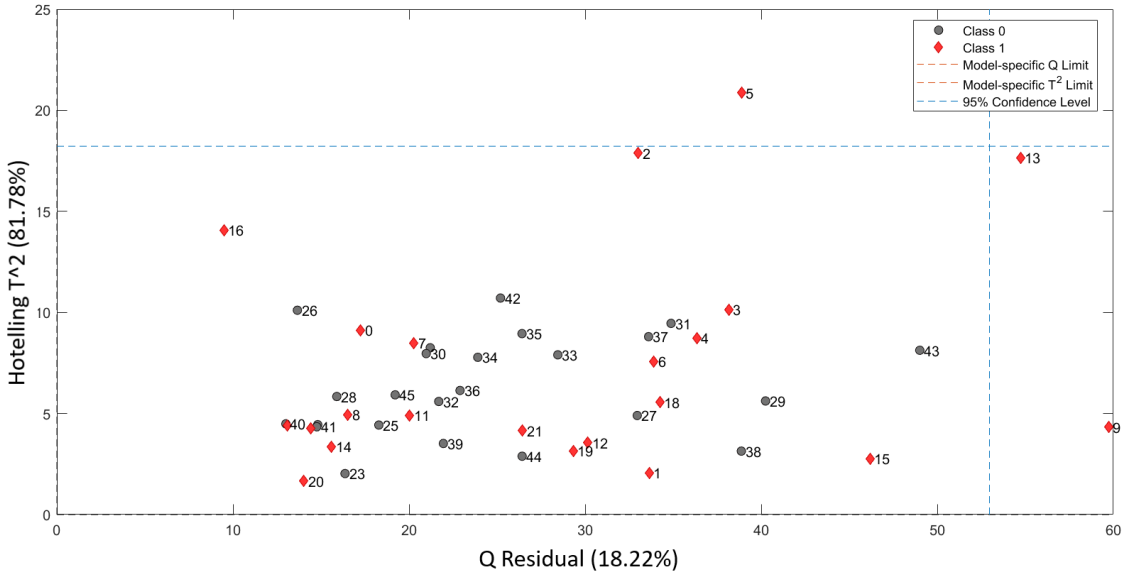


Figure 4.20: The Q-T plot of the caudate's 64-bin texture feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

4.4 Predicting the Treatment Method

For each structure, two classification experiments were performed, predicting the treatment method based on the shape features, 64-bin texture features and 128-bin

texture features. The first experiment was based upon the radiomic features of the left-sided and right-sided structure separately, resulting in 178 features analysed for each structure. The second experiment predicted the subjects' treatment group based on the mean values of the left- and right-sided structures' features, halving the number of features compared to the first experiment performed.

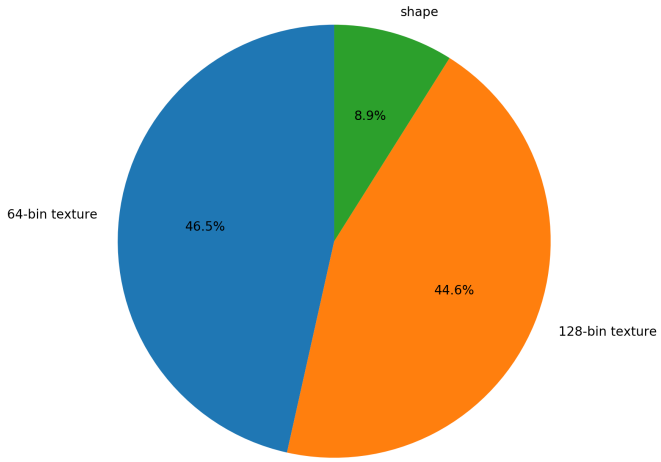
As described in Chapter 3.7.3, the subjects were labelled according to their treatment group as either "MPH" or "placebo". In each experiment, the class labels of the subjects were predicted through thirty classifications, one for each classification and feature selection algorithm combination available. The applied classification and feature selection algorithms are listed in Chapter 3.7.3. Table 3.4 contains the abbreviations of the classifier algorithms used in this chapter.

The AUC evaluating the experiments were illustrated through heatmaps. Pie diagrams were created to examine the connection between the AUC and the features selected by the feature selection algorithms by each experiment. The distributions of the feature categories were weighted with respect to the number of times features were chosen. Moreover, the selection rate of the five most frequent chosen variables based on the entire experiment was illustrated.

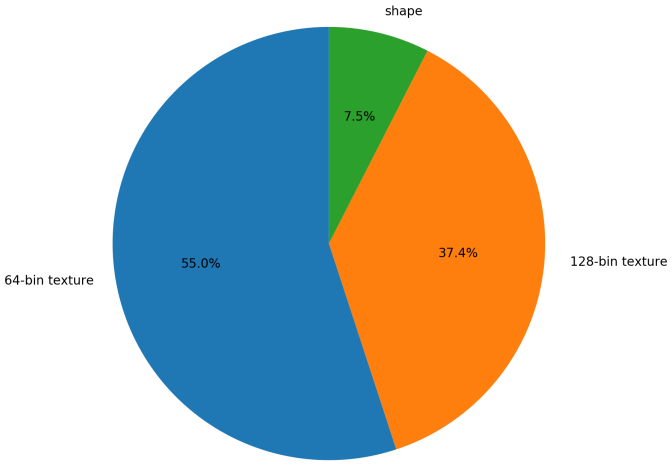
4.4.1 The Caudate

According to Figures 4.21a and 4.21b, approximately 40% of the features selected by the classification experiments were categorised as 128-bin texture features. The same figure yielded that approximately 50% of the chosen features were 64-bin texture features, and the remaining 10% shape features. The distribution between the feature categories could be simplified to a 5:4:1 relationship, and differed just slightly between the two caudate experiments. The dominance of the texture features was also reflected by the remaining structures' classification experiments.

Figure 4.22 illustrated the five most selected features by the classification models. According to Figure 4.22a, 64-bin texture features extracted from the right caudate dominated when performing the classifications with feature selection. This was confirmed by Figure 4.22b, which illustrated the five features selected most often by the mean-based classification experiment. Only 64-bin texture features were included in the figure, although these features were partly different features than highlighted in Figure 4.22a.

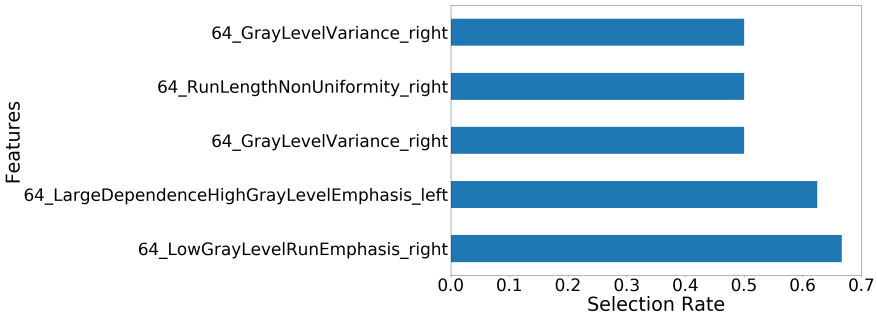


(a) The selected features distribution based upon the features of the left- and right sided ROIs separately.

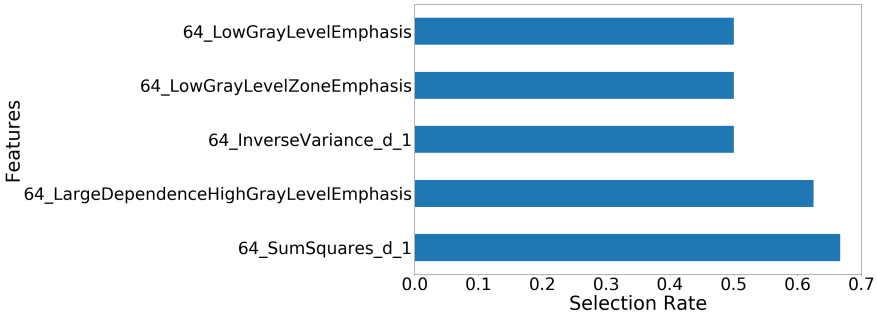


(b) The selected features distribution based upon the mean value of the left- and right sided ROI's features.

Figure 4.21: The figure illustrates the overall distribution of the features selected by the caudate's classification experiments. The features were sorted by their feature categories; shape, 128-bin texture and 64-bin texture features. The pie chart in subfigure (a) displays the selected features distribution from the experiment considering the right- and left-sided ROI's separately. Subfigure (b) illustrates the distribution of the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.



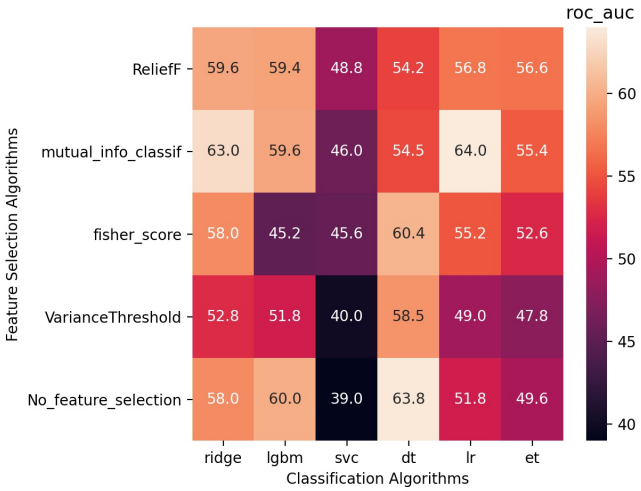
(a) The five most selected features based upon the features of the left- and right sided ROIs separately.



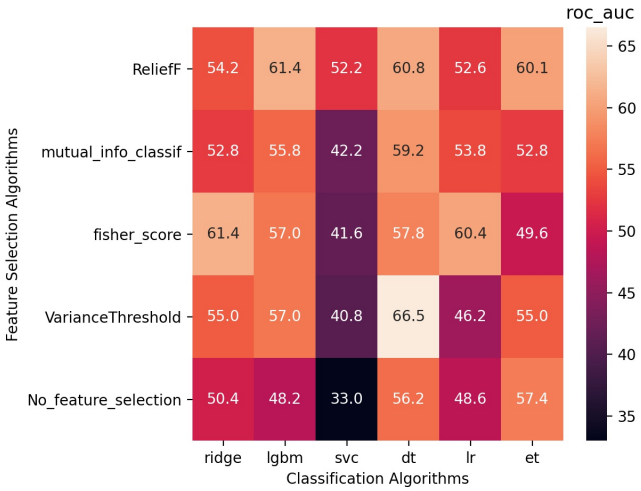
(b) The five most selected features based upon the mean value of the left- and right sided ROI's features.

Figure 4.22: The figure illustrates the selection rate of the five most selected features by the caudate's classification experiments. Only 64-bin texture features were included in the figures. Subfigure (a) displays the selected features from the experiment considering the right- and left-sided ROI's separately, while subfigure (b) illustrates the distribution of the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.

The heatmaps in Figure 4.23 both present AUC scores ranging approximately from 33 to 67, reflecting a considerable variation in prediction precision depending on the choice of feature selection and classification algorithm. The Support Vector Classifier (SVC) yields on average the lowest AUCs amongst the classification algorithms in both experiments. In Figure 4.23a the highest scores were achieved by the Ridge Classification, the Decision Tree classifier (DT) and the Logistic Regression algorithm (LR). Especially the Mutual Information Classifier proved to return good results in combination with the Ridge classifier and LR classifier when considering the right- and left-sided ROIs separately, yielding an AUC of 63 and 64 respectively. When combining the left- and right-sided ROIs, the decision tree classifier still performed best amongst the classification algorithms according to Figure 4.23b. The ReliefF feature selector was the only feature selection algorithm to not return AUC scores below 50 in Figure 4.23b, independent of the classification algorithm applied.



(a) The AUC scores from the prediction models based upon the features of the left- and right sided ROIs separately.



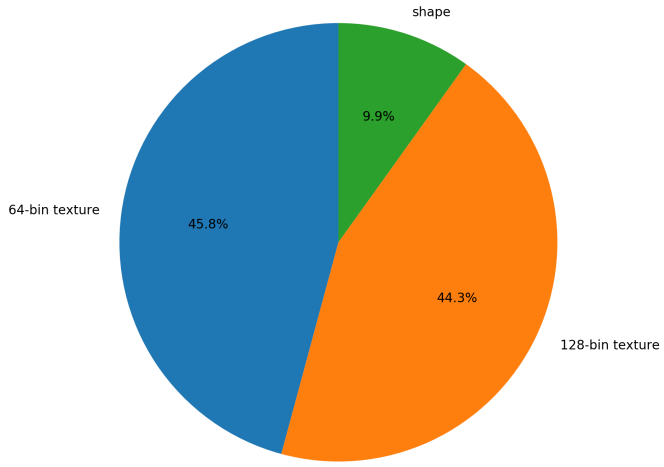
(b) The AUC scores from the prediction models based upon the mean value of the left- and right sided ROI's features.

Figure 4.23: The average AUC scores achieved by the thirty prediction models generated by the caudate's classification experiments. Four feature selectors were implemented together with six classification algorithms. In addition, classifications using the six algorithms was performed without feature selection. Subfigure (a) displays the AUC scores from the experiment considering the right- and left-sided ROI's separately, while subfigure (b) illustrates the scores from the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.

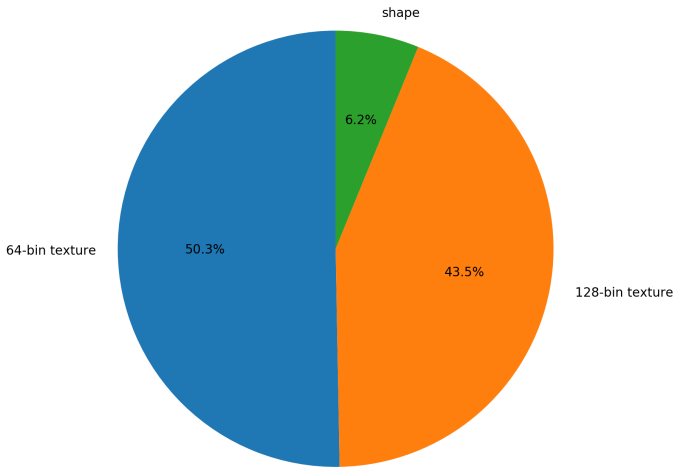
4.4.2 The Hippocampus

Similar to the caudet's experiments, the texture feature categories dominated the feature selection and were selected an approximately equal amount of times according to Figure 4.24. Predicting the treatment groups based on the mean of the left and right ROI's features resulted in a slight decrease in the selection of shape features, compared to the first experiment. This reduction The reduced was reflected by the selection rate Figure 4.25, were 4.25a included two shape features and 4.25b none.

In contrast to the distribution of the selected features presented in Figure 4.24, two out of the five features listed in Figure 4.25a were shape features. In addition, two 128-bin texture features and one 64-bin texture feature were included in the figure. All five features in Figure 4.25a were extracted from the right hippocampus of the subjects.

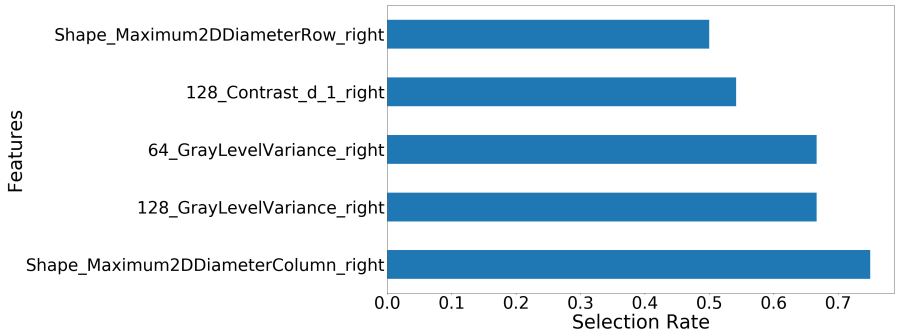


(a) The selected features distribution based upon the features of the left- and right sided ROIs separately.

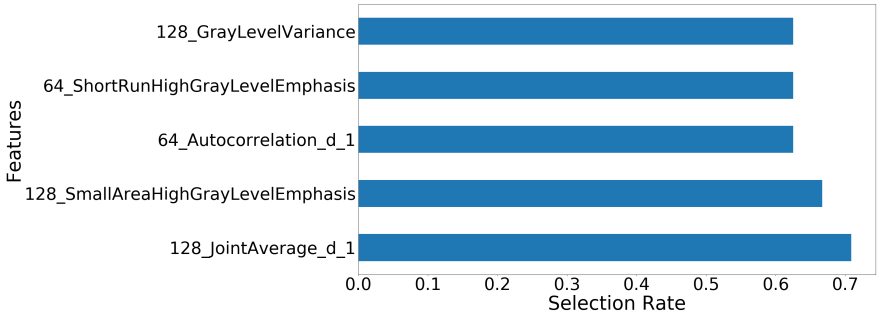


(b) The selected features distribution based upon the mean value of the left- and right sided ROI's features.

Figure 4.24: The figure illustrates the overall distribution of the features selected by the hippocampus' classification experiments. The features were sorted by their feature categories; shape, 128-bin texture and 64-bin texture features. The pie chart in subfigure (a) displays the selected features distribution from the experiment considering the right- and left-sided ROI's separately. Subfigure (b) illustrates the distribution of the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.



(a) The five most selected features based upon the features of the left- and right sided ROIs separately.

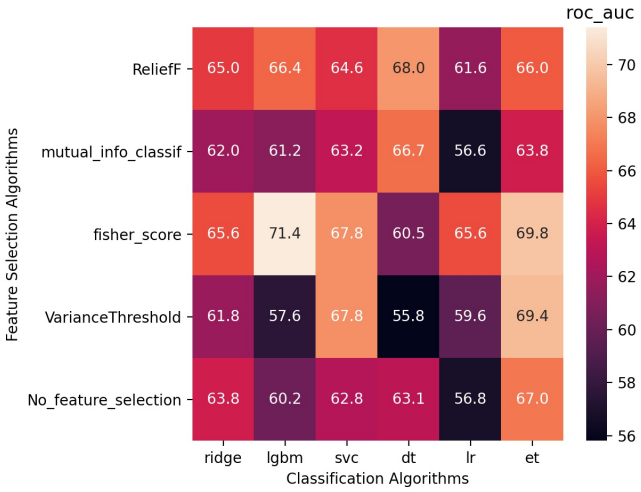


(b) The five most selected features based upon the mean value of the left- and right sided ROI's features.

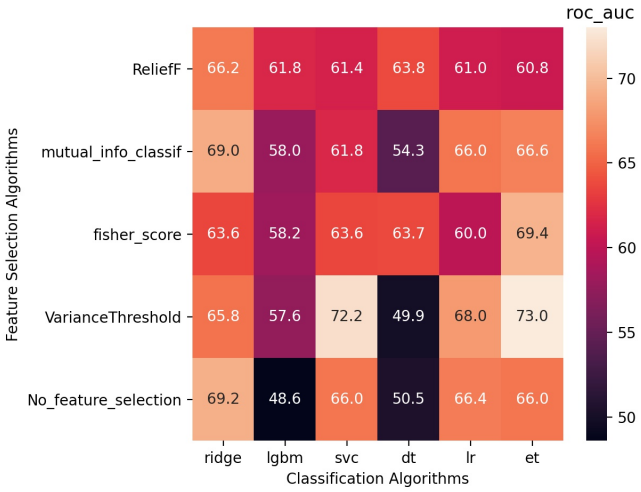
Figure 4.25: The figure illustrates the selection rate of the five most selected features by the hippocampus' classification experiments. The name of the features indicate the feature type, 128 corresponding to 128-bin texture features and 64 to 64-bin texture features. Subfigure (a) displays the selected features from the experiment considering the right- and left-sided ROI's separately, while subfigure (b) illustrates the distribution of the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.

Figure 4.26 illustrates the heatmap of the hippocampus' experiments. The AUC scores in Figure 4.26a varied from approximately 56 to 71, in contrast to the larger range observed in 4.26b (from 49 to 73). The Fisher score and ReliefF algorithm were the only feature selection algorithms achieving results solely above 60, with one exception in Figure 4.26b. The same applies to the Ridge Classifier, the SVC and the Extremely Randomised Tree Classification (ET) algorithm.

Although the Light Gradient Boosting Machine (LGBM) algorithm yielded the highest AUC of the experiment when combined with the fischer score in 4.26a (AUC $\approx$ 71), the classifier returned the lowest AUC scores on average in both heatmaps in 4.26. In Figure 4.26b the highest AUC scores were achieved by combining the Variance Threshold selector with either the SVC or the ET algorithm. These combinations achieved AUC scores of approximately 72 and 73, respectively.



(a) The AUC scores from the prediction models based upon the features of the left- and right sided ROIs separately.



(b) The AUC scores from the prediction models based upon the mean value of the left- and right sided ROI's features.

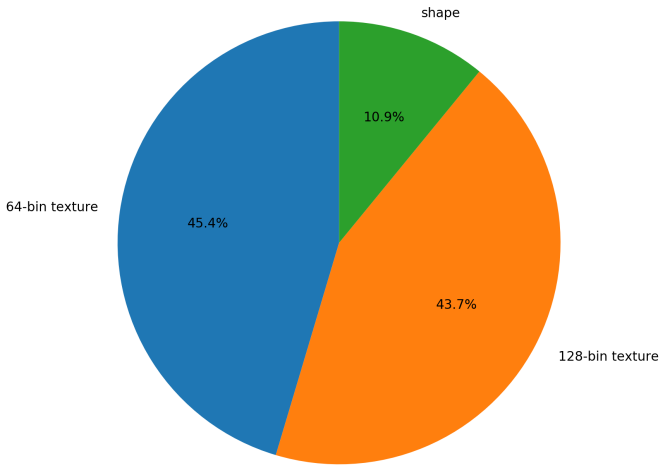
Figure 4.26: The average AUC scores achieved by the thirty prediction models generated by the hippocampus' classification experiments. Four feature selectors were implemented together with six classification algorithms. In addition, classifications using the six algorithms was performed without feature selection. Subfigure (a) displays the AUC scores from the experiment considering the right- and left-sided ROI's separately, while subfigure (b) illustrates the scores from the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.

4.4.3 The Pallidum

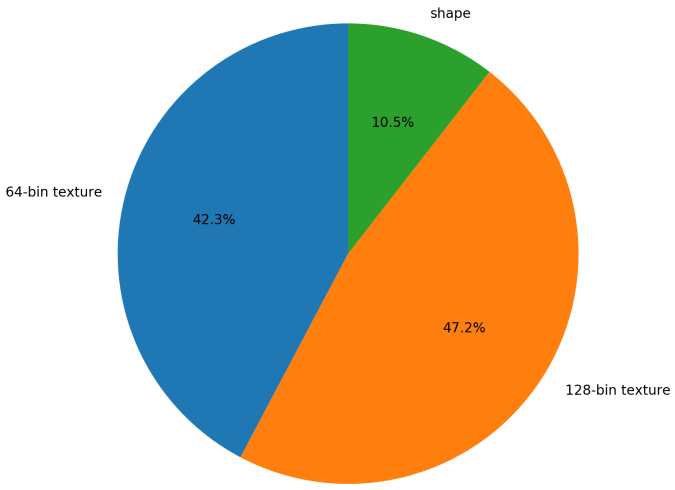
The distributions of the selected features illustrated in Figure 4.27 did not change noticeably across the experiments, and resembled the distribution of the other structures' experiments with approximately 40% covering 128-bin texture features, approximately 40% covering 64-bin texture features and 10% shape features.

According to Figure 4.28a, illustrating the five most selected features by the experiment considering the left- and right-sided ROI separately, the shape features

included were based upon the right pallidum of the trial subjects. The remaining features listed in Figure 4.28a belonged to the texture feature categories and were extracted from the left pallidum. While the selected features in Figure 4.28a were distributed across the 64-bin texture features, the 128-bin texture features and the shape features, Figure 4.28b reflected an increased amount of selected 128-bin texture features. One of the shape features, Maximum2DDiameterRow, and two 64-bin texture features, ClusterTendency_d_1 and SizeZoneNonUniformity, were replaced by 128-bin texture features.

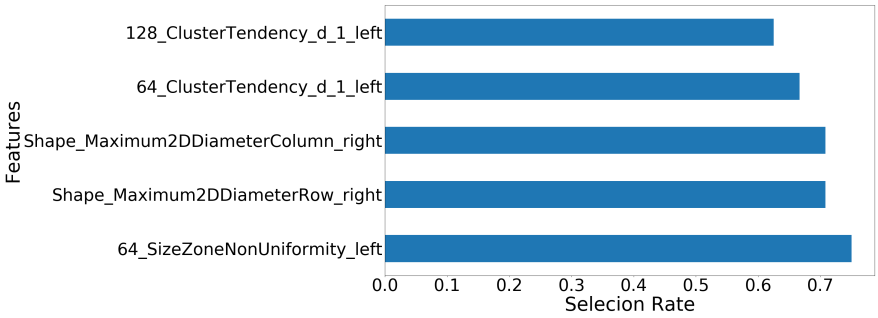


(a) The selected features distribution based upon the features of the left- and right sided ROIs separately.

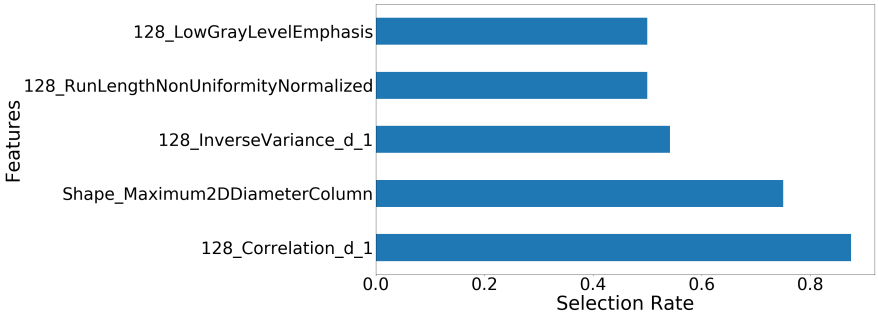


(b) The selected features distribution based upon the mean value of the left- and right sided ROI's features.

Figure 4.27: The figure illustrates the overall distribution of the features selected by the pallidum's classification experiments. The features were sorted by their feature categories; shape, 128-bin texture and 64-bin texture features. The pie chart in subfigure (a) displays the selected features distribution from the experiment considering the right- and left-sided ROI's separately. Subfigure (b) illustrates the distribution of the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.



(a) The five most selected features based upon the features of the left- and right sided ROIs separately.

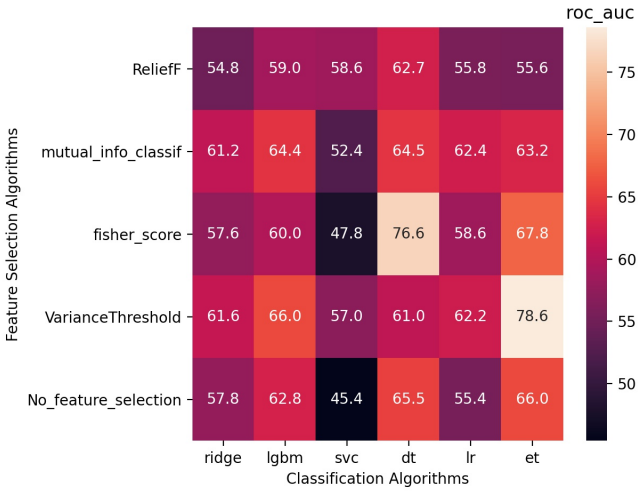


(b) The five most selected features based upon the mean value of the left- and right sided ROI's features.

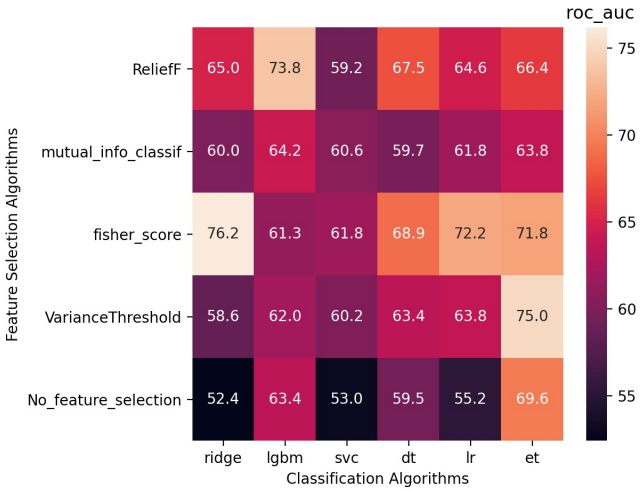
Figure 4.28: The figure illustrates the selection rate of the five most selected features by the pallidum's classification experiments. The name of the features indicate the feature type, shape corresponding to shape features and 128 to 128-bin texture features. Subfigure (a) displays the selected features from the experiment considering the right- and left-sided ROI's separately, while subfigure (b) illustrates the distribution of the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.

The AUC scores of the experiment considering the left- and right ROI separately, illustrated by the heatmap in Figure 4.29a, varied between approximately 45 and 79. The AUCs of the mean-based experiment in Figure 4.29b were, however, distributed across a smaller range from approximately 53 to 76. The highest scores identified in the first experiment were achieved by combining the Fisher Score with the DT algorithm (AUC $\approx$ 77) or the Variance Threshold selector with the ET algorithm (AUC $\approx$ 78) (see Figure 4.29a).

The Fisher score also achieved relative high AUC scores in the second experiment in Figure 4.29b, especially combined with the Ridge classifier, the LR algorithm and the ET algorithm. Similar to Figure 4.29a, the combination of the Variance Threshold algorithm and the ET classifier also yielded an AUC score above 70, even though the Variance Threshold on average did return AUC scores below 65 in both experiments.



(a) The AUC scores from the prediction models based upon the features of the left- and right sided ROIs separately.



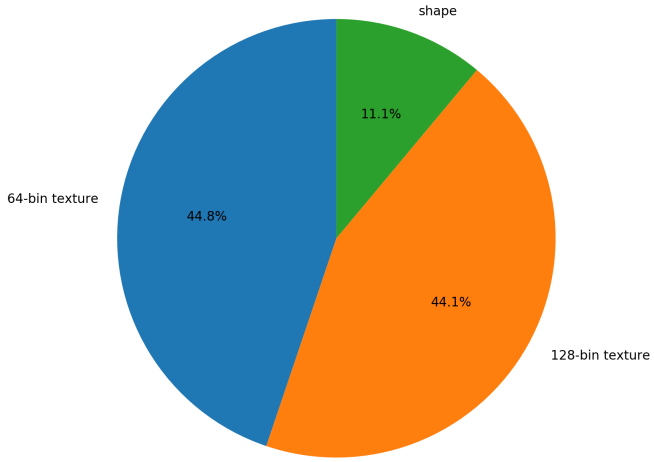
(b) The AUC scores from the prediction models based upon the mean value of the left- and right sided ROI's features.

Figure 4.29: The average AUC scores achieved by the thirty prediction models generated by the pallidum's classification experiments. Four feature selectors were implemented together with six classification algorithms. In addition, classifications using the six algorithms was performed without feature selection. Subfigure (a) displays the AUC scores from the experiment considering the right- and left-sided ROI's separately, while subfigure (b) illustrates the scores from the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.

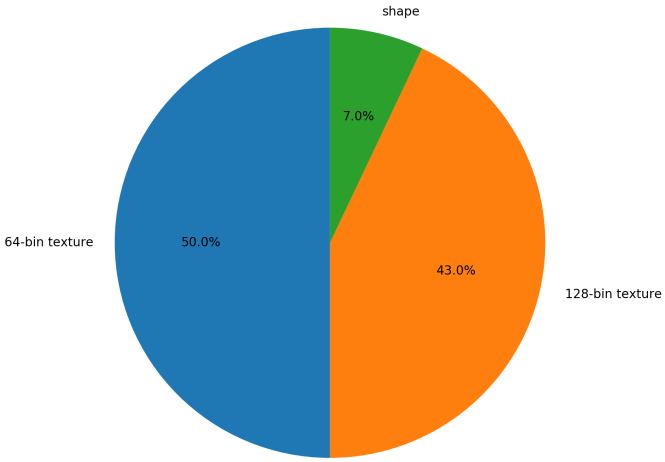
4.4.4 The Putamen

Figure 4.30 illustrates the distribution of the selected features in each classification experiment across the feature categories. According to Figure 4.30a, approximately 45% of the features selected by the feature selection algorithms were categorised as 128-bin texture features. An equal amount of 64-bin texture features and approximate 10% shape features were selected. The distribution only changed slightly by considering the mean of the right- and left-sided ROIs' features, according to Figure 4.30b.

Further, the five features with the highest selection rate were listed together with the corresponding selection rate in Figure 4.31 for each of the experiments. According to Figure 4.31a, one shape feature, two 128-bin texture features and two 64-bin texture features were highlighted by the experiment based on the left and right ROIs separately. The mean based experiment (see Figure 4.31b) highlighted two and three 64-bin and 128-bin texture features, respectively, and none of the same as in the first experiment. It appears from both the pie charts in Figure 4.30 and the selection rates in Figure 4.31, that texture features dominated the classification experiments.

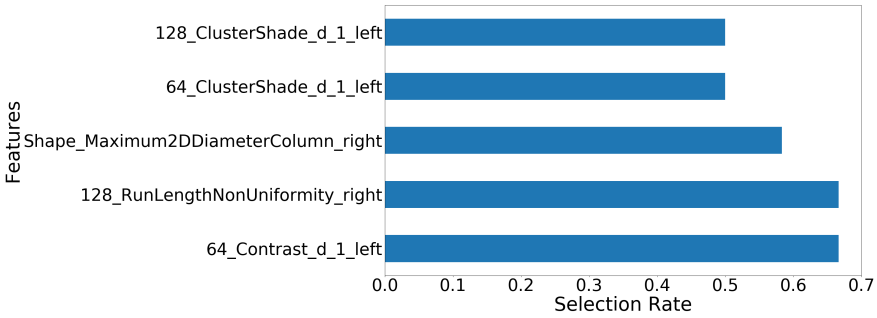


(a) The selected features distribution based upon the features of the left- and right sided ROIs separately.

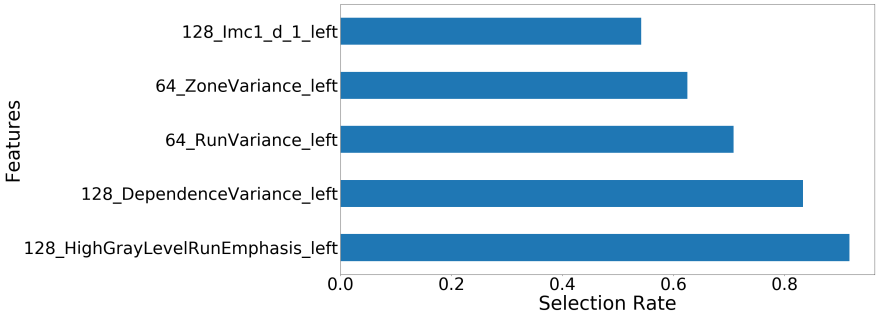


(b) The selected features distribution based upon the mean value of the left- and right sided ROI's features.

Figure 4.30: The figure illustrates the overall distribution of the features selected by the putamen's classification experiments. The features were sorted by their feature categories; shape, 128-bin texture and 64-bin texture features. The pie chart in subfigure (a) displays the selected features distribution from the experiment considering the right- and left-sided ROI's separately. Subfigure (b) illustrates the distribution of the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.



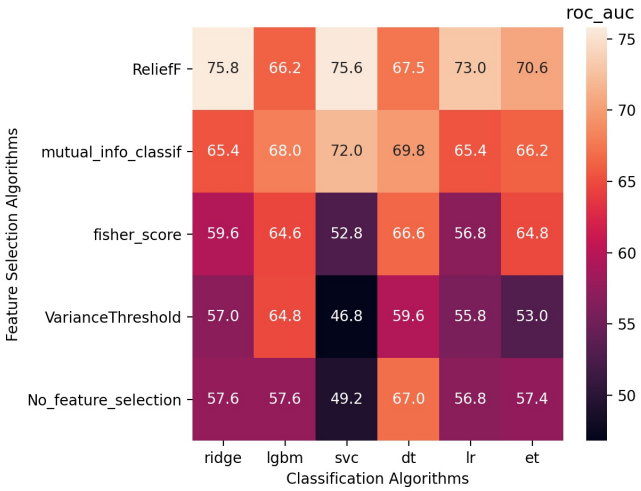
(a) The five most selected features based upon the features of the left- and right sided ROIs separately.



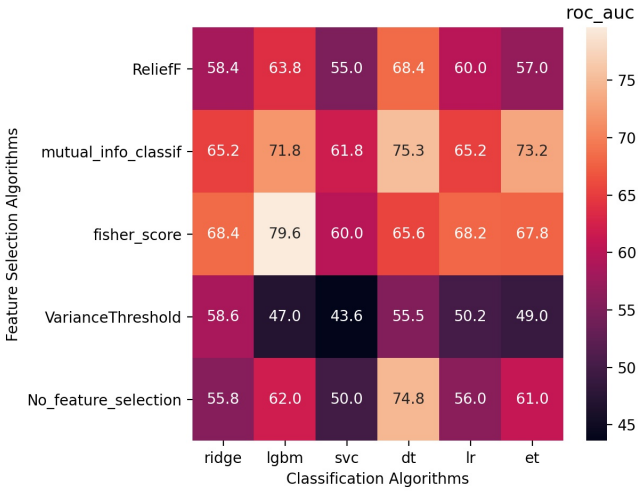
(b) The five most selected features based upon the mean value of the left- and right sided ROI's features.

Figure 4.31: The figure illustrates the selection rate of the five most selected features by the putamen's classification experiments. The name of the features indicate the feature type, shape corresponding to shape features, 128 to 128-bin texture features and 64 to 64-bin texture features. Subfigure (a) displays the selected features from the experiment considering the right- and left-sided ROI's separately, while subfigure (b) illustrates the distribution of the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.

The heatmaps illustrating the performance of the classification models were illustrated in Figure 4.32. The AUCs achieved by the experiments' models ranged from approximately 44 to 80. Overall, the feature selector Mutual Information Classifier performed best in both experiments, regardless of the classification algorithm chosen. The highest AUC scores in Figures 4.32a and 4.32b were achieved by the feature selectors ReliefF and the Fisher Score, respectively. The performance of the models appeared to be connected to feature selectors, rather than the classification algorithms in both experiments.



(a) The AUC scores from the prediction models based upon the features of the left- and right sided ROIs separately.



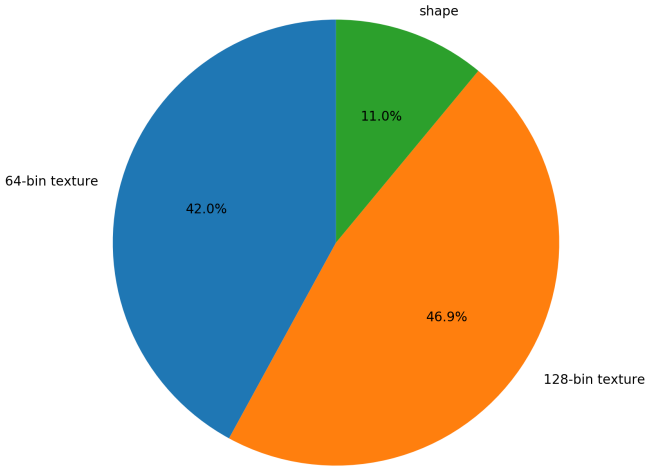
(b) The AUC scores from the prediction models based upon the mean value of the left- and right sided ROI's features.

Figure 4.32: The average AUC scores achieved by the thirty prediction models generated by the putamen's classification experiments. Four feature selectors were implemented together with six classification algorithms. In addition, classifications using the six algorithms was performed without feature selection. Subfigure (a) displays the AUC scores from the experiment considering the right- and left-sided ROI's separately, while subfigure (b) illustrates the scores from the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.

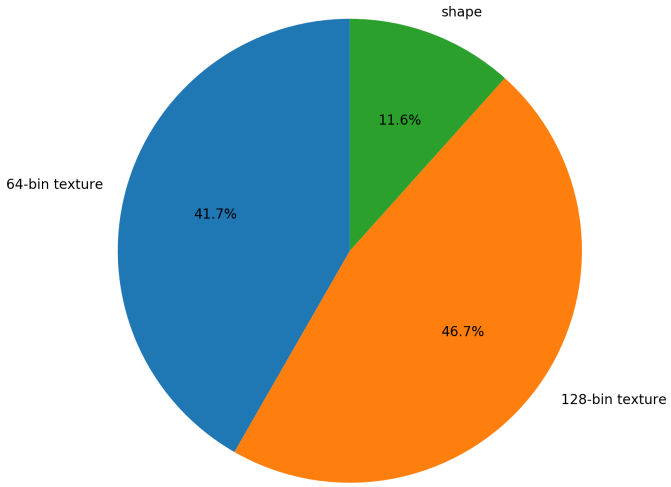
4.4.5 The Thalamus

In the thalamus classification experiment, the distribution of the selected features across the feature categories resembled the distributions of the other brain structures presented in this chapter. The distributions of the thalamus' experiments were almost identical according to Figure 4.33. Despite of the number of 64-bin texture features selected according to Figure 4.33, only one of them had a high enough selection rate to be included in Figures 4.34a and 4.34b, illustrating the five most selected features by both experiments. The 128-bin texture features and the shape

features included in the Selection Rate Figure 4.34 were all extracted based on the subjects' right thalamus.

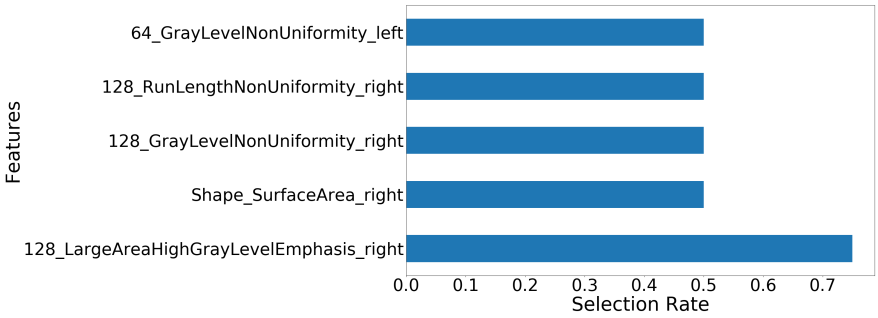


(a) The selected features distribution based upon the features of the left- and right sided ROIs separately.

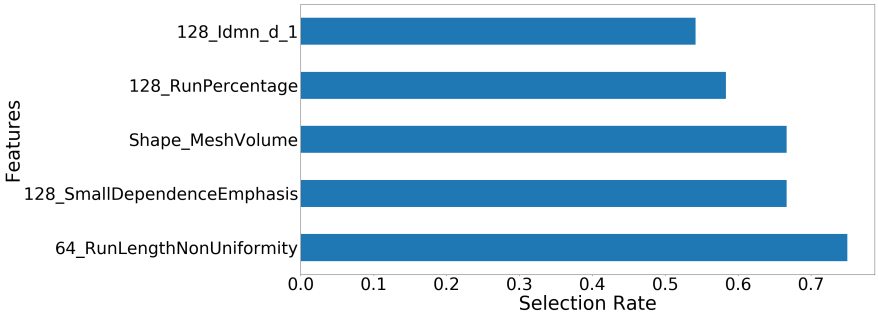


(b) The selected features distribution based upon the mean value of the left- and right sided ROI's features.

Figure 4.33: The figure illustrates the overall distribution of the features selected by the thalamus' classification experiments. The features were sorted by their feature categories; shape, 128-bin texture and 64-bin texture features. The pie chart in subfigure (a) displays the selected features distribution from the experiment considering the right- and left-sided ROI's separately. Subfigure (b) illustrates the distribution of the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.



(a) The five most selected features based upon the features of the left- and right sided ROIs separately.

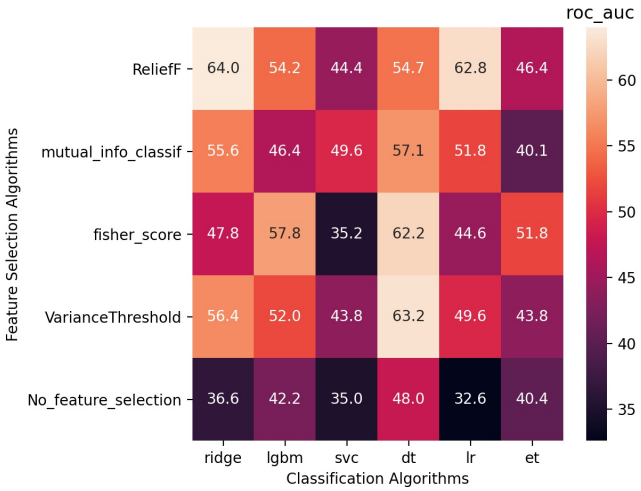


(b) The five most selected features based upon the mean value of the left- and right sided ROI's features.

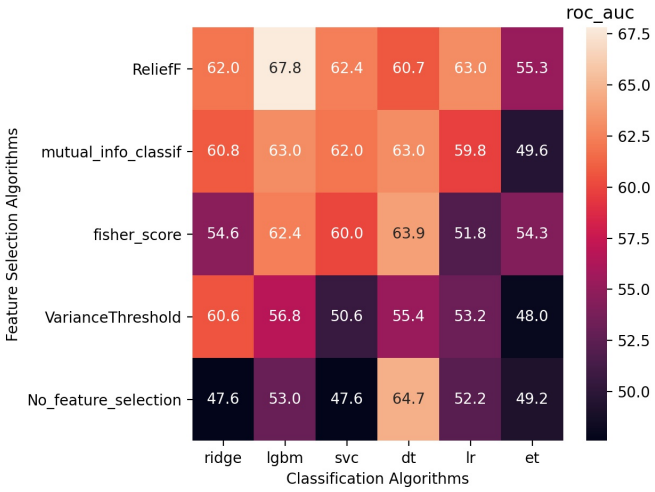
Figure 4.34: The figure illustrates the selection rate of the five most selected features by the thalamus' classification experiments. The name of the features indicate the feature type, shape corresponding to shape features and 64 to 64-bin texture features. Subfigure (a) displays the selected features from the experiment considering the right- and left-sided ROI's separately, while subfigure (b) illustrates the distribution of the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.

The performance of the experiments were evaluated by calculating the AUC scores of the classification models, and were illustrated as heatmaps in Figure 4.35. The first experiment, considering the left- and right-sided ROIs' separatley, yielded AUC scores ranging from approximately 33 to 64, while the mean based experiment spanned a slightly smaller range from approximately 48 to 68. Overall, all classifiers in both experiments returned two or more AUC scores approximately equal or below 60.

Classifications performed without feature selection solely yielded AUC scores below 50 in Figure 4.35a, and below 60 in Figure 4.35b with one exception. The four highest scoring classifications in Figure 4.35a were performed by either using the ReflefF selection algorithm or the DT classification algorithm. The ReflefF algorithm also performed better compared to the remaining selectors in the mean-based experiment according 4.35b.



(a) The AUC scores from the prediction models based upon the features of the left- and right sided ROIs separately.



(b) The AUC scores from the prediction models based upon the mean value of the left- and right sided ROI's features.

Figure 4.35: The average AUC scores achieved by the thirty prediction models generated by the thalamus' classification experiments. Four feature selectors were implemented together with six classification algorithms. In addition, classifications using the six algorithms was performed without feature selection. Subfigure (a) displays the AUC scores from the experiment considering the right- and left-sided ROI's separately, while subfigure (b) illustrates the scores from the experiment predicting the subjects' treatment group based on the mean of the right- and left-sided ROI's features.

Chapter 5

Discussion

The main objective of this thesis was to examine the effects of methylphenidate on regional locations in the grey matter of brains of ADHD diagnosed children. Analyses performed on magnetic resonance images suggested significant changes in the volume of the right pallidum and right hippocampus as well as changes in the surface area of the pallidum and the right thalamus. Results from classification experiments indicated further detectable changes due to the ADHD-treatment in the pallidum, the putamen and the hippocampus. These findings will be discussed, along with the findings of the performed principal component analyses. Uncertainties introduced due to brain structure segmentation and image pre-processing will also be evaluated.

5.1 Evaluation of the Data Pre-Processing

Before being able to perform the analyses or extract features from the images, the dataset had to be cleaned and pre-processed. MRIs should not be analysed raw, due to their intensity values being arbitrary units and therefore not comparable based on their voxels' intensity values [16, 32, 46]. An analysis of images across various patients, scanners or protocols, would require a form of standardisation [32, 46]. Establishing a suitable pre-processing of the images proved challenging, as multiple techniques have been applied by previous studies involving MRIs [14, 15, 29, 46, 47].

The standardisation method applied was implemented based on recommendations and techniques from Collewet et al. (2004), Isensee et al. (2018) and Duron et al. (2019), normalising the images based on the intensity values' z-scores and rescaling the images of the ROIs to 8-bit images, as described in Chapter 3.4.3. Also, voxels with intensity values outside of the range $[\mu_{ROI} - 3\sigma_{ROI}, \mu_{ROI} + 3\sigma_{ROI}]$ were removed, as recommended by Collewet et al. (2004), Zwanenburg et al. (2016) and Duron et al. (2019). The process was referred to as the autoscaling process.

5.1.1 The Autoscaling Process

Normalising the Intensity Level Distribution

The autoscaling process had three main effects on the MR images. Firstly, the images' intensity distribution was altered. As seen in Figures 4.3 and D.1-D.9, the minimum, maximum and mean intensity values of the structures varied considerably from subject to subject before performing the autoscaling process. Applying the autoscaling aligned the intensity statistics of the structures to equal minimum and maximum values, and approximate equal mean values across the subjects, allowing further analyses of the intensity relationships of the ROIs.

Rescaling the Intensity Value Range

Secondly, rescaling images down to 8-bit images enhanced the intensity contrasts of the structures, as seen in Figure 4.1 and visualised new patterns in the intensity distribution. As explained in Chapter 3.5, rescaling or discretising an image, affects the level of details included in the images. Due to the nature of the radiomic texture features extracted from the ROIs, the number of patterns detectable in the extracted features is directly correlated to the intensity value range of the ROIs' image volumes [76].

The ROIs analysed in this thesis were relatively small compared to the brain's total size. Hence, the intensity level range of the complete brain MRI would not represent the relationship of the ROIs intensity levels optimally. A new intensity level range was created for every ROI by rescaling the intensity levels to values in the range [0,255]. The rescaling of the images of the ROIs ensured a larger intensity level variation enabeling detection of subtle texture differences between the structures.

Exclusion of Artefact Voxels

The last effect of the autoscaling was the exclusion of possible outlier voxels, referred to as artefact voxel exclusion in this thesis. According to Perlaki et al. (2017) [42], FreeSurfer tends to overestimate the volume of the caudate and putamen compared to manual segmentations. To correct for the possible uncertainties in segmentation created by FreeSurfer, voxels with intensity values outside of range $[\mu_{ROI}-3\sigma_{ROI}, \mu_{ROI}+3\sigma_{ROI}]$ were removed, as described in Chapter 3.4.3. According to a recommendation made by Gillies et al. (2016) [12], the automated segmentation should be followed by a manual curation in order to achieve optimum reproducible segmentation. However, an automated correction method was chosen instead to reduce time usage and eliminate further involvement of physicians.

While the rescaling of the images only affected the radiomic texture features, the exclusion of voxels affected the shape features as well due to their respective definitions defined by Zwanenburg et al. (2020) [33]. By excluding artefact voxels, a decrease in the ROIs' surface area was expected. To examine whether a such a reduction took place, the percentage change of the ROIs' surface area due to the

voxel exclusion was plotted. As described in Chapter 4.1.1 the surface area of the subjects' caudate and thalamus decreased on average. In contrast, the surface area of the subjects' remaining structures increased. Such a behaviour can be explained by the structures' changing surface due to the voxel removal. To examine the effect of the voxel exclusion closer, the ROIs' volume change also should have been examined. As the observed average increase in surface area did not exceed 2%, no such examination was performed.

5.2 Evaluating the Performed Analyses

A variation of papers have been published in the field of psychological radiomics, assessing changes in brain structures due to psychological disorders such as alzheimer's disease [77], bipolar disorder [78] and dementia [79]. A majority of papers addressing ADHD, focus mainly on discriminating ADHD diagnosed patients from a control group. One of the papers examining structural changes in ADHD patients was Hoogman et al. (2017) [22]. Similar to this thesis, subcortical structures were analysed, examining the effect of ADHD on the structures' volume as well as the effect of stimulation such as MPH or amphetamine [22]. However, no radiomic features were involved in the analysis. In contrast, Sun et al. (2018) [5] applied radiomic features to identify characteristics discriminating the ADHD group from healthy control subjects. This paper assessed the possibility of classifying subjects as ADHD diagnosed or healthy, achieving an accuracy of 73.7% [5].

It is important to point out that no earlier studies tied to the ePOD-MPH trial examined the structural differences between the non-treated and the treated test-subjects' subcortical anatomies based on radiomic features. Schrantee et al. (2020) discussed structural changes connected to MPH-treatment based on the same dataset, but not associated to the difference between the treatment groups. Instead the possibility of using the caudate volume, amongst other brain structures, as a biomarker to predict nonresponse in MPH-treatment patients was examined. Schrantee et al. (2016) established a connection between MPH-treatment and change in cerebral blood flow in the striatum and thalamus. However, no structural examinations were performed.

Due to the lack of studies examining the effects of MPH based on radiomic features, the findings of this thesis were compared to studies and papers examining ADHD without ties to radiomics or MPH treatment, such as Hoogman et al. (2017), Frodl and Skokauskas (2012) and Ellison-Wright et al. (2008).

It is also important to note the difference in the complexity of the performed analyses. While the boxplot and PCA analyses both served as exploratory analyses, the classification experiments were limited to an initial screening. Optimally, the best performing models would have been examined closer to detect why they perform better compared to the rest. Due to time restraints, an overall evaluation of the classification models was prioritised in order to establish whether closer examination of the models in future work would be useful.

5.2.1 Comparing the Structure Volumes and Surface Areas

As multiple studies have been performed researching the effect of ADHD on the brain's structural anatomy in children [20, 22, 23, 26, 40, 41], boxplot analyses were performed assessing whether there was a connection between MPH-treatment of ADHD and structural brain changes in this study. For each of the structures listed in Chapter 3.3 an analysis was performed comparing the surface area and volume of the left and right ROIs across the treatment groups. In addition, the mean surface area and volume of the treated and non-treated subjects' left and right ROI were compared.

As described in Chapter 4.2, significant differences in volume were detected in the right hippocampus (Figure 4.5a) and the right pallidum (Figure 4.6a). No significant differences were detected in the mean volume of the structures, suggesting the difference in volume between the groups was not as prominent as implied by Figures 4.5a and 4.6a. Based on Figure 4.8, the same can be concluded for the surface area of the thalamus. In contrast, the surface area of the MPH-treated subjects' pallidum as well as the mean surface area of the left and right pallidum appeared to be significantly larger compared to the placebo-treated subjects' pallidum.

It should be noted that in contrast to the right-sided ROIs, no left-sided ROIs suggested significant changes discriminating the treatment groups. The lack of significant differences in the left-sided ROIs was supported by the review papers Frodl and Skokauskas (2012) [26] and Ellison-Wright et al. (2008) [39], investigating the structural volume changes in the brain caused by ADHD. None of the structures highlighted in this papers were restricted to the left side of the brain.

In addition, it must be noted that uncertainties arose when performing shape feature based analyses in this thesis. The lack of the FreeSurfer segmentation's and the artefact voxel exclusion's precision need to be considered when evaluating the surface area and the volume of the examined structures. The influence of the uncertainty on the findings may be reduced by implementing better pre-processing segmentation corrections, such as manual curation as suggested by Gillies et al. (2016) [12]. The necessity of implementing such a correction tool becomes clear when comparing the findings in this thesis with the findings from Hoogman et al. (2017) [22]. This study mainly examined the effect of ADHD on regional brain structures in the grey matter, but a comparison of treated subjects to non-treated subjects was also performed. No significant differences in structure volume discriminating the groups were identified, in contrast to this thesis' findings [22].

When evaluating the boxplots, it is important to note that an approximate normal distribution of the observations was assumed. According to literature, the Welch's t-test is robust against non-normal distributions given a homogenous variance or an equal distribution of subjects between the classes [56]. Despite the possibility of a

skew distribution, Welch's t-test was applied due to the sample sizes of the classes only being slightly dissimilar.

5.2.2 The Principal Component Analyses

Principal component analyses were performed in this thesis to search for patterns in the extracted radiomic features distinguishing the treatment groups. The shape features, 64-bin texture feature and 128-bin texture features were analysed separately. The mean values of the left and right ROIs' radiomic features were not analysed using PCA due to time constraints. For each structure three PCA models were created, as explained in Chapter 3.7.2. Score plots were created to identify characteristics separating the classes, and subjects standing out were listed in Table H.1 for possible future studies of the ePOD-MPH dataset.

Before deciding upon the number of components to include in the PCA models, the RMSEC and the RMSECV were examined. As described in Chapter 4.3, the decreasing RMSEC and the increasing RMSECV of the shape feature models suggested the models kept collecting unnecessary information from the features and overfitted quickly. This behaviour was observed in all shape feature models, independent of the structure in question. Due to the overfitting, the information extracted from the shape feature models could have been left out from further analyses. No such exclusion was performed, and score plots were created and visually examined based on the shape feature models. No combination of principal components included in the score plots distinguished the classes noticeably, suggesting no effects due to methylphenidate treatment were detected by the shape feature PCAs.

The PCA models based on the 64-bin and 128-bin texture features also proved to behave similarly to each other across structures. Instead of overfitting early, the RMSEC and RMSECV kept decreasing before flattening. The flattening of the RMSECV curve indicated at which point the models stopped catching new, utilizable information and started overfitting too. Compared to the shape feature models, on average more principal components were included in the texture feature models, suggesting the texture features should be preferred in possible future work related to this analysis as they may contain useful information about the subjects. Although the texture feature based PCA models seemed more promising compared to the shape feature models based on the RMSECV curves, no characteristics distinguishing the treatment classes were detected in either models.

Outlier detection

None of the score plots included in the thesis managed to present clear distinctions between the MPH- and placebo treated subjects. Nevertheless, smaller groups or individuals stood out in the score plots, suggesting characteristics were captured highlighting the subjects. Selected subjects were identified according to Chapter 3.7.2 and listed in Table H.1, enabling future studies to examine the characteristics

making the trial subjects stand out.

Causes for standing out could be anatomical and tied to the MPH-treatments, or associated with MR image qualities, the segmentation accuracy or the image preprocessing performed. In Chapter 4.1.1, trial participants with relative high changes in surface area due to voxel exclusion were listed. These subjects, except subject nr. 1, were also listed at least once in Table H.1, suggesting a connection between the voxel exclusion and the principal component analyses.

If examining the subjects listed in Table H.1 in future studies, it is important to note that all selections were performed based on visual inspections and are therefore to be considered relatively subjective. To limit the subjectivity, the highlighted subjects listed in Table H.1 were compared to Hotelling's T^2 - Q residual plots. Subjects were considered as possible outliers if they stood out in all three PCA models of a structure and exceeded the 95% confidence limit.

Future studies of the dataset should examine the possible outliers and determine whether they should be excluded from further analyses or examined in further detail. The list of possible outliers could also be used as a tool in future studies to compare the features of the subjects to new patients. By coordinating with physicians, not only anatomical characteristics could be examined, but also include and analyse clinical factors tied to the patients.

5.2.3 The Classification Experiments

Before examining the classification experiments, it should once again be noted that the classification experiments only covered an initial screening to evaluate whether closer examinations of the classification models are worthwhile.

When extracting the texture features using Biorad [36], image discretisation was applied. Performing discretisation decreases the sparsity of the texture matrices generating the texture features, and can be performed either using a fixed number of bins or a fixed bin size [71]. The method of discretisation was already implemented in Biorad [36], but the bin size still had to be decided upon. Bins ranging two and four intensities were chosen, resulting in discretised images ranging 128 and 64 intensity levels, respectively.

The pie charts illustrating the distribution of the features selected across the feature categories, suggest that texture features were preferred to the shape features by the feature selectors. Approximately 85-90% of the selected features in every experiment belonged to either the 64-bin texture features or the 128-bin texture features. The distribution between the two texture feature categories was approximately equal in all experiments according to the provided pie charts. Based on the findings of the classification experiments, it could not be decided which of the bin sizes contributed

most to the high-scoring models.

The texture features also dominated the figures illustrating the selection rate of the five most selected features of every experiment, but due to the lack of more detailed examinations of the models it could not be determined whether the texture features contributed to the best performing models. Similar to the RMSECV of the PCAs, the figures illustrating the features' selection rate suggest future radiomic feature analyses should prioritise analysing the texture features rather than shape features.

The figures illustrating the selection rate of the five most selected features in every experiment also suggested that the right ROI of the structures dominated the classifications. The boxplots in Chapter 4.2 reflect the importance of the right ROIs of the structures, as significant differences between the MPH-treated group and placebo-treated groups only were detected in the right ROIs and the mean of the left and right ROIs. The dominance of the right-sided ROIs can be traced back to the review papers by Frodl and Skokauskas (2012) [26] and Ellison-Wright et al. (2008) [39], indicating that the right-sided ROIs were affected by ADHD and therefore also may be influenced by methylphenidate.

As described in Chapter 4.4 and seen in the corresponding figures, the AUC of the experiments varied across feature selectors and classifiers, suggesting a possible overfit in good performing models or arbitrary high achieved AUCs. No structure presented AUC scores solely above 60, and only a few models surpassed an AUC of 70. While the caudate and thalamus experiments did not return any AUC scores above 70, the pallidum and putamen had top scores close to 80. The hippocmapus' experiments' highest achived AUC equaled 73.

The models achieving AUC scores above 70 were based on varying combination of feature selectors and classifiers across the structures. Therefore, no best performing feature selector-classifier combination could be identified. No classification algorithms achieved AUC scores consistently close to 70, but the Decision Tree Classifier and Extremely Randomised Tree Classification performed relatively better than the remaining classifiers. ReliefF and the Fisher Score selectors returned overall relatively high AUC scores on average in all experiments, independent of the selected classification algorithm. The experiments clarified the need of feature selection, as the classifications performed without selection on average returned lower scores compared to the models performed with feature selection.

Overall, the experiments performed well enough to suggest further examinations of the models of all structures. Especially the pallidum and putamen experiments should be examined closer in future studies as the high AUC scores indicated the MPH and placebo groups may be distinguishable based on the radiomic features. A closer examination of the pallidum's, the hippocampus' and the thalamus' models was also supported by the boxplot analyses performed, as significant differences between the treatment groups were detected.

In addition to recommend performing the classifications only based on texture features, it would also be interesting to examine whether a model based on radiomic features from multiple structures could identify more patterns than the existing experiments.

5.3 Further Work

In order to compare and reproduce findings across studies, a guideline for precise segmentations of ROIs in MRIs is of essence. As explained in Chapter 3.4.3 and according to Perlaki et al. (2017) [42], automated segmentation techniques such as FreeSurfer may cause inaccurate renderings of the ROIs and produce artefact voxels. In this thesis the exclusion of said voxels was limited to the exclusion technique presented by Collewet et al. (2004) [14], removing voxels outside of the intensity range $[\mu_{ROI} - 3\sigma_{ROI}, \mu_{ROI} + 3\sigma_{ROI}]$. As structural differences between treated and non-treated subjects' brain structures may be relative minor, there is a need for precise segmentations. Future studies should examine whether a more precise exclusion of artefact voxels is possible without performing manual corrections.

As this thesis is limited to an initial screening of the classification experiments, no detailed examinations of the models and their individual selected features were performed. The highest scoring models of the hippocampus (AUC$\approx$70), the pallidum (AUC$\approx$80) and the putamen (AUC$\approx$80) indicate the possibility of differentiating MPH-treated patients from placebo-treated patients based on radiomic shape and texture features. In order to asses this possibility, the classification models have to be closer examined. Specifically, the contributions of the different features and feature categories to the models have to be examined. However, before performing such an assessment, the models should be tested for overfitting. AUC scores above 70 could have been caused by overfitting the data available. Overfitting can be caused by various factors, such as a high number of included features in a model or a small sized dataset [59]. The models should be inspected for both factors, excluding or confirming a possible overfitting.

It should also be examined whether the structures' right-sided ROI are directly linked to structural changes due to MPH treatment, as suggested by the boxplot analyses and the features highlighted in the selection rate figures in Chapter 4.4. Review papers by Ellison-Wright et al. (2008) [39] and Frodl and Skokauskas (2012) [26] suggested an decrease in the right putamen and the right pallidum of ADHD diagnosed children, compared to non-diagnosed patients. According to the findings from the boxplot analyses, the right pallidum responded to the MPH-treatment by increasing. Future studies should therefore assess the possibility that medicating children diagnosed with ADHD reduced the structural effects caused by ADHD in brain structures, as suggested by Wilens and Spencer (2010) [18]. However, this only applied for the pallidum, as no such connections were identified in the right putamen in this thesis.

In addition, other brain structures tied to ADHD and MPH should be analysed and compared to the structures already analysed. Eligible structures for initial screening and analyses could be the corpus callosum, the prefrontal cortex and the cerebellum. The structures in question have all been connected to changes caused by ADHD [20] and non-response to MPH-treatment [10].

In this thesis, possible outliers and subjects standing out based on PCA models were identified. Assessing possible connections between the subjects standing out and their treatment group affiliation could shed further light upon the effects of MPH-treatment on children. In addition to examining structural features, clinical factors should be analysed by physicians to search for similarities.

Effects of MPH are not restricted to structural changes [9, 10, 25], and an inclusion of features extracted from other MRI modalities in future analyses may increase the probability of identifying brain structure characteristics tied to the MPH-treatment. Schrantee et al. (2016) [9] used pharmacological MRIs (phMRI) to examine and state differences between the treatment groups on the basis of the cerebral blood flow. An inclusion of features and information from other MRI modalities such as phMRIs and functional MRIs (fMRI) should therefore be considered included in future classification experiments.

Chapter 6

Conclusions

In this thesis, the effects of methylphenidate on ADHD diagnosed children were assessed, utilising radiomic shape and texture features. The analyses were restricted to examining structural changes in the caudate, the hippocampus, the pallidum, the putamen, and the thalamus. In addition to examining the right and left regions of the brain structures separately, the structures' regions were assessed as a whole. T1-weighted MRIs analysed in this thesis were collected as part of the ePOD-MPH study [11].

The analyses performed indicated significant changes in the brain structures' shape due to MPH-treatment. According to the boxplot analyses, the surface area of the right pallidum and the right thalamus, as well as the mean surface area of the pallidum, differed significantly between the treatment groups. The same applied to the volume of the right hippocampus and right pallidum. Classification experiments predicted the treatment group of the trial subjects based on the radiomic features extracted from the ROIs. The highest AUC score amongst the structures was achieved by the putamen experiment based on the mean of the left- and right-sided ROI (AUC$\approx$80), by combining the Fisher score with the Light Gradient Boosting Machine-classifier. By combining the variance threshold selector with the decision tree classifier, a similar score was achieved by the pallidum experiment based on the features of the left and right ROIs separately (AUC$\approx$80). The best performing model of the caudate, the hippocampus and the thalamus scored approximately 65, 75 and 70, respectively. The highest AUC scores in the caudate's, hippocampus' and pallidum's classification experiments were achieved by combining the Variance Threshold feature selector with either the Decision Tree or the Extremely Randomised Tree classifier. The combination of the Light Gradient Boosting Machine classifier with the ReliefF and the Fisher Score feature selector algorithm performed best in the thalamus' and the putamen's experiments, respectively. Furthermore, the Fisher Score algorithm and the ReliefF algorithm displayed the potential for stable and relative high AUC scores across all performed experiments, independent of the classifiers included.

Models with AUC scores above approximately 70 suggested detectable changes in the hippocampus, the pallidum and the putamen due to MPH-treatment, based on

analyses of radiomic shape and texture features. However, based on the lack of detected patterns distinguishing the treatment groups in the PCAs and the number of models returning AUC scores close to or below 50, further examinations of the features and classification models are necessary before a detectable change due to MPH can be confirmed for certain.

Due to the absence of a standardised normalisation method for MRIs before extracting radiomic features from brain regions, an autoscaling process based on Collewet et al. (2004) and Isensee et al. (2018) was implemented. As a part of the autoscaling, artefact voxels were excluded from further analyses to correct for possible segmentation inaccuracies. In order to reduce intra- and interobservation variability and securing reproducible and comparable findings across studies and projects, preprocessing guidelines for handling brain MRIs segmented using automated software should be established.

Bibliography

[1] Cheima Bouziane, Olena G. Filatova, Anouk Schrantee, Matthan W. A. Caan, Frans M. Vos, and Liesbeth Reneman. White matter by diffusion mri following methylphenidate treatment: A randomized control trial in males with attention-deficit/hyperactivity disorder. *Radiology*, 293(1):186–192, 2019. doi: 10.1148/radiol.2019182528. URL `https://doi.org/10.1148/radiol.2019182528`. PMID: 31407970.

[2] Todd Elder. The importance of relative standards in adhd diagnoses: Evidence based on exact birth dates. *Journal of health economics*, 29:641–56, 09 2010. doi: 10.1016/j.jhealeco.2010.06.003.

[3] Gigi Loo-Neus, Nanda Rommelse, and Jan Buitelaar. To stop or not to stop? how long should medication treatment of attention-deficit hyperactivity disorder be extended? *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology*, 21:584–99, 04 2011. doi: 10.1016/j.euroneuro.2011.03.008.

[4] Angelika Schlarb, Martina Starck, and Julia Grünwald. Occurrence of adhd in parents of adhd children in a clinical sample. *Neuropsychiatric Disease and Treatment*, 12:581, 03 2016. doi: 10.2147/NDT.S100238.

[5] Huaiqiang Sun, Ying Chen, Qiang Huang, Su Lui, Xiaoqi Huang, Yan Shi, Xin Xu, John A. Sweeney, and Qiyong Gong. Psychoradiologic utility of mr imaging for diagnosis of attentio deficit hyperactivity disorder: A radiomics analysis. *Radiology*, 287(2):620–630, 2018. doi: 10.1148/radiol.2017170226. URL `https://doi.org/10.1148/radiol.2017170226`. PMID: 29165048.

[6] John Port. Diagnosis of attention deficit hyperactivity disorder by using mr imaging and radiomics: A potential tool for clinicians. *Radiology*, 287:631–632, 05 2018. doi: 10.1148/radiol.2018172804.

[7] T. Grund, K. Lehmanm, N. Bock, A. Rothenberger, and G. Teuchert-Noodt. Influence of methylphenidate on brain development–an update of recent animal experiments. *Behavioral and brain functions : BBF*, 2:2, 2006. URL `https://doi.org/10.1186/1744-9081-2-2`.

[8] Klaus W. Lange. The evidence of the benefits and harms of methylphenidate in the treatment of attention deficit/hyperactivity disorder is inconclusive. *Journal of Pharmacology Clinical Toxicology*, 6(4), 2018.

[9] Anouk Schrantee, Hyke G. H. Tamminga, Cheima Bouziane, Marco A. Bottelier, Esther E. Bron, Henk-Jan M. M. Mutsaerts, Aeilko H. Zwinderman,

Inge R. Groote, Serge A. R. B. Rombouts, Ramon J. L. Lindauer, Stefan Klein, Wiro J. Niessen, Brent C. Opmeer, Frits Boer, Paul J. Lucassen, Susan L. Andersen, Hilde M. Geurts, and Liesbeth Reneman. Age-dependent effects of methylphenidate on the human dopaminergic system in young vs adult patients with attention-deficit/hyperactivity disorder: A randomized clinical trial. *JAMA psychiatry*, 73(9):955–962, Sep 2016. ISSN 2168-6238. doi: 10.1001/jamapsychiatry.2016.1572. URL `https://www.ncbi.nlm.nih.gov/pubmed/27487479`.

[10] Anouk Schrantee, Henricus Gerardus Ruhé, and Liesbeth Reneman. Psychoradiological biomarkers for psychopharmaceutical effects. *Neuroimaging Clinics*, 30(1):53–63, 2020.

[11] Marco Bottelier, Marieke Schouw, Anne Klomp, Hyke Tamminga, Anouk Schrantee, Cheima Bouziane, Michiel de Ruiter, Frits Boer, Henricus Ruhé, Damiaan Denys, Roselyne Rijsman, Ramón Lindauer, Hans Reitsma, Hilde Geurts, and Liesbeth Reneman. The effects of psychotropic drugs on developing brain (epod) study: Methods and design. *BMC psychiatry*, 14, 02 2014. doi: 10.1186/1471-244X-14-48.

[12] Robert J. Gillies, Paul E. Kinahan, and Hedvig Hricak. Radiomics: Images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2016. doi: 10.1148/radiol.2015151169. URL `https://doi.org/10.1148/radiol.2015151169`. PMID: 26579733.

[13] Marius E Mayerhoefer, Andrzej Materka, Georg Langs, Ida Häggström, Piotr Szczypinski, Peter Gibbs, and Gary Cook. Introduction to radiomics. *Journal of Nuclear Medicine*, 2020. doi: 10.2967/jnumed.118.222893.

[14] Guylaine Collewet, Michal Strzelecki, and François Mariette. Influence of mri acquisition protocols and image intensity normalization methods on texture classification. *Magnetic resonance imaging*, 22:81–91, 2004. doi: 10.1016/j.mri.2003.09.001.

[15] Fabian Isensee, Jens Petersen, André Klein, David Zimmerer, Paul Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, S. Wirkert, and Klaus Maier-Hein. nnu-net: Self-adapting framework for u-net-based medical image segmentation. 2018. *arXiv:1809.10486v1* ***[cs.CV]***.

[16] L. Duron, D. Balvay, S. Vande Perre, A. Bouchouicha, J. Savatovsky, J. C. Sadik, I. Thomassin-Naggara, L. Fournier, and A. Lecler. Gray-level discretization impacts reproducible mri radiomics texture features. *PLoS one*, 14(3), 2019. doi: 10.1371/journal.pone.0213459. URL `https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0213459`.

[17] L. B. Sollaci and M. G. Pereira. The introduction, methods, results, and discussion (imrad) structure: a fifty-year survey. *Journal of the Medical Library Association : JMLA*, 92(3):364–367, 2004.

[18] T. E. Wilens and T. J. Spencer. Understanding attention-deficit/hyperactivity disorder from childhood to adulthood. *Postgraduate medicine*, 122(5):97–109, 2010. URL `https://doi.org/10.3810/pgm.2010.09.2206`.

[19] Ajay Singh, Chia Jung Yeh, Nidhi Verma, and Ajay Kumar Das. Overview of attention deficit hyperactivity disorder in young children. *Health psychology research*, 3(2), 2015. URL https://doi.org/10.4081/hpr.2015.2115.

[20] Gail Tripp and Jeffery R. Wickens. Neurobiology of adhd. *Neuropharmacology*, 57(7):579 – 589, 2009. ISSN 0028-3908. doi: https://doi.org/10.1016/j.neuropharm.2009.07.026 URL http://www.sciencedirect.com/science/article/pii/S0028390809002445.

[21] Frank Amthor. *Das Menschliche Gehirn für Dummies [The Human Brain for Dummies]*. Wiley-VCH Verlag GmbH und Co., 2 edition, 2019.

[22] M. Hoogman, J. Bralten, D. P. Hibar, M. Mennes, M. P. Zwiers, L. Schweren, van Hulzen K., S. E. Medland, E. Shumskaya, N. Jahanshad, P. Zeeuw, E. Szekely, G. Sudre, T. Wolfers, A. Onnink, J. T. Dammers, J. C. Mostert, Y. Vives-Gilabert, G. Kohls, E. Oberwelland, and B. Franke. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *The Lancet. Psychiatry*, 4(4):310–319, 2017.

[23] Iliyan Ivanov, Ravi Bansal, Xuejun Hao, Hongtu Zhu, Cristoph Kellendonk, Loren Miller, Juan Sanchez-Pena, Ann M. Miller, M. Mallar Chakravarty, Kristin Klahr, Kathleen Durkin, Laurence L. Greenhill, and Bradley S. Peterson. Morphological abnormalities of the thalamus in youths with attention deficit hyperactivity disorder. *American Journal of Psychiatry*, 167(4):397–408, 2010. doi: 10.1176/appi.ajp.2009.09030398. URL https://doi.org/10.1176/appi.ajp.2009.090303.

[24] Xueyi Shen, Lianne M. Reus, Simon R. Cox, Mark J. Adams, David C. Liewald, Mark E. Bastin, Daniel J. Smith, Ian J. Deary, Heather C. Whalley, and Andrew M. McIntosh. Subcortical volume and white matter integrity abnormalities in major depressive disorder: findings from uk biobank imaging data. *Sci Rep*, 7, 2017. URL https://doi.org/10.1038/s41598-017-05507-6.

[25] Veronika Engert and Jens C. Preussner. Dopaminergic and noradrenergic contributions to functionality in adhd: the role of methylphenidate. *Curr Neuropharmacol*, 6(4):322–328, 2008. doi: 10.2174/157015908787386069.

[26] T. Frodl and N. Skokauskas. Meta-analysis of structural mri studies in children and adults with attention deficit hyperactivity disorder indicates treatment effects. *Acta Psychiatrica Scandinavica*, 125(2):114–126, 2012. doi: 10.1111/j.1600-0447.2011.01786.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1600-0447.2011.01786.x.

[27] Mohammad Balafar, Abdul Ramli, M. Iqbal Saripan, and Syamsiah Mashohor. Review of brain mri image segmentation methods. *Artif. Intell. Rev.*, 33:261–274, 03 2010. doi: 10.1007/s10462-010-9155-0.

[28] M. A. Flower, editor. *Webb's Physics of Medical Imaging (Second Edition)*. CRC Press, second edition edition, 2012.

[29] László Nyúl and Jayaram Udupa. On standardizing the mr image intensity scale. *Magnetic resonance in medicine*, 42(6):1072–81, 12 1999.

[30] A. Zwanenburg, S. Leger, M. Vallières, and S. Löck. Image biomarker standardisation initiative: Reference manual. 2016. URL https://arxiv.org/abs/1612.07003v10. *arXiv preprint arXiv:1612.07003. v10.*

[31] Kieran Maher. T1t2pd.jpg. From Wikimedia Commons. Retrieved 05.08.2020 from https://commons.wikimedia.org/wiki/File:T1t2PD.jpg.

[32] S. Rizzo, F. Botta, S. Raimondi, D. Origgi, C. Fanciullo, A. G. Morganti, and M. Bellomi. Radiomics: the facts and the challenges of image analysis. *European radiology experimental*, 2(36), 2018. doi: https://doi.org/10.1186/s41747-018-0068-z.

[33] Alex Zwanenburg, Martin Vallières, Mahmoud A. Abdalah, Hugo J. W. L. Aerts, Vincent Andrearczyk, Aditya Apte, Saeed Ashrafinia, Spyridon Bakas, Roelof J. Beukinga, Ronald Boellaard, Marta Bogowicz, Luca Boldrini, Irène Buvat, Gary J. R. Cook, Christos Davatzikos, Adrien Depeursinge, Marie-Charlotte Desseroit, Nicola Dinapoli, Cuong Viet Dinh, Sebastian Echegaray, Issam El Naqa, Andriy Y. Fedorov, Roberto Gatta, Robert J. Gillies, Vicky Goh, Michael Götz, Matthias Guckenberger, Sung Min Ha, Mathieu Hatt, Fabian Isensee, Philippe Lambin, Stefan Leger, Ralph T.H. Leijenaar, Jacopo Lenkowicz, Fiona Lippert, Are Losnegård, Klaus H. Maier-Hein, Olivier Morin, Henning Müller, Sandy Napel, Christophe Nioche, Fanny Orlhac, Sarthak Pati, Elisabeth A.G. Pfaehler, Arman Rahmim, Arvind U.K. Rao, Jonas Scherer, Muhammad Musib Siddique, Nanna M. Sijtsema, Jairo Socarras Fernandez, Emiliano Spezi, Roel J.H.M. Steenbakkers, Stephanie Tanadini-Lang, Daniela Thorwarth, Esther G.C. Troost, Taman Upadhaya, Vincenzo Valentini, Lisanne V. van Dijk, Joost van Griethuysen, Floris H.P. van Velden, Philip Whybra, Christian Richter, and Steffen Löck. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. *Radiology*, 295(2):328–338, 2020. doi: 10.1148/radiol.2020191145. URL https://doi.org/10.1148/radiol.2020191145. PMID: 32154773.

[34] PyRadiomics Community. Pyradiomics documentation. URL http://pyradiomics.readthedocs.io (accessed: 03.08.2020).

[35] Alise Danielle Midtfjord. Prediksjon av behandlingsutfall for hode- og halskreft ved bruk av radiomics av pet/ct-bilder [prediciton of treatment outcome of head and throat cancer using radiomics of pet/ct images], 2018.

[36] Ahmed Albuni. *Biorad*, 2020. Version: 0.01. URL https://github.com/ahmedalbuni/biorad.

[37] Ahmed Albuni. Development of a user-friendly radiomics framework. M.sc. thesis, Norwegian University of Life Sciences, Ås, 2020.

[38] Geir Severin Rakh Elvatun Langberg. Searching for biomarkers of disease-free survival in head and neck cancers using pet/ct radiomics. M.sc. thesis, Norwegian University of Life Sciences, Ås, 2019.

[39] I. Ellison-Wright, Z. Ellison-Wright, and E. Bullmore. Structural brain change in attention deficit hyperactivity disorder identified by meta-analysis. *BMC psychiatry*, 8, 51, 2008. URL https://doi.org/10.1186/1471-244X-8-51.

[40] Kerstin J. Plessen, Ravi Bansal, Hongtu Zhu, Ronald Whiteman, Jose Amat, Georgette A. Quackenbush, Laura Martin, Kathleen Durkin, Clancy Blair, Jason Royal, Kenneth Hugdahl, and Bradley S. Peterson. Hippocampus and amygdala morphology in attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, 63(7):795–807, 07 2006. URL https://doi.org/10.1001/archpsyc.63.7.795.

[41] M. J. Batty, L. Palaniyappan, G. Scerif, M. J. Groom, E. B. Liddle, P. F. Liddle, and C. Hollis. Morphological abnormalities in prefrontal surface area and thalamic volume in attention deficit/hyperactivity disorder. *Psychiatry research*, 233(2):225–232, 2015. URL https://doi.org/10.1016/j.pscychresns.2015.07.004.

[42] G. Perlaki, R. Horvath, S. A. Nagy, P. Bogner, T. Doczi, J. Janszky, and G. Orsi. Comparison of accuracy between fsl's first and freesurfer for caudate nucleus and putamen segmentation. *Scientific Reports*, 7(2418), 2017. URL https://doi.org/10.1038/s41598-017-02584-5.

[43] Michelle Tang, Pulkit Kumar, Hao CHen, and Abhinav Shrivastava. Deep multimodal learning for the diagnossis of autism spectrum disorder. *J.Imaging*, 47, 6 2020. doi: https://doi.org/10.3390/jimaging6060047.

[44] H. Knipe, K. Greeway, D. J. Bell, I. Bickle, A.K. Banerjee, Y. Weerakkody, and T. Luijkx. Hounsfield units, n.d. URL https://radiopaedia.org/articles/hounsfield-unit (accessed: 13.05.2020).

[45] D. J. Bell, K. Yap, T. Foster, D. Smith, A. Goel, and J. Jones. Standard uptake value, n.d. URL https://radiopaedia.org/articles/standard-uptake-value?lang=gb (accessed: 13.05.2020).

[46] Jessica Goya-Outi, Fanny Orlhac, Raphael Calmon, Agusti Alentorn, Christophe Nioche, Cathy Philippe, Stéphanie Puget, Nathalie Boddaert, Irène Buvat, Jacques Grill, Vincent Frouin, and Frédérique Frouin. Computation of reliable textural indices from multimodal brain MRI: suggestions based on a study of patients with diffuse intrinsic pontine glioma. *Physics in Medicine and Biology*, 63(10):105003, May 2018. doi: 10.1088/1361-6560/aabd21. URL https://hal.sorbonne-universite.fr/hal-01801219.

[47] S. Schindler, J. Schreiber, P. L. Bazin, R. Trampel, A. Anwander, S. Geyer, and P. Schönknecht. Intensity standardisation of 7t mr images for intensity-based segmentation of the human hypothalamus. *PLoS one*, 12(3), 2017. doi: 10.1371/journal.pone.0173344. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5333904/.

[48] Guido Van Rossum and Fred L. Drake. *Python 3 Reference Manual*. CreateSpace, Scotts Valley, CA, 2009. ISBN 1441412697.

[49] M. Shafiq-Ul-Hassan, G. G. Zhang, K. Latifi, G. Ullah, D. C. Hunt, Y. Balagurunathan, M. A. Abdalah, M. B. Schabath, D. G. Goldgof, D. Mackin, L. E. Court, R. J. Gillies, and E. G. Moros. Intrinsic dependencies of ct radiomic features on voxel size and number of gray levels. *Medical physics*, 44(3):1050–1062, 2017. doi: https://doi.org/10.1002/mp.12123.

[50] F. Tixier, C. C. Le Rest, M. Hatt, N. Albarghach, O. Pradier, J. P. Metges, L. Corcos, and D. Visvikis. Intratumor heterogeneity characterized by textural features on baseline 18f-fdg pet images predicts response to concomitant radiochemotherapy in esophageal cancer. *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, 52(3):369–378, 2011. doi: ://-doi.org/10.2967/jnumed.110.082404.

[51] J. J. M. van Griethuysen, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. G. H. Beets-Tan, J. C. Fillon-Robin, S. Pieper, and H. J. W. L. Aerts. Computational radiomics system to decode the radiographic phenotype. *Cancer Research*, 77(21):e104–e107, 2017. URL `https://doi.org/10.1158/0008-5472.CAN-17-0339`.

[52] Michael Galarnyk. Understanding boxplots, 2018. URL `https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51` (accessed: 01.05.2020).

[53] Khan Academy. Box plot review, n.d. URL `https://www.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/box-whisker-plots/a/box-plot-review` (accessed: 01.05.2020).

[54] N. A. Ahad and S. S. S. Yahaya. Sensitivity analysis of welch's t-test. *AIP Conference Proceedings*, 1605(1):888–893, 2014. doi: 10.1063/1.4887707. URL `https://aip.scitation.org/doi/abs/10.1063/1.4887707`.

[55] B. Derrick and P. White. Why welch's test is type i error robust. *The Quantitative Methods for Psychology*, 12(1):30–38, 2016. doi: 10.20982/tqmp.12.1.p030. URL `http://www.tqmp.org/RegularArticles/vol12-1/p030/p030.pdf`.

[56] G. D. Ruxton. The unequal variance t-test is an underused alternative to Student's t-test and the Mann–Whitney U test. *Behavioral Ecology*, 17(4):688–690, 05 2006. ISSN 1045-2249. doi: 10.1093/beheco/ark016. URL `https://doi.org/10.1093/beheco/ark016`.

[57] Tom A.B. Snijders. Statistical methods: Robustness, 2011. URL `http://www.stats.ox.ac.uk/~snijders/SM_robustness.pdf` (accessed: 10.05.2020).

[58] Marc Weber. *Statannot*, 2018. URL `https://github.com/webermarcolivier/statannot`. Accessed: 15.04.2020.

[59] Sebastian Raschka and Vahid Mirjalili. *Python Machine Learning*. Packt Publishing, second edition, 2017.

[60] B. M. Wise, N. B. Gallagher, R. Bro, J. M. Shaver, W. Windig, and R. Scott Koch. *Chemomterics Tutorial for PLS_Toolbox ans Solo*. Eigenvector Research, Manson, WA, USA, 2020. software available at `http://www.eigenvector.com`.

[61] Rasmus Bro and Age K. Smilde. Principal component analysis. *Analytical Methods*, 6:2812–2831, 03 2014. doi: 10.1039/C3AY41907J.

[62] Bent Jørgensen and Yuri Goegebeur. Department of statistics st02: Multivariate data analysis and chemometrics, 2007.

[63] Linh Ngo. How to read pca biplots and scree plots,

2018. URL https://blog.bioturing.com/2018/06/18/how-to-read-pca-biplots-and-scree-plots/ (accessed: 15.04.2020).

[64] MATLAB. *9.7.0.1296695 (R2019b)*. The MathWorks Inc., Natick, Massachusetts, 2019.

[65] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL https://www.R-project.org/.

[66] RStudio Team. *RStudio: Integrated Development Environment for R*. RStudio, Inc., Boston, MA, 2015. URL http://www.rstudio.com/.

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

[68] Trevor Hastie, Tibshirani Robert, and Friedman Jerome. *The elements of statistical learning*. Springer, second edition, 2001.

[69] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 3146–3154. Curran Associates, Inc., 2017. URL http://papers.nips.cc/paper/6907-lightgbm-a-highly-efficient-gradient-boosting-decision-tree.pdf.

[70] Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremly randomized trees. *Mach Learn*, 63:3–42, 2006. URL https://doi.org/10.1007/s10994-006-6226-1.

[71] Alex Zwanenburg. Radiomics in nuclear medicine: robustness, reproducibility, standardization, and how to avoid data analysis traps and replication crisis. *European Journal of Nuclear Medicine and Molecular Imaging*, 46:638–2655, 2019. doi: https://doi.org/10.1007/s00259-019-04391-8.

[72] Max Kuhn and Kjell Johnson. *Applied Predictive Modeling*. Springer, fifth edition, 2016.

[73] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *PHYSICAL REVIEW E*, 69, 2004.

[74] Jundong Li, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P. Trevino, Jiliang Tang, and Huan Liu. Feature selection: A data perspective. *ACM Comput. Surv.*, 50(6), December 2017. ISSN 0360-0300. doi: 10.1145/3136625. URL https://doi.org/10.1145/3136625.

[75] Karimollah Hajian-Tilaki. Receiver operating characteristic (roc) curve analysis for medical diagnostic test evaluation. *Physics in Medicine and Biology*, 4(2): 627–635, 2013.

[76] Vishwa S. Parekh and Michael A. Jacobs. Deep learning and radiomics in precision medicine. *Expert Review of Precision Medicine and Drug Development*, 4(2):59–72, 2019. doi: 10.1080/23808993.2019.1585805. URL https://doi.org/10.1080/23808993.2019.1585805. PMID: 31080889.

[77] Qi Feng, Mei Wang, Qiaowei Song, Zhengwang Wu, Hongyang Jiang, Peipei Pang, Zhengluan Liao, Enyan Yu, and Zhongxiang Ding. Correlation between hippocampus mri radiomic features and resting-state intrahippocampal functional connectivity in alzheimer's disease. *Frontiers in Neuroscience*, 13: 435, 2019. ISSN 1662-453X. doi: 10.3389/fnins.2019.00435. URL https://www.frontiersin.org/articles/10.3389/fnins.2019.00435.

[78] Ying Wang, Kai Sun, Zhenyu Liu, Guanmao Chen, Yanbin Jia, Shuming Zhong, Jiyang Pan, Li Huang, and Jie Tian. Classification of Unmedicated Bipolar Disorder Using Whole-Brain Functional Activity and Connectivity: A Radiomics Analysis. *Cerebral Cortex*, 30(3):1117–1128, 08 2019. ISSN 1047-3211. doi: 10.1093/cercor/bhz152. URL https://doi.org/10.1093/cercor/bhz152.

[79] Sara Ranjbar, Stefanie N. Velgos, Amylou C. Dueck, Yonas E. Geda, and J. Ross and Mitchell. Brain mr radiomics to differentiate cognitive disorders. *The Journal of Neuropsychiatry and Clinical Neurosciences*, 31(3):210–219, 2019. doi: 10.1176/appi.neuropsych.17120366. URL https://doi.org/10.1176/appi.neuropsych.17120366. PMID: 30636564.

Appendix A

Subject ID key

Table A.1: The original IDs of the ePOD-MPH study's trial subjects and the corresponding IDs used in this thesis.

(a) IDs of MPH treated participants

ePOD-MPH ID	Thesis ID
1	0
4	1
6	2
10	3
11	4
22	5
24	6
27	7
34	8
36	9
38	10
41	11
44	12
49	13
52	14
54	15
59	16
63	17
67	18
70	19
71	20
75	21

(b) IDs of placebo treated participants

ePOD-MPH ID	Thesis ID
3	22
7	23
8	24
17	25
18	26
20	27
21	28
29	29
32	30
33	31
37	32
39	33
42	34
45	35
47	36
50	37
51	38
56	39
62	40
64	41
68	42
69	43
72	44
74	45

Appendix B

The Autoscaling Program

```
1
2
3  import nibabel as nib
4  import numpy as np
5  import os
6  from copy import deepcopy
7  from scipy.stats import zscore
8  from skimage import img_as_ubyte
9  from skimage.exposure import rescale_intensity
10 from datetime import datetime
11
12
13
14 def import_data(img_path, affine_header = True):
15     """
16     Importing nifti images
17
18     :param img_path: path to folder containing nifti images
19     :param affine_header: Whether to store the affine
       transformation and the
20         header of the images
21
22     :return: list of 3D images. List of affine transformations and
       headers
23         are returned if affine_header is true
24     """
25
26     image_list = []
27     if affine_header:
28         image_affine = []
29         image_header = []
30
31
32     for dirname, _, filenames in os.walk(img_path):
33         for file in sorted(filenames):
34             img = nib.load(os.path.join(dirname, file))
35             image = img.get_fdata()
36             image_swap = np.transpose(image)
37             image_list.append(image_swap)
38             if affine_header:
39                 image_affine.append(img.affine)
```

```
                image_header.append(img.header)

    if affine_header:
        return image_list, image_affine, image_header
    else:
        return image_list

def dimension_3to1(image_list):
    """
    Tranfsorming list of 3D images to 1D

    :param image_list: List of 3D images
    :return: List of 1D images
    """

    D1_image = []
    shape = image_list[0].shape
    dim = shape[0]*shape[1]*shape[2]

    for element in image_list:
        D1_image.append(np.reshape(element, (dim,1)))

    return D1_image

def dimension_1to3(image_list):
    """
    Tranfsorming list of 1D images to 3D with shape (120, 256, 256)

    :param image_list: List of 1D images
    :return: List of 3D images
    """

    D3_image = []

    for element in image_list:
        D3_image.append(np.reshape(element, (120, 256, 256)))

    return D3_image

def create_segment(D1_image, D1_mask):
    """
    Creating a list of segments based on a list og 1D images and a
   list of
    1D binary masks

    :param D1_image: List of 1D images
    :param 1D1_mask: List of 1D binary masks

    :return: List of 1D segments
    """

    segment = []
    volume = 0

    for element in D1_image:
        val_ind = 0
        segment.append(np.zeros(0))

```

```
97          for value in element:
98              if D1_mask[volume][val_ind] == 1:
99
100                 segment[volume] = np.append(segment[volume], value)
101
102             val_ind += 1
103
104         volume += 1
105
106     return segment
107
108 def zscore_calc(segment_list):
109     """
110     Calculating the z-score of elements in a list
111
112     :param segment_list: List of 1D segments
113
114     :return: the normalised input list
115     """
116
117     segment_z = []
118     for ind in range(len(segment_list)):
119         segment_z.append(zscore(segment_list[ind]))
120
121     return segment_z
122
123 def threshold(D1_mask, segment):
124     """
125     Defining new binary masks. Voxels with z-score outside of range
    [-3,3] are
126     changed to 0 in new binary masks
127
128     :parma D1_mask: List of binary masks
129     :param segment: List of segments containing normalised values
    as z-scores
130
131     :return: List of new binary masks for segments
132     """
133
134     new_mask = []
135
136     for ind in range(len(D1_mask)):
137         new_mask.append(np.ones(int(D1_mask[ind].sum())))
138
139     for element in range(len(segment)):
140         for ind in range(len(segment[element])):
141             if segment[element][ind] <-3:
142                 new_mask[element][ind] = 0
143             elif segment[element][ind] > 3:
144                 new_mask[element][ind] = 0
145
146     return new_mask
147
148 def create_full_mask(new_mask, original_mask):
149     """
150     Creating full, new binary masks based on the new masks of the
    segments
151
```

```
152     :param new_mask: List of binary masks of segments
153     :param original_mask: List of the original masks of the
    imported images
154
155     :return: List of the new, full masks
156     """
157
158     full_mask = deepcopy(original_mask)
159     mask_copy = deepcopy(new_mask)
160
161     for ind in range(len(original_mask)):
162         for value in range(len(original_mask[ind])):
163             if original_mask[ind][value] == 1:
164                 full_mask[ind][value] = mask_copy[ind][0]
165                 mask_copy[ind] = np.delete(mask_copy[ind],[0])
166
167     return full_mask
168
169 def saving(D3_list, affine_list, header_list, path):
170     """
171     Saving 3D images.
172
173     :param 3D_list: List of 3D images
174     :param affine_list: List of affine transformations
175     :param header_list: List of headers
176     :param path: List of paths to save-folder
177     """
178
179     for index in range(len(D3_list)):
180         transposed = np.transpose(D3_list[index])
181         temp_seg = nib.Nifti1Image(transposed, affine_list[index],
    header_list[index])
182
183         if index < 10:
184             name = path+'/00'+str(index)+'_seg.nii'
185         elif index >= 10 and index < 100:
186             name = path + '/0'+str(index)+'_seg.nii'
187
188         nib.save(temp_seg, name)
189
190 def minimum(D1_image):
191     """
192     Detecting the minimium value in a list
193
194     :param D1_image: List of 1D images
195
196     :return: The minimum
197     """
198
199     min_seg = []
200
201     for element in D1_image:
202         min_seg.append(element.min())
203
204     global_min = 0
205     for element in min_seg:
206         if element<= global_min:
207             global_min = element
```

```
208
209     return global_min
210
211 def adjust_zscore(D1_zscore, global_min):
212     """
213     Adding the global minimun to the normalised images
214
215     :param D1_zscore: List of normalised 1D images
216     :param global_minimum: The global minimum
217
218     :return: List of shifted, normalised 1D images
219     """
220
221     seg_adjusted = []
222     for ind in range(len(D1_zscore)):
223         seg_adjusted.append(D1_zscore[ind] - global_min)
224
225     return seg_adjusted
226
227 def z_to_8bits(zscore_list):
228     """
229     Rescaling the normalised intensity values of 1D images to 1D 8-
    bit images
230
231     :param zscore_list: List of normalised 1D images
232
233     :return: List of 8-bit images
234     """
235
236     float_list = []
237     for ind in range(len(zscore_list)):
238         float_list.append(rescale_intensity(zscore_list[ind]))
239
240
241     uint8_list = []
242     for ind in range(len(float_list)):
243         uint8_list.append(img_as_ubyte(float_list[ind]))
244
245     return uint8_list
246
247 def segment_to_image(bits8_list, mask_list):
248     """
249     Transforming segments back to full 1D images based binary 1D
    masks
250
251     :param bits8_list: List of 8-bit segments
252     :param mask_list: List of binary 1D masks
253
254     :return: 1D images of the segments
255     """
256
257     uint8 = []
258     u8_copy = deepcopy(bits8_list)
259     dim=256*256*120
260
261     for ind in range(len(u8_copy)):
262         uint8.append(np.full(dim, 0))
263
```

```
264         for value in range(len(uint8[ind])):
265             if mask_list[ind][value] == 1:
266                 uint8[ind][value] = u8_copy[ind][0]
267                 u8_copy[ind] = np.delete(u8_copy[ind],[0])
268 
269     return uint8
270 
271 
272 if __name__=="__main__":
273     # The structures to be autoscaled:
274     brain_parts = ['caudate', 'hippocampus', 'pallidum', 'putamen',
    'thalamus' ]
275 
276 
277     # The autoscaling process:
278     for brain in brain_parts:
279 
280 
281         # The paths:
282         path_image=[r'D:\Master_2020\Data\Segmented data\children\
    MPH\image\bl\nifti',
283                     r'D:\Master_2020\Data\Segmented data\children\
    placebo\image\bl\nifti',
284                     r'D:\Master_2020\Data\Segmented data\children\
    MPH\image\bl\nifti',
285                     r'D:\Master_2020\Data\Segmented data\children\
    placebo\image\bl\nifti',
286                     r'D:\Master_2020\Data\Segmented data\children\
    MPH\image\pt\nifti',
287                     r'D:\Master_2020\Data\Segmented data\children\
    placebo\image\pt\nifti',
288                     r'D:\Master_2020\Data\Segmented data\children\
    MPH\image\pt\nifti',
289                     r'D:\Master_2020\Data\Segmented data\children\
    placebo\image\pt\nifti']
290 
291         path_mask = ['D:/Master_2020/Data/Segmented data/children/
    MPH/mask/left_'+brain+'/bl/nifti',
292                      'D:/Master_2020/Data/Segmented data/children/
    placebo/mask/left_'+brain+'/bl/nifti',
293                      'D:/Master_2020/Data/Segmented data/children/
    MPH/mask/right_'+brain+'/bl/nifti',
294                      'D:/Master_2020/Data/Segmented data/children/
    placebo/mask/right_'+brain+'/bl/nifti',
295                      'D:/Master_2020/Data/Segmented data/children/
    MPH/mask/left_'+brain+'/pt/nifti',
296                      'D:/Master_2020/Data/Segmented data/children/
    placebo/mask/left_'+brain+'/pt/nifti',
297                      'D:/Master_2020/Data/Segmented data/children/
    MPH/mask/right_'+brain+'/pt/nifti',
298                      'D:/Master_2020/Data/Segmented data/children/
    placebo/mask/right_'+brain+'/pt/nifti']
299 
300         end_path_image=['D:/Master_2020/Data/autoscaled_data/'+
    brain+'/left/mph/bl/image',
301                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/left/placebo/bl/image',
302                         'D:/Master_2020/Data/autoscaled_data/'+
```

```
    brain+'/right/mph/bl/image',
303                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/right/placebo/bl/image',
304                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/left/mph/pt/image',
305                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/left/placebo/pt/image',
306                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/right/mph/pt/image',
307                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/right/placebo/pt/image']
308
309         end_path_mask=['D:/Master_2020/Data/autoscaled_data/'+brain
    +'/left/mph/bl/mask',
310                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/left/placebo/bl/mask',
311                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/right/mph/bl/mask',
312                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/right/placebo/bl/mask',
313                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/left/mph/pt/mask',
314                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/left/placebo/pt/mask',
315                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/right/mph/pt/mask',
316                         'D:/Master_2020/Data/autoscaled_data/'+
    brain+'/right/placebo/pt/mask']
317
318
319         path_dict = {0:[], 1:[], 2:[], 3:[], 4:[], 5:[], 6:[],
    7:[]}
320
321         # 1) Creating segments and new binary masks
322         for path in range(len(path_image)):
323             print(path_mask[path])
324
325             # 1.1) Importing images and masks
326             org_image_list, org_image_affine, org_image_header =
    import_data(path_image[path])
327             org_mask_list, org_mask_affine, org_mask_header =
    import_data(path_mask[path])
328
329             d1_org_img = dimension_3to1(org_image_list)
330             d1_org_mask = dimension_3to1(org_mask_list)
331
332             # 1.2) Creating segments of the brain structures
333             segment_list = create_segment(D1_image=d1_org_img,
    D1_mask=d1_org_mask)
334
335             # 1.3) Normalising and exporting the segments
336             z_segments = zscore_calc(segment_list)
337             z_cut_mask = threshold(d1_org_mask, z_segments)
338
339             d1_mask2 = create_full_mask(z_cut_mask, d1_org_mask)
340             d3_mask2 = dimension_1to3(d1_mask2)
341             segment_list2 = create_segment(D1_image=z_segments,
    D1_mask=z_cut_mask)
```

```
342
343             path_dict[path].append(segment_list2)
344             path_dict[path].append(d1_mask2)
345             path_dict[path].append(org_image_affine)
346             path_dict[path].append(org_image_header)
347
348             saving(d3_mask2, org_mask_affine, org_mask_header,
    end_path_mask[path])
349
350         # 2) Identifying the global minimum
351
352         global_min = 1000
353         for path in path_dict:
354             new_min = minimum(path_dict[path][0])
355             if new_min < global_min:
356                 global_min = new_min
357
358         # 3) Shifting and rescaling the segments to 8-bit images
359
360         for path in range(len(end_path_image)):
361             segment_adjusted = adjust_zscore(path_dict[path][0],
    global_min)
362             segment_uint8 = z_to_8bits(segment_adjusted)
363             uint8_list = segment_to_image(segment_uint8, path_dict[
    path][1])
364             d3_uint8 = dimension_1to3(uint8_list)
365
366             saving(d3_uint8, path_dict[path][2], path_dict[path
    ][3], end_path_image[path])
367
368             print(datetime.now().time())
```

Appendix C

The Change in Surface Area Due to Voxel Exclusion

The effect of the voxel exclusion performed as part of the autoscaling process, was evaluated by illustrating the change in surface area of the ROIs. In order to be able to compare the change in surface area across the ROIs, the average change (black lines) and three standrad deviations (red lines) were marked. A closer examination of the figures was performed in Chapter 4.1.1.

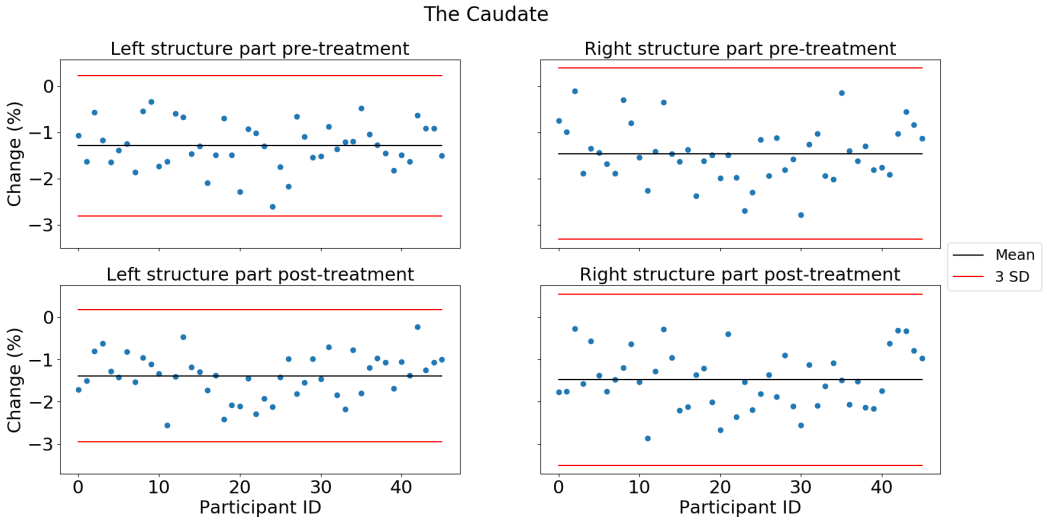


Figure C.1: The figure illustrates the change in surface area of the left and right caudate due to artefact voxel exclusion. The mean of the percentage change of the structures (black line) and three standard deviations (red lines) were included in the plots. On average, the subjects' caudate changed by approximately 1.5 %.

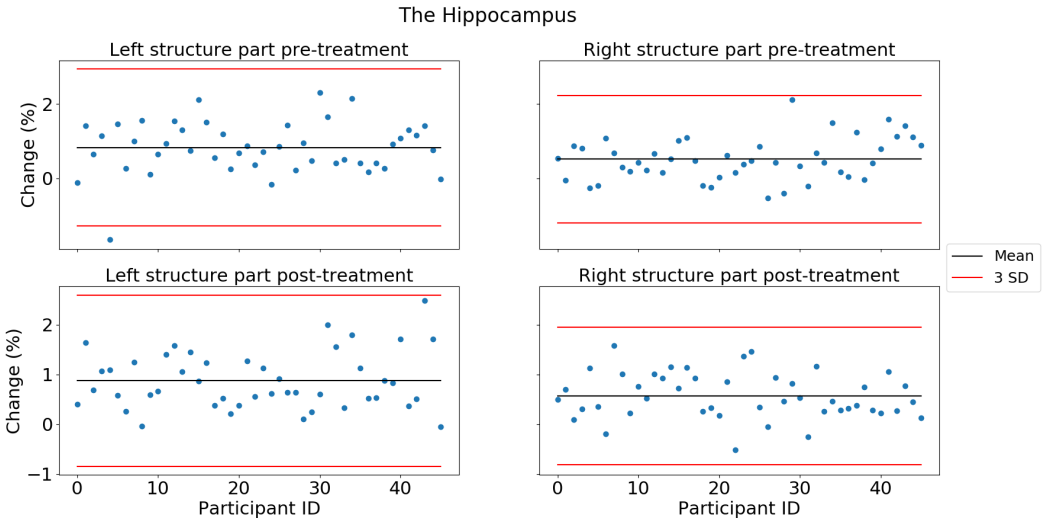


Figure C.2: The figure illustrates the change in surface area of the left and right hippocampus due to artefact voxel exclusion. The mean of the percentage change of the structures (black line) and three standard deviations (red lines) were included in the plots. On average, the subjects' hippocampus changed by approximately 1 %. Subject nr. 4 deviated from the remaining subjects as its left hippocampus was reduced by more than three standard deviations pre-treatment.

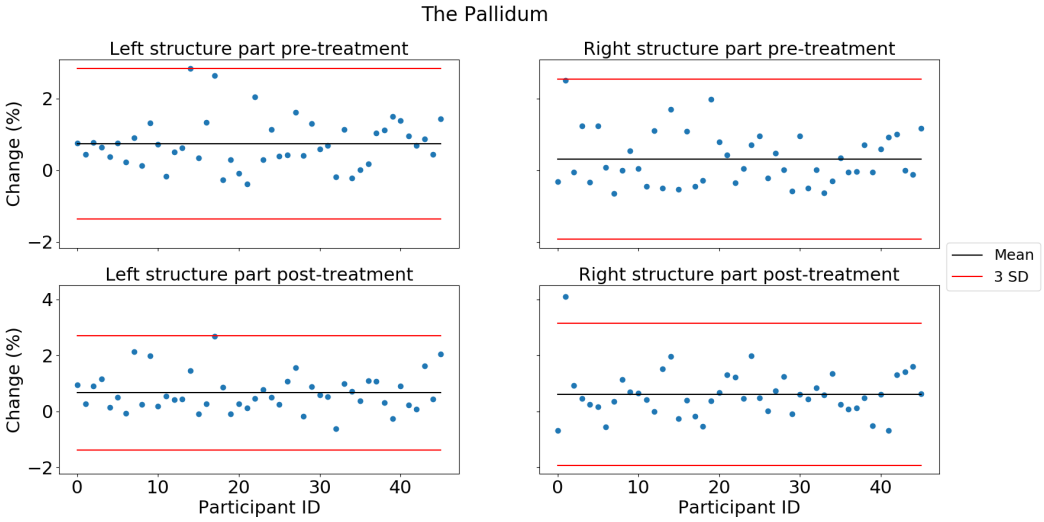


Figure C.3: The figure illustrates the change in surface area of the left and right pallidum due to artefact voxel exclusion. The mean of the percentage change of the structures (black line) and three standard deviations (red lines) were included in the plots. On average, the subjects' pallidum changed by approximately 1 %. Subject nr. 1 deviated from the remaining subjects as its right pallidum increased by more than three standard deviations post-treatment.

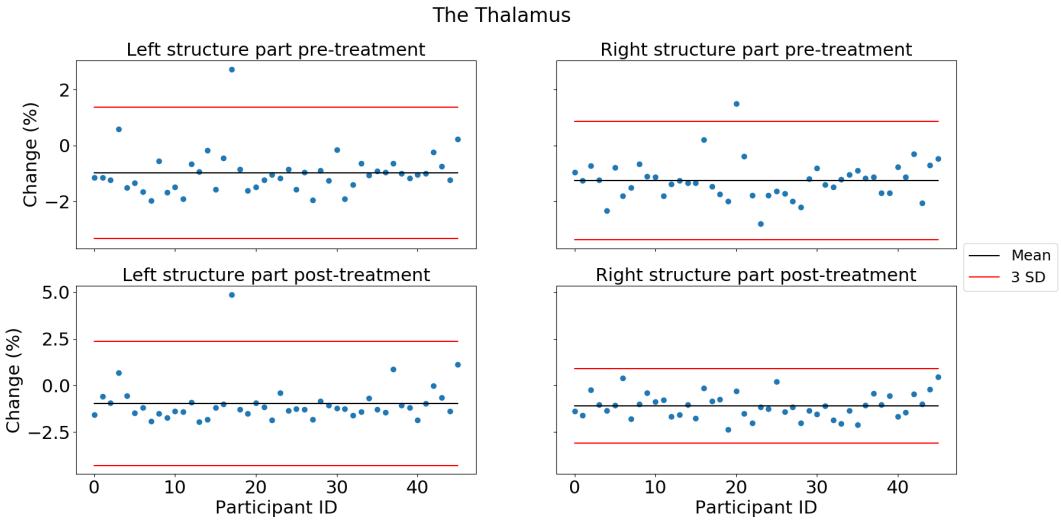


Figure C.4: The figure illustrates the change in surface area of the left and right thalamus due to artefact voxel exclusion. The mean of the percentage change of the structures (black line) and three standard deviations (red lines) were included in the plots. On average, the subjects' thalamus changed by approximately 1 %. Subject nr. 17 deviated from the remaining subjects as its left thalamus increased by more than three standard deviations pre- and post-treatment. The same apllied for the right thalamus of subject nr. 20 pre-treatment.

Appendix D

The Effect of Intensity Value Normalisation

In order to examine the effect of the autoscaling process on the ROIs' intensity level distribution, the distributions from before and after normalisation were plotted. In general, the normalisation and rescaling of the images resulted in approximate alignment of the intensity values across the subjects, as seen in the appendices' figures.

The Caudate

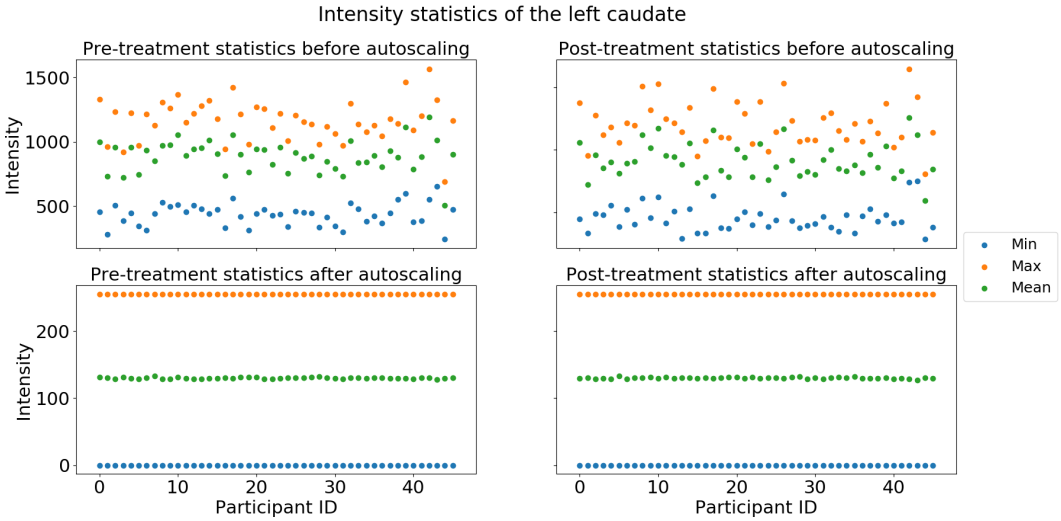


Figure D.1: The change of the intensity distribution of the left caudate due to normalisation. The top and bottom rows display the minimum, maximum and mean intensity values of the ROI across the subjects before and after normalisation, respectively. The intensity levels of the subjects were normalised and rescaled to a common intensity level range [0, 255]. After normalisation the maximum and minimum intensity values were equal across all subjects, and the mean values approximately equal.

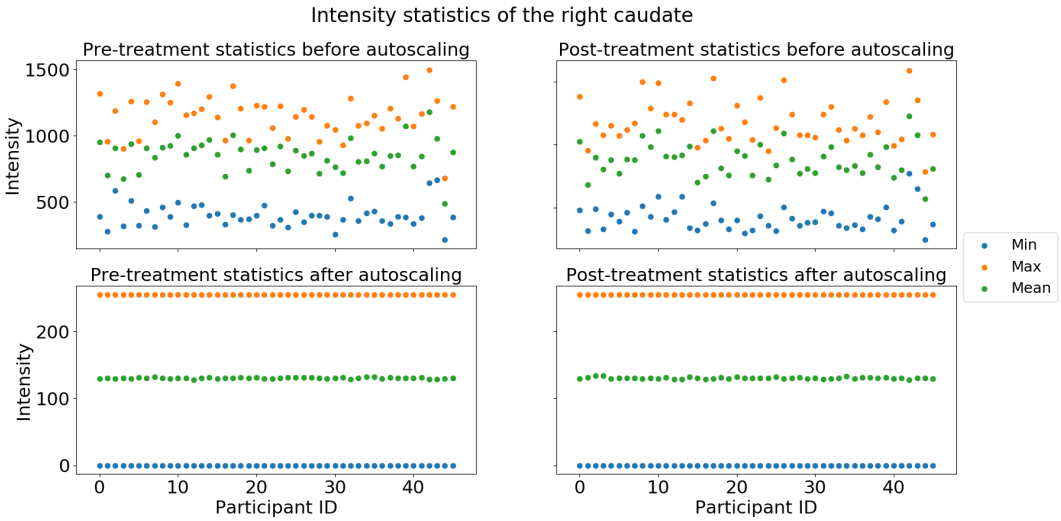


Figure D.2: The change of the intensity distribution of the right caudate due to normalisation. The top and bottom rows display the minimum, maximum and mean intensity values of the ROI across the subjects before and after normalisation, respectively. The intensity levels of the subjects were normalised and rescaled to a common intensity level range [0, 255]. After normalisation the maximum and minimum intensity values were equal across all subjects, and the mean values approximately equal.

The Hippocampus

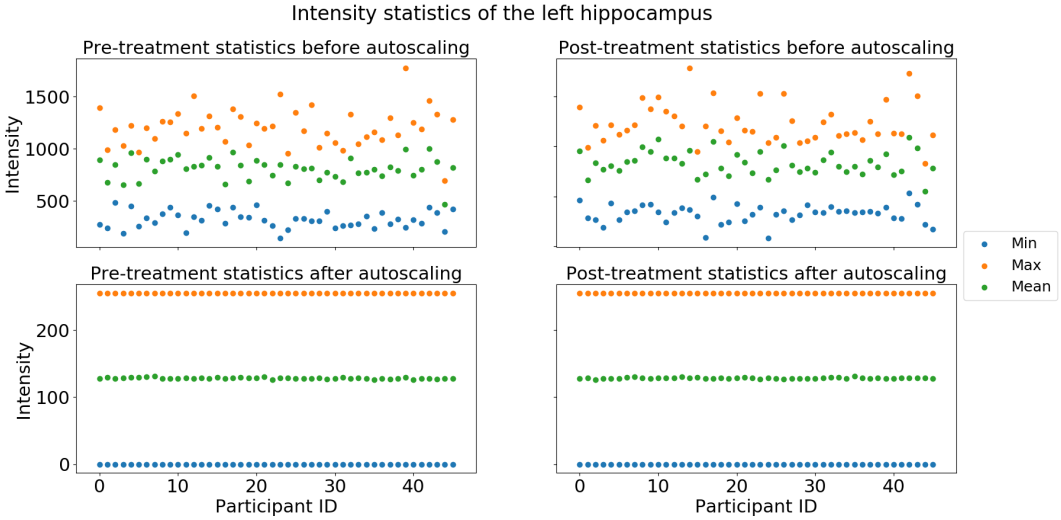


Figure D.3: The change of the intensity distribution of the left hippocampus due to normalisation. The top and bottom rows display the minimum, maximum and mean intensity values of the ROI across the subjects before and after normalisation, respectively. The intensity levels of the subjects were normalised and rescaled to a common intensity level range [0, 255]. After normalisation the maximum and minimum intensity values were equal across all subjects, and the mean values approximately equal.

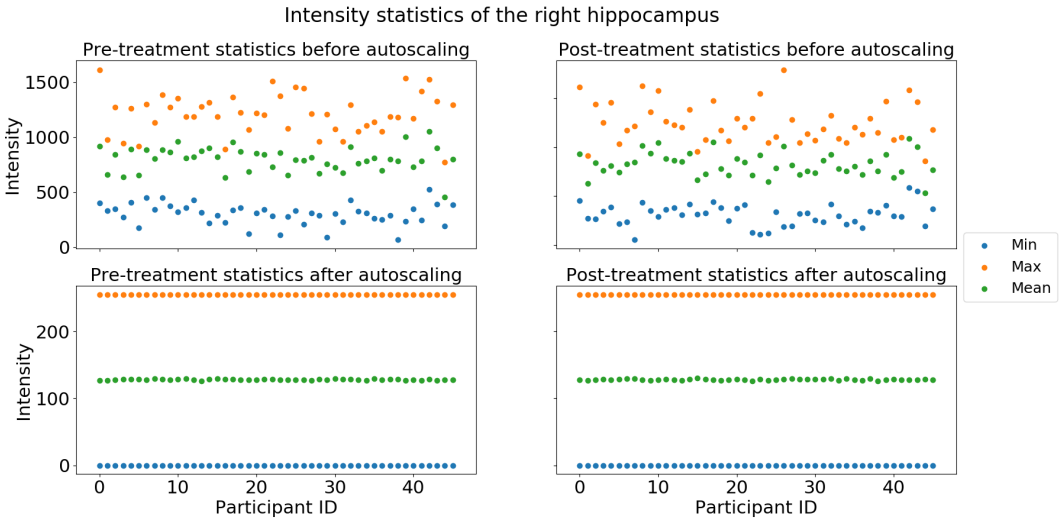


Figure D.4: The change of the intensity distribution of the right hippocampus due to normalisation. The top and bottom rows display the minimum, maximum and mean intensity values of the ROI across the subjects before and after normalisation, respectively. The intensity levels of the subjects were normalised and rescaled to a common intensity level range [0, 255]. After normalisation the maximum and minimum intensity values were equal across all subjects, and the mean values approximately equal.

The Pallidum

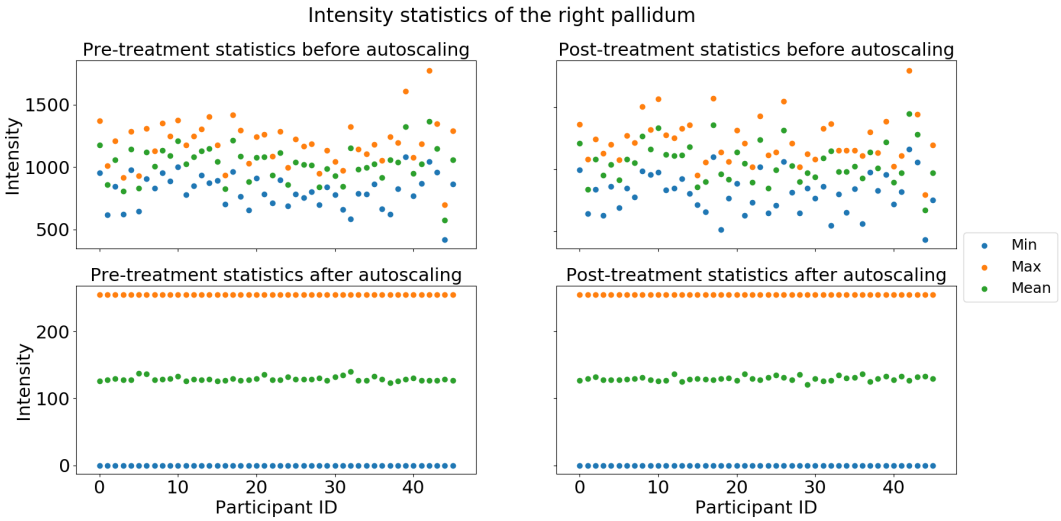


Figure D.5: The change of the intensity distribution of the right pallidum due to normalisation. The top and bottom rows display the minimum, maximum and mean intensity values of the ROI across the subjects before and after normalisation, respectively. The intensity levels of the subjects were normalised and rescaled to a common intensity level range [0, 255]. After normalisation the maximum and minimum intensity values were equal across all subjects, and the mean values approximately equal.

The Putamen

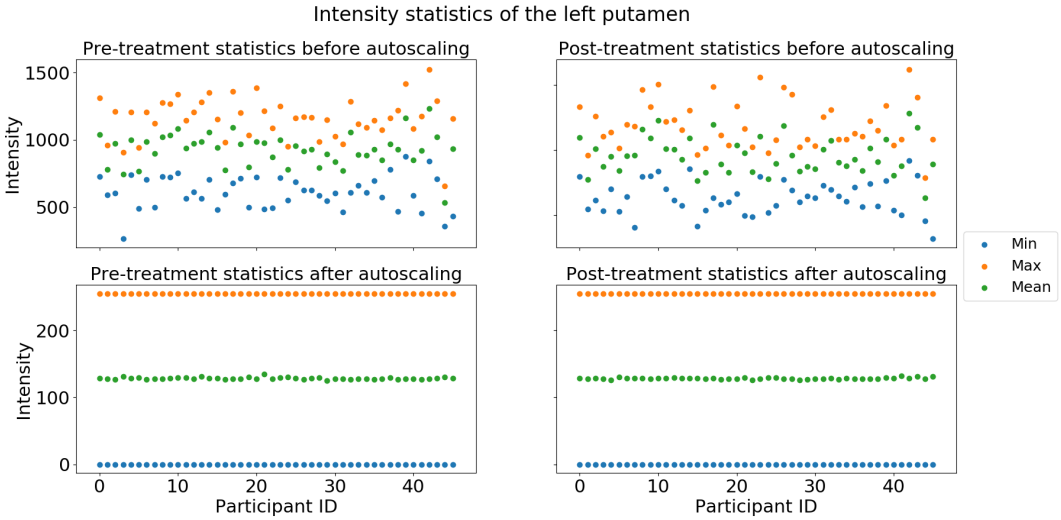


Figure D.6: The change of the intensity distribution of the left putamen due to normalisation. The top and bottom rows display the minimum, maximum and mean intensity values of the ROI across the subjects before and after normalisation, respectively. The intensity levels of the subjects were normalised and rescaled to a common intensity level range [0, 255]. After normalisation the maximum and minimum intensity values were equal across all subjects, and the mean values approximately equal.

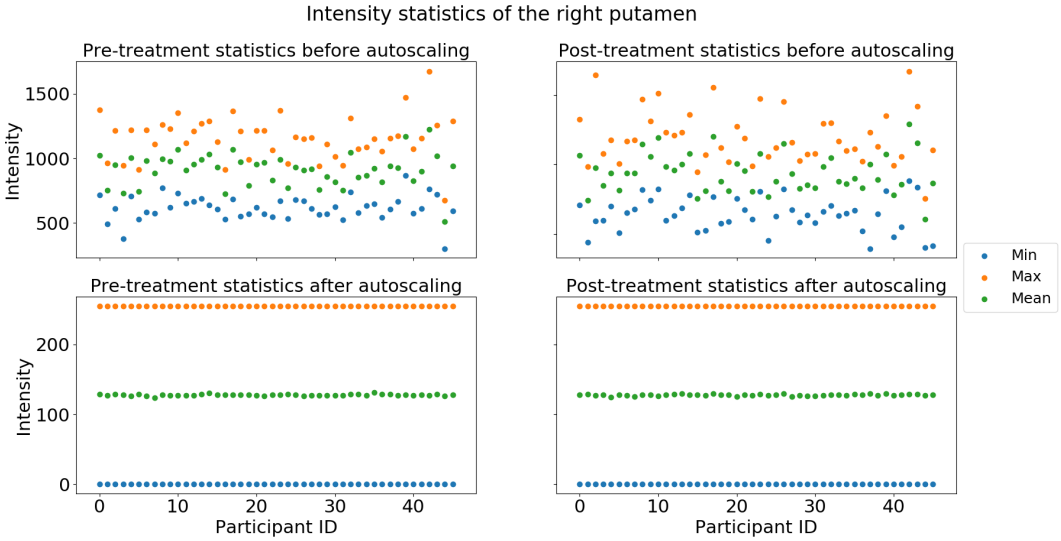


Figure D.7: The change of the intensity distribution of the right putamen due to normalisation. The top and bottom rows display the minimum, maximum and mean intensity values of the ROI across the subjects before and after normalisation, respectively. The intensity levels of the subjects were normalised and rescaled to a common intensity level range [0, 255]. After normalisation the maximum and minimum intensity values were equal across all subjects, and the mean values approximately equal.

The Thalamus

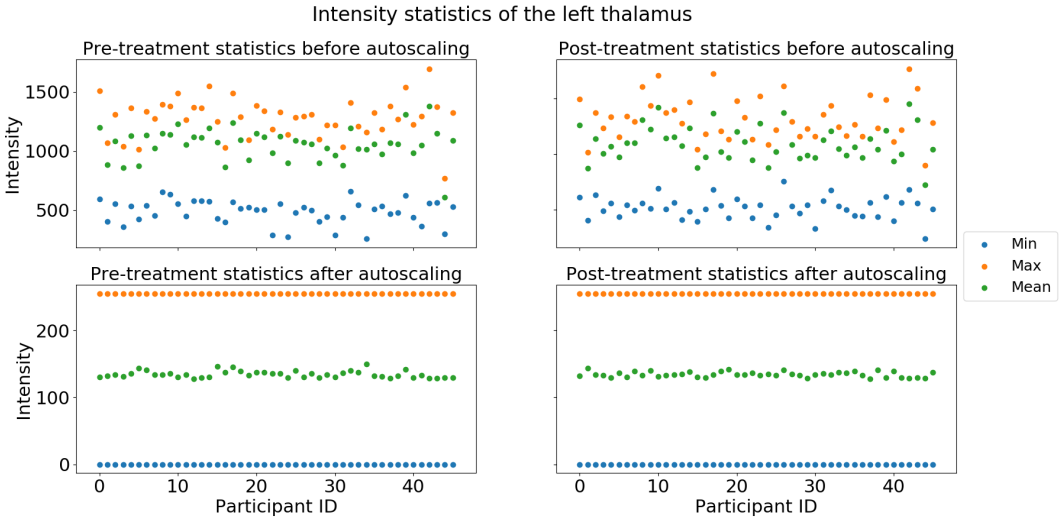


Figure D.8: The change of the intensity distribution of the left thalamus due to normalisation. The top and bottom rows display the minimum, maximum and mean intensity values of the ROI across the subjects before and after normalisation, respectively. The intensity levels of the subjects were normalised and rescaled to a common intensity level range [0, 255]. After normalisation the maximum and minimum intensity values were equal across all subjects, and the mean values approximately equal.

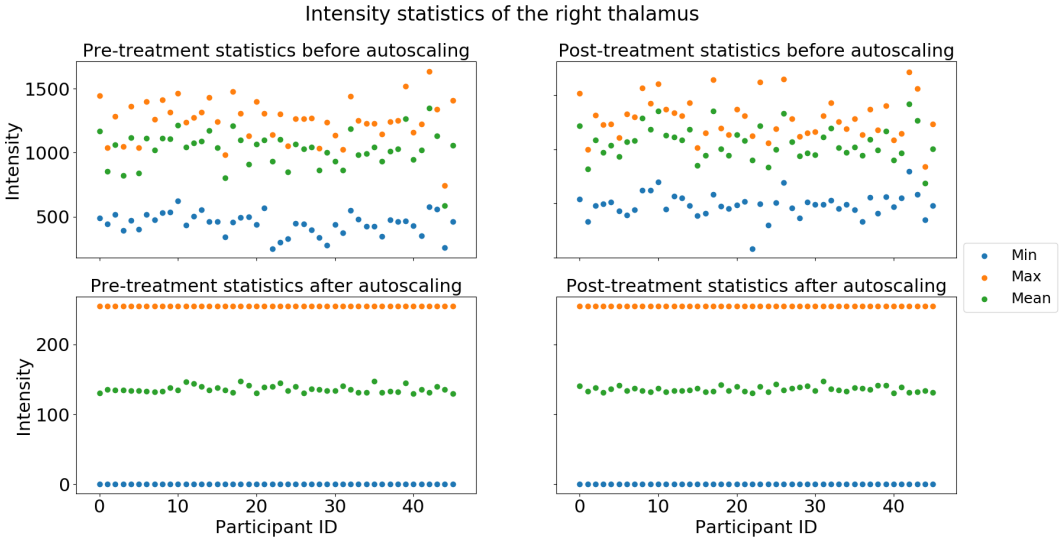


Figure D.9: The change of the intensity distribution of the right thalamus due to normalisation. The top and bottom rows display the minimum, maximum and mean intensity values of the ROI across the subjects before and after normalisation, respectively. The intensity levels of the subjects were normalised and rescaled to a common intensity level range [0, 255]. After normalisation the maximum and minimum intensity values were equal across all subjects, and the mean values approximately equal.

Appendix E

Evaluating the PCA Models of the Hippocampus, Pallidum, Putamen and Thalamus

E.1 The Hippocampus

The RMSECV and RMSEC curves of the hippocampus' shape feature PCA model are shown in Figure E.1, indicated strong overfitting. Based on the model's statistics in Figure F.4, seven components were included in the shape feature model. PLS-Toolbox suggested using seven and five PCs for the 128-bin and 64-bin texture feature, respectively. The suggestions were implemented.

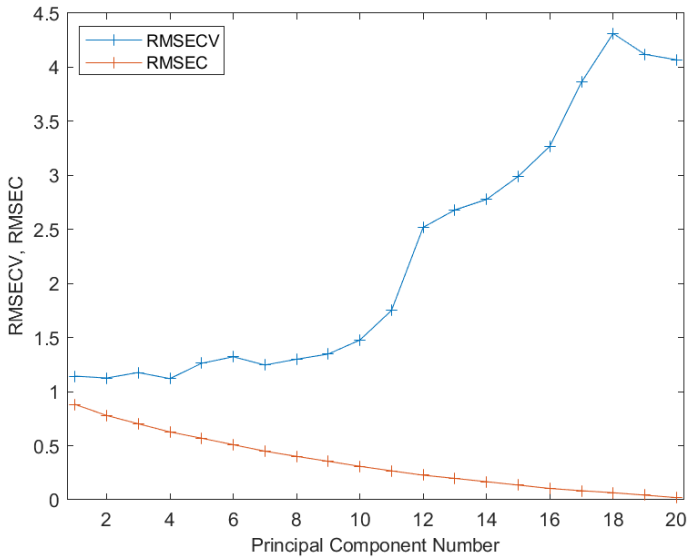


Figure E.1: The RMSEC and RMSECV curve of the hippocampus' shape feature model. While the RMSEC curve decreased constantly, the RMSECV increased after four components, indicating the model overfited stronlgy.

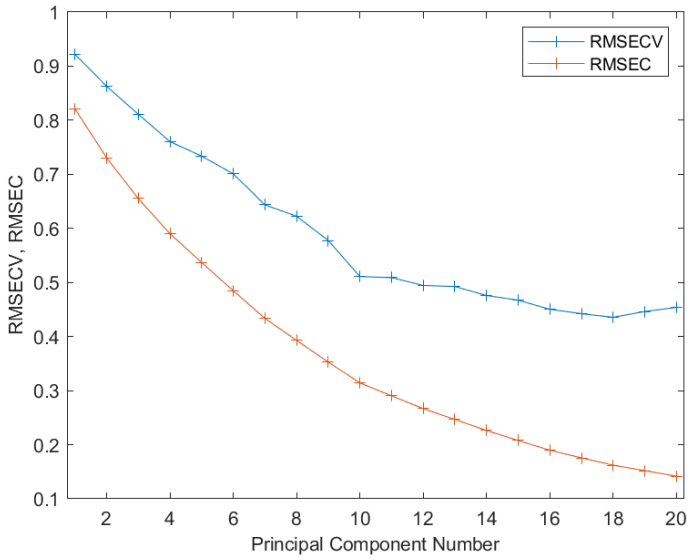


Figure E.2: The RMSEC and RMSECV curve of the hippocampus' 128-bin texture feature model. Both curves decreased constantly, until the RMSECV curve flattened out slightly.

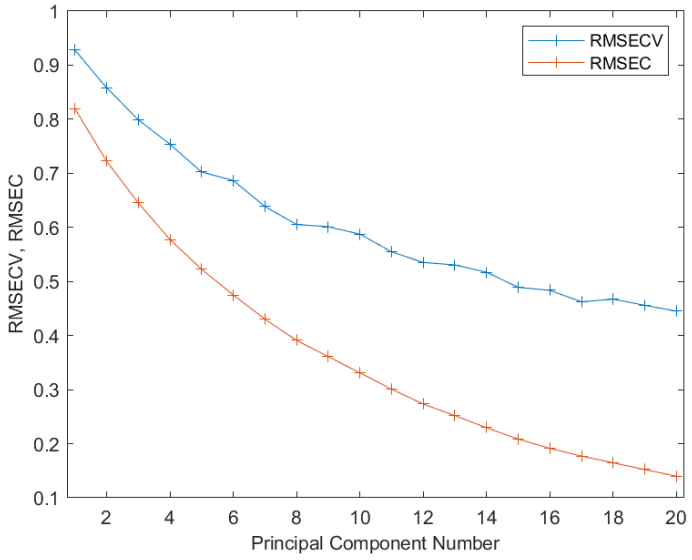


Figure E.3: The RMSEC and RMSECV curve of the hippocampus' 64-bin texture feature model. Both curves decreased constantly, until the RMSECV curve flattened out slightly.

E.1.1 Pattern Detection in the Score Plots

Identical to the caudate, no characteristics were identified in the corresponding score plots in Figures E.4-E.8.

The Shape Feature-based PCA model

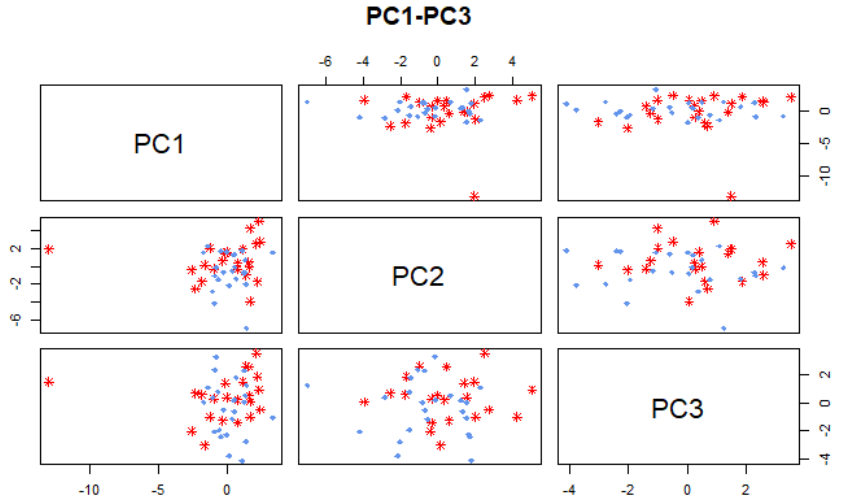


Figure E.4: The hippocampus' pair plot of the shape feature principal components nr 1-3. No principal components discriminating the treatment groups were detected.

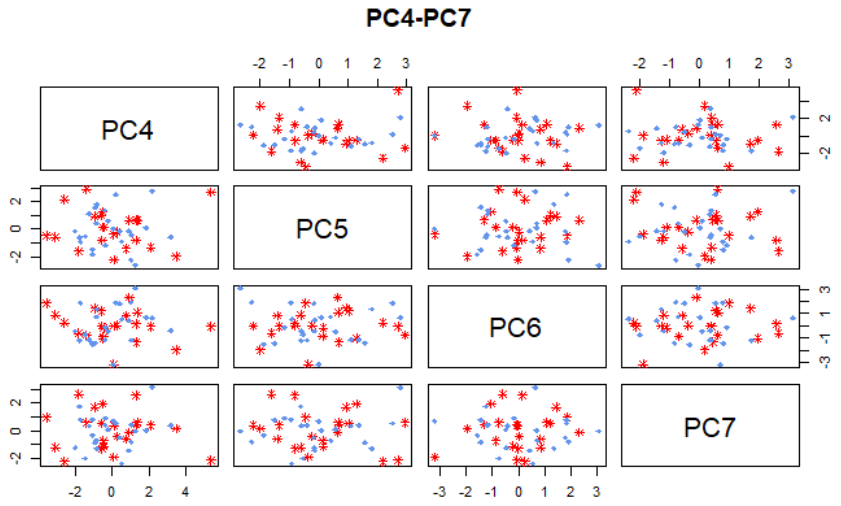


Figure E.5: The hippocampus' pair plot of the shape feature principal components nr 4-7. No principal components discriminating the treatment groups were detected.

The 128-bin Texture Feature-based PCA model

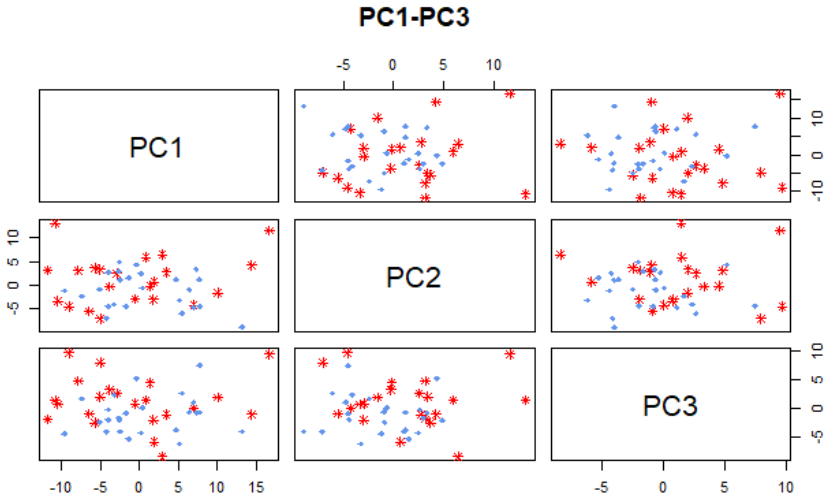


Figure E.6: The hippocampus' pair plot of the 128-bin texture feature principal components nr 1-3. No principal components discriminating the treatment groups were detected.

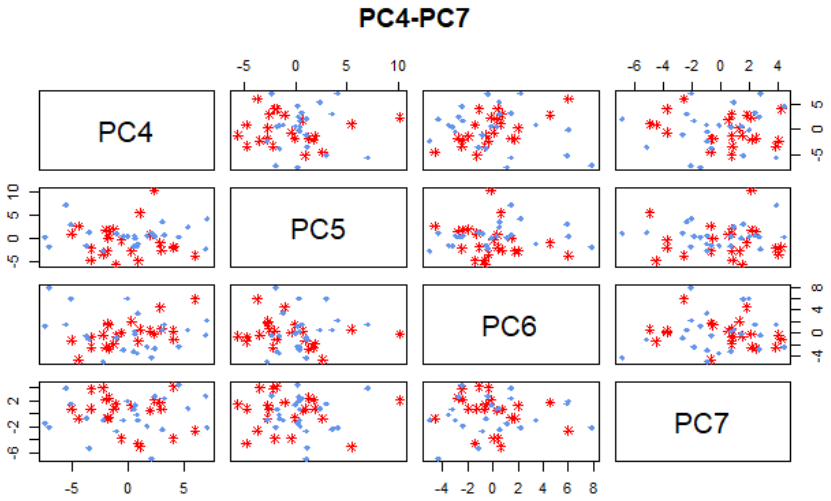


Figure E.7: The hippocampus' pair plot of the 128-bin texture feature principal components nr 4-7. No principal components discriminating the treatment groups were detected.

The 64-bin Texture Feature-based PCA model

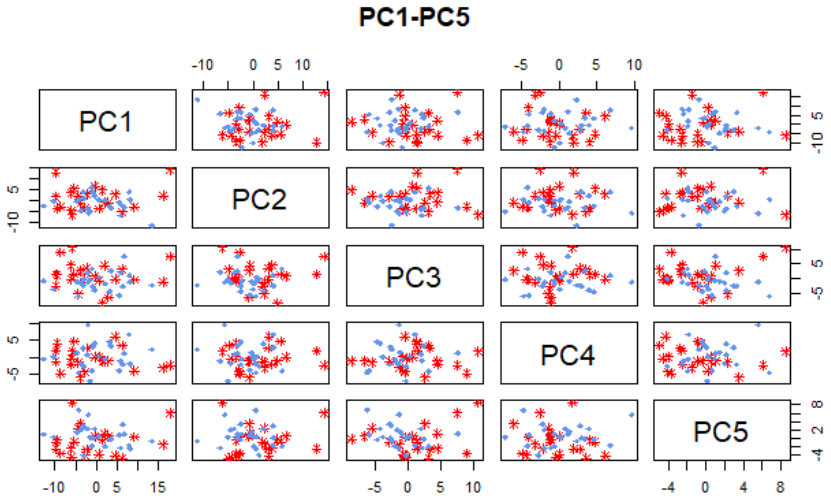


Figure E.8: The hippocampus' pair plot of the 64-bin texture feature principal components nr 1-5. No principal components discriminating the treatment groups were detected.

E.1.2 Identifying Possible Outliers in the Score and Q-T Plots

Eight, ten and five subjects from class 1 stood out in the shape, 128-bin texture and 64-bin texture feature model, respectively. In the same models, five, seven and six placebo-treated subjects were selected and listed in Table H.1. From the detected subjects, nr. 3, 4, 7 and 29 stood out in all mentioned models. Out of these, only nr. 4 should be considered as a possible outlier according to the Q - T plots in Figures E.9-E.11.

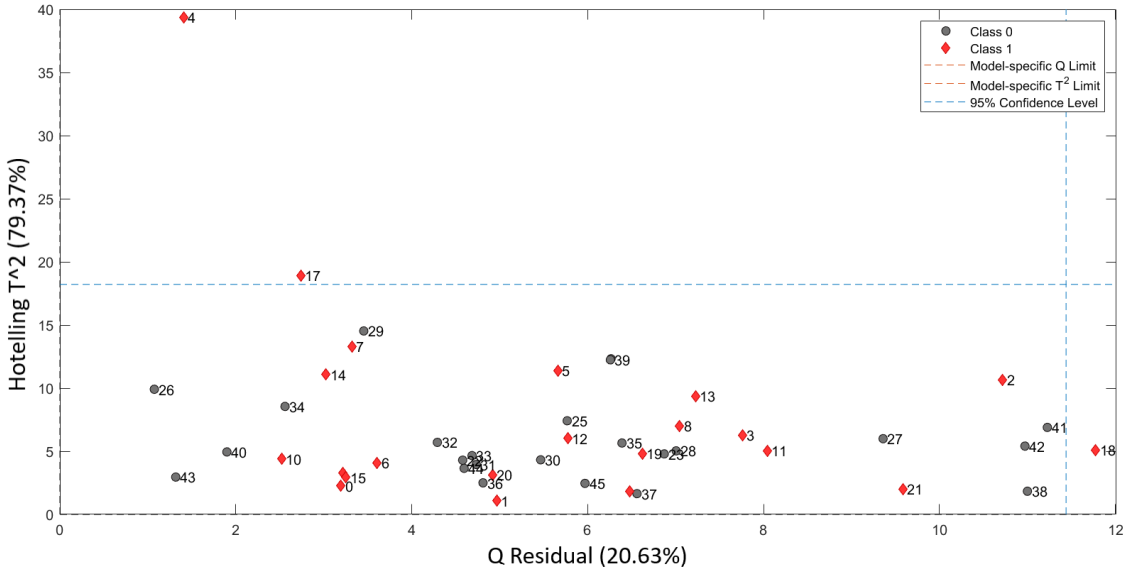


Figure E.9: The Q-T plot of the hippocampus' shape feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

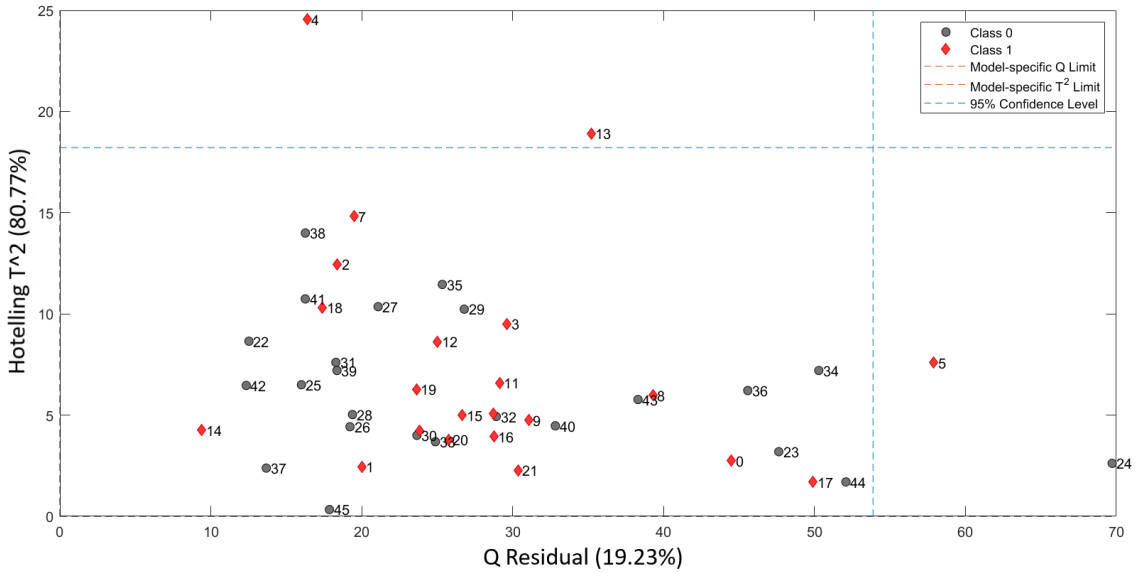


Figure E.10: The Q-T plot of the hippocampus' 128-bin texture feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

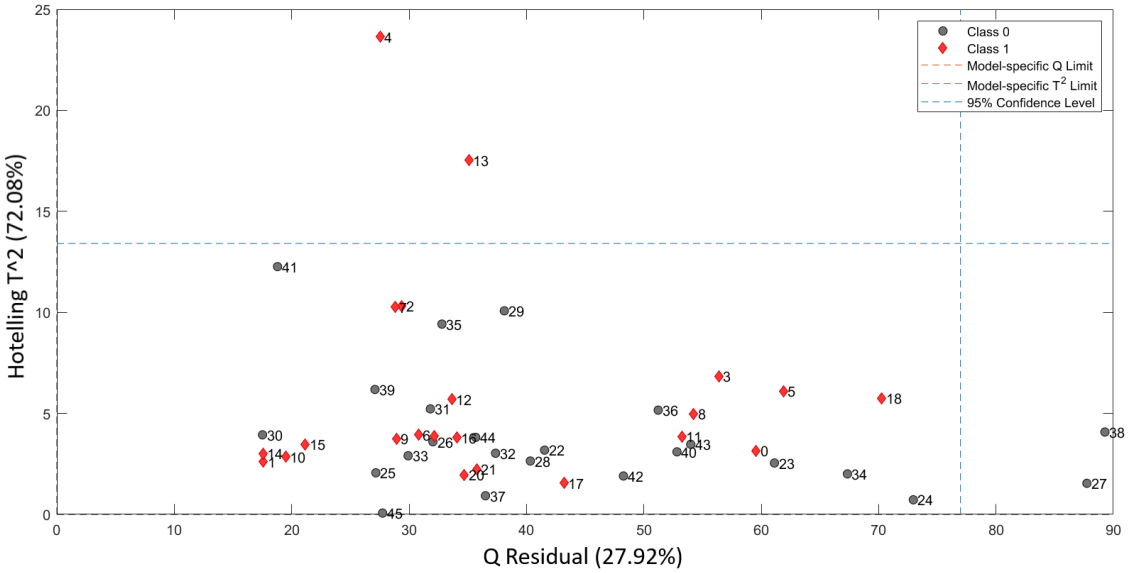


Figure E.11: The Q-T plot of the hippocampus' 64-bin texture feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

E.2 The Pallidum

PLS-Toolbox recommends one PC for the pallidum shape feature model, eleven for the 128-bin texture feature model and eight for the 64-bin texture feature model. The suggestions for the texture features were implemented. However, the shape feature model's "cutoff"-component was set to six. The decision was grounded in the model's RMSEC in Figure E.12, where the sixth component corresponds to a local minimum.

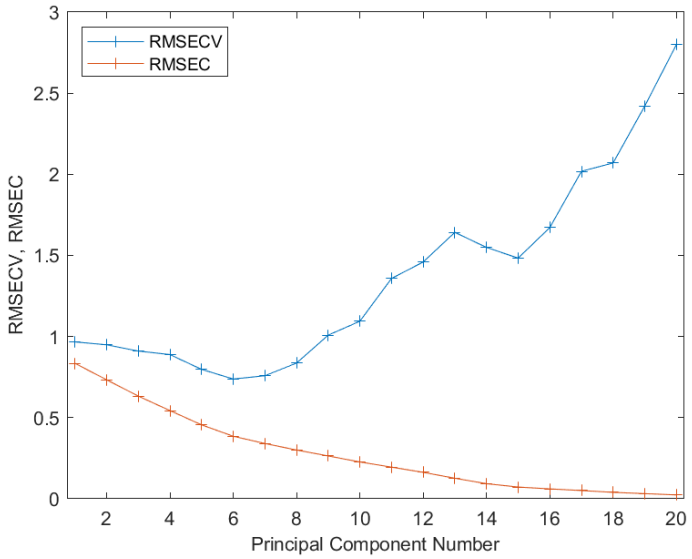


Figure E.12: The RMSEC and RMSECV curve of the pallidum's shape feature model. While the RMSEC curve decreased constantly, the RMSECV increased after six components, indicating the model overfited stronlgy.

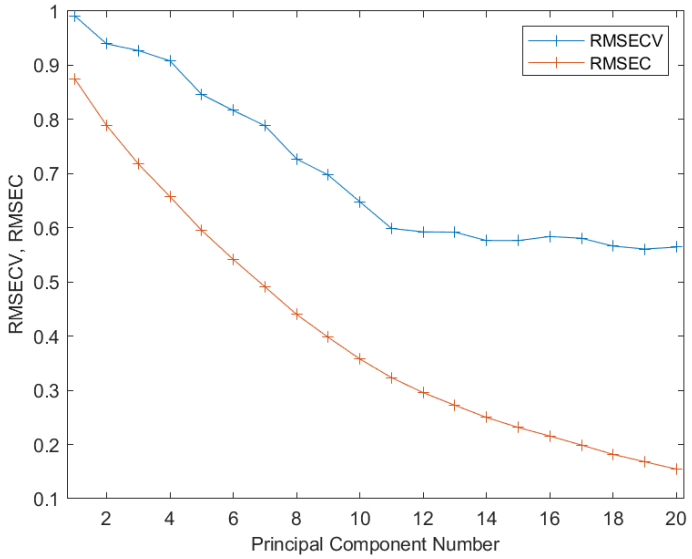


Figure E.13: The RMSEC and RMSECV curve of the pallidum's 128-bin texture feature model. Both curves decreased constantly, until the RMSECV curve flattened out slightly.

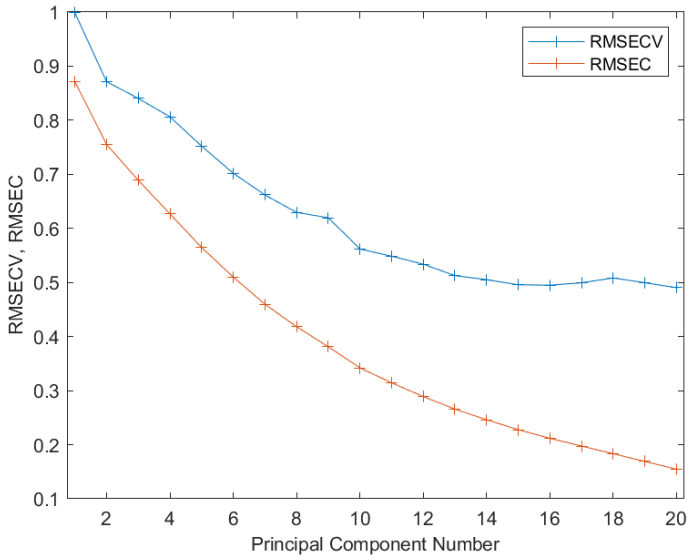


Figure E.14: The RMSEC and RMSECV curve of the pallidum's 64-bin texture feature model. Both curves decreased constantly, until the RMSECV curve flattened out slightly.

E.2.1 Pattern Detection in the Score Plots

No principle components discriminating the treatment groups were identified in the pallidum's score plots in Figures E.15-E.21 by visual examination.

The Shape Feature-based PCA model

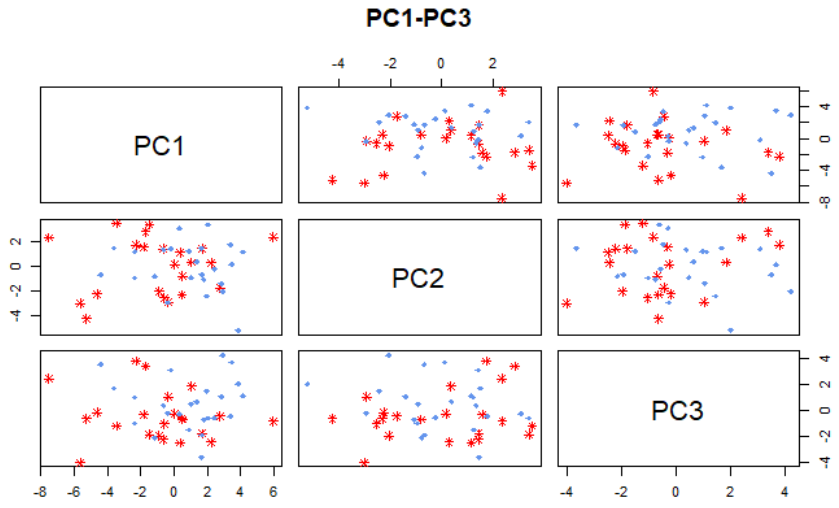


Figure E.15: The pallidum's pair plot of the shape feature principal components nr 1-3. No principal components discriminating the treatment groups were detected.

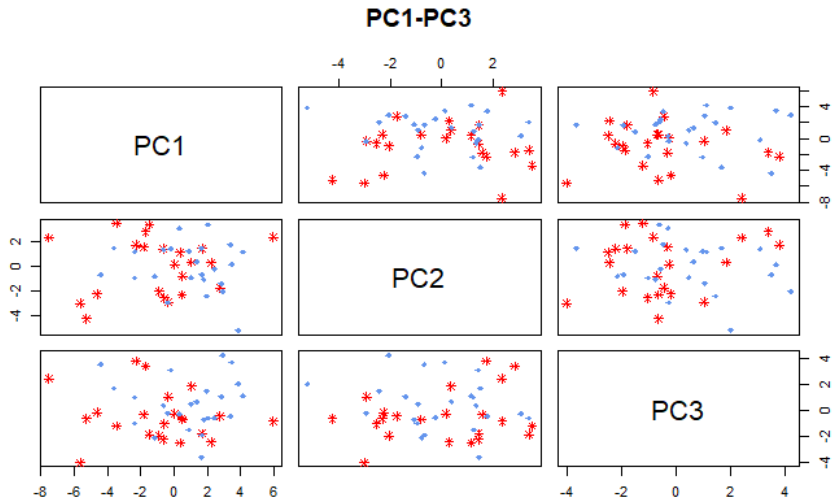


Figure E.16: The pallidum's pair plot of the shape feature principal components nr 4-6. No principal components discriminating the treatment groups were detected.

The 128-bin Texture Feature-based PCA model

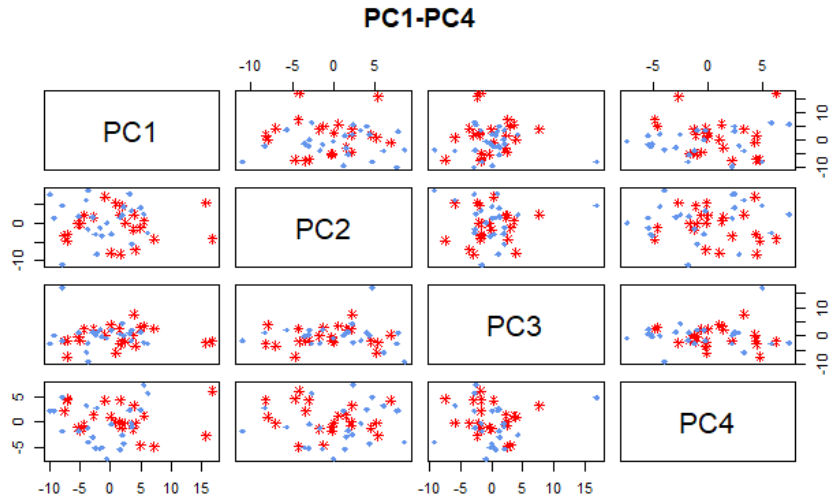


Figure E.17: The pallidum's pair plot of the 128-bin texture feature principal components nr 1-4. No principal components discriminating the treatment groups were detected.

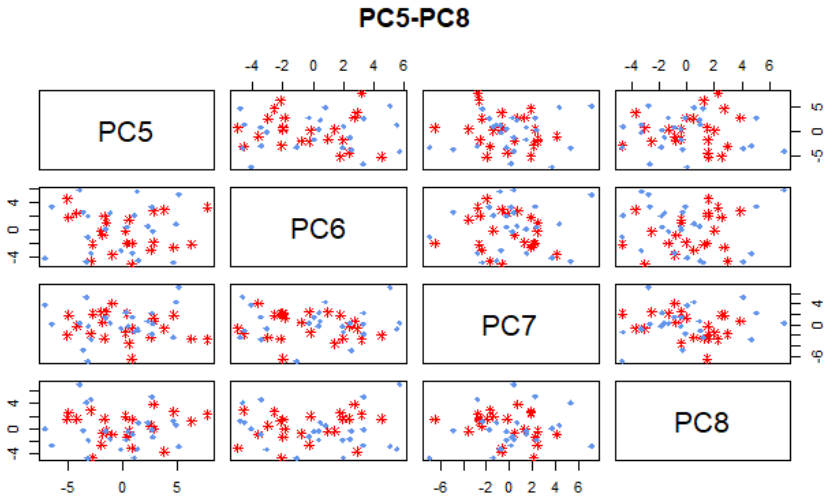


Figure E.18: The pallidum's pair plot of the 128-bin texture feature principal components nr 5-8. No principal components discriminating the treatment groups were detected.

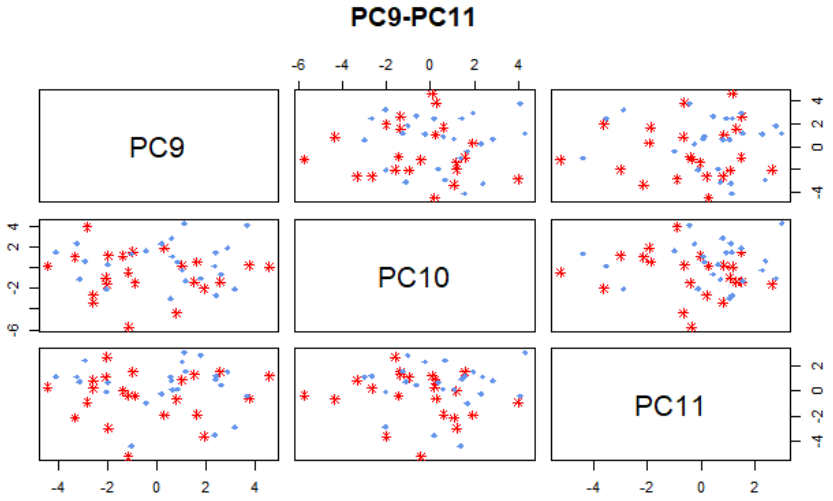


Figure E.19: The pallidum's pair plot of the 128-bin texture feature principal components nr 9-11. No principal components discriminating the treatment groups were detected.

The 64-bin Texture Feature-based PCA model

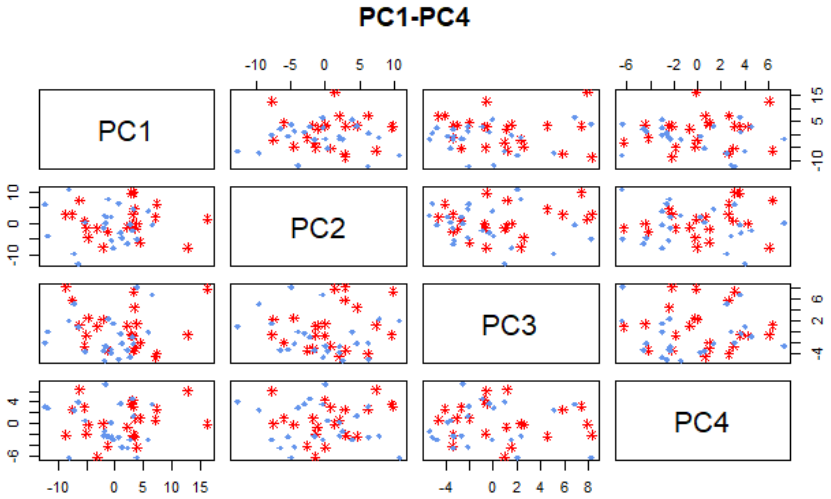


Figure E.20: The pallidum's pair plot of the 64-bin texture feature principal components nr 1-4. No principal components discriminating the treatment groups were detected.

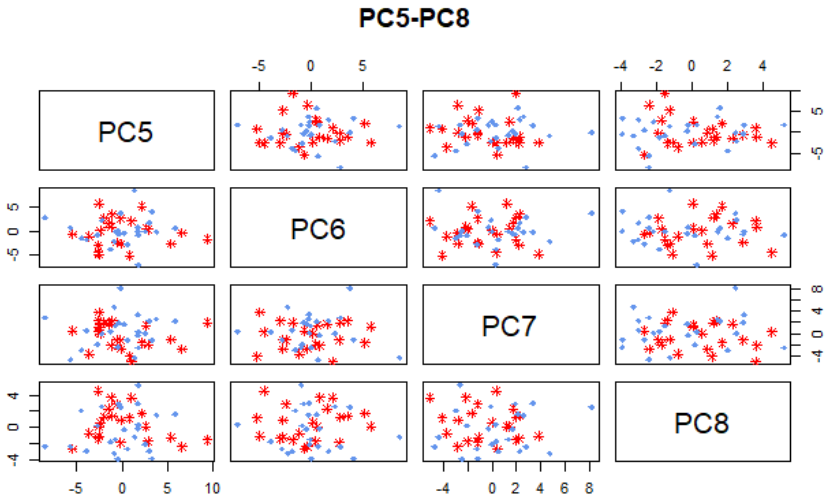


Figure E.21: The pallidum's pair plot of the 64-bin texture feature principal components nr 5-8. No principal components discriminating the treatment groups were detected.

E.2.2 Identifying Possible Outliers in the Score and Q-T Plots

Ten subjects stood out in the shape feature model, three of them from the placebo group. Eight placebo-treated and ten MPH-treated subjects were selected in the 64-bin texture feature model. The 128-bin texture feature model highlighted eight subjects, evenly distributed between the groups. Subjects nr. 5 and 29 were selected in all models. Out of these two, only nr. 5 was classified as a possible outlier by the

corresponding Q - T plots.

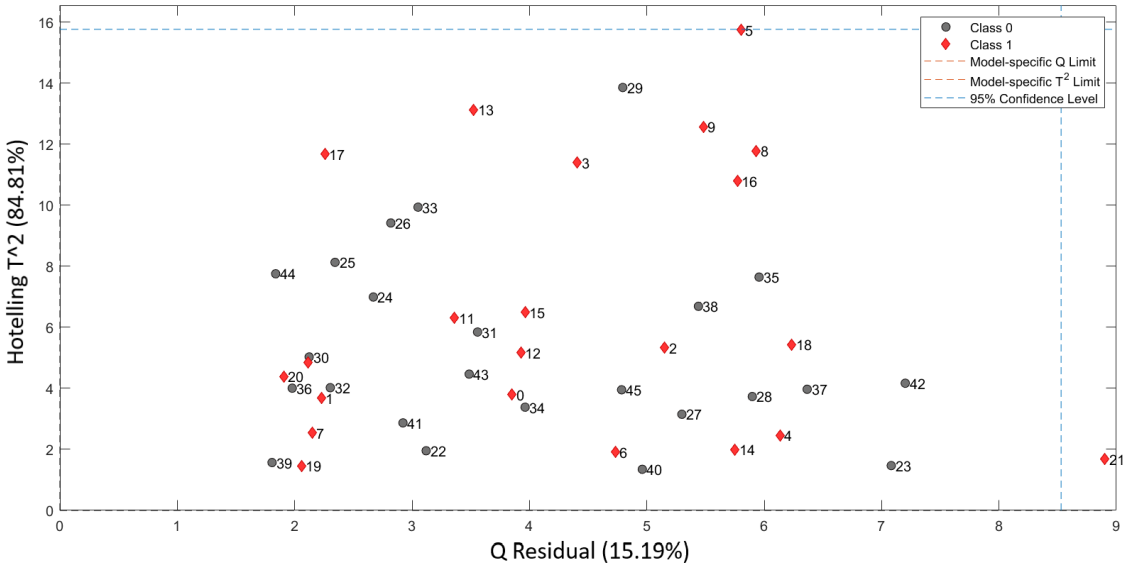


Figure E.22: The Q-T plot of the pallidum's shape feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

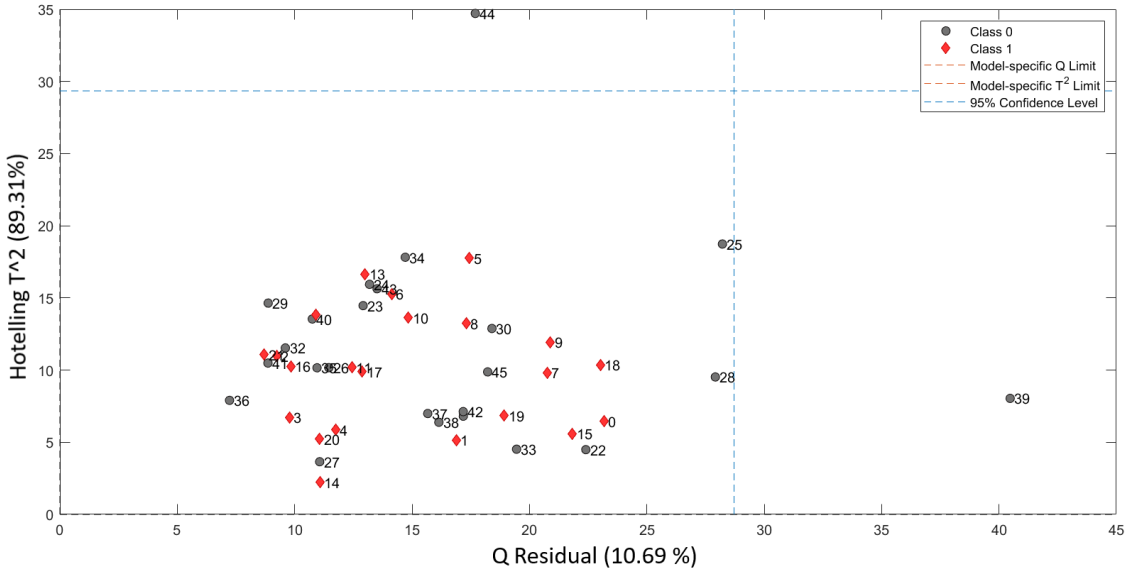


Figure E.23: The Q-T plot of the pallidum'd shape-bin texture feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

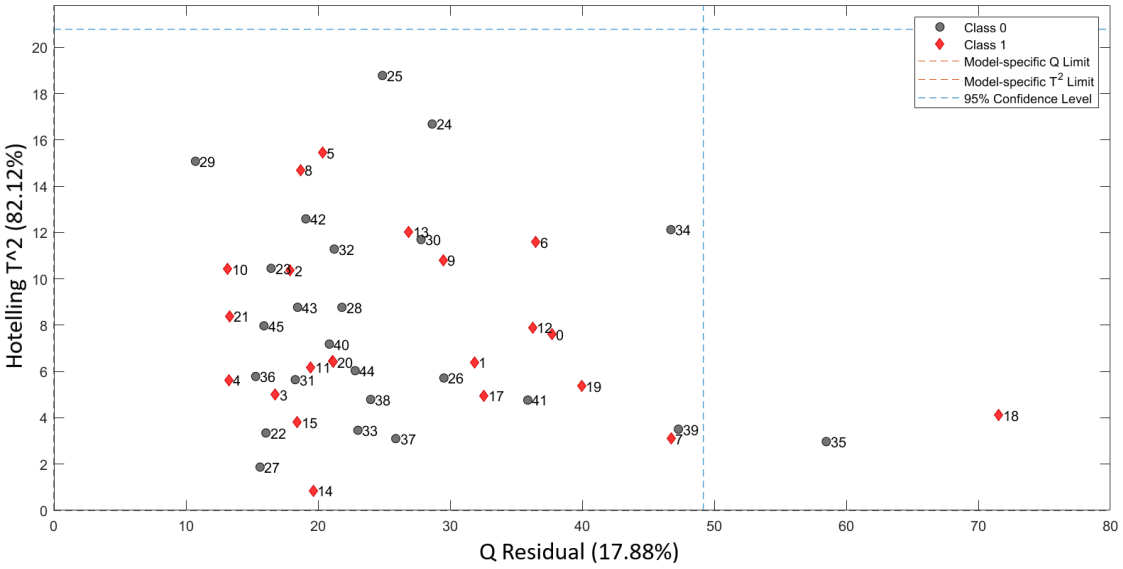


Figure E.24: The Q-T plot of the pallidum's 64-bin texture feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

E.3 The Putamen

Establishing the number of components to include in the putamen models based on the statistics or the RMSECV and RMSEC curves in Figure E.25 proved to be challenging. The PLS-Toolbox suggested less than three components for each of the models, covering under 40% of the cumulative explained variance according to Figures F.10, F.11, F.12. Seven principal components were chosen, covering approximately 80% of the cumulative variance of the models.

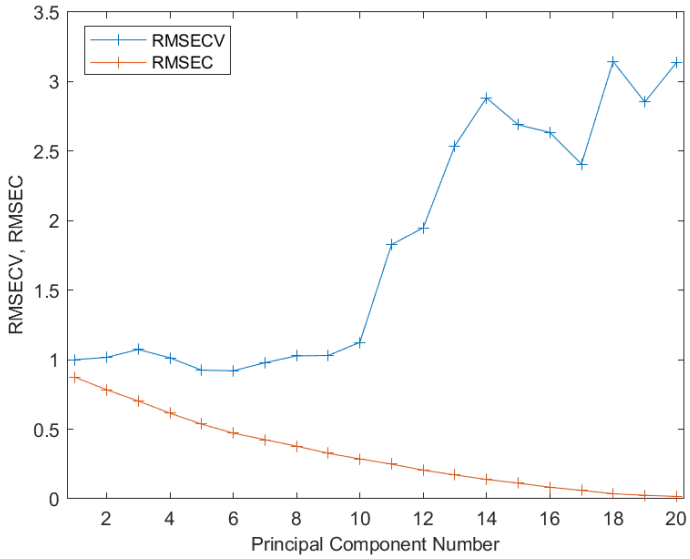


Figure E.25: The RMSEC and RMSECV curve of the putamen's shape feature model. While the RMSEC curve decreased constantly, the RMSECV increased after five components, indicating the model overfited stronlgy.

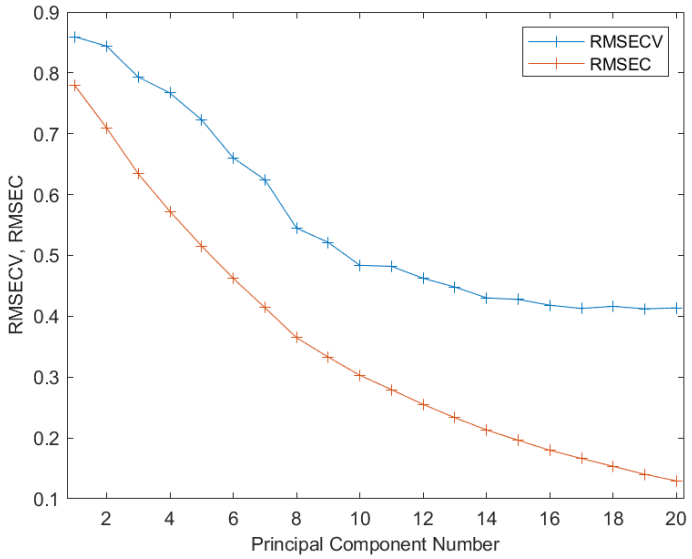


Figure E.26: The RMSEC and RMSECV curve of the putamen's 128-bin texture feature model. Both curves decreased constantly, until the RMSECV curve flattened out slightly.

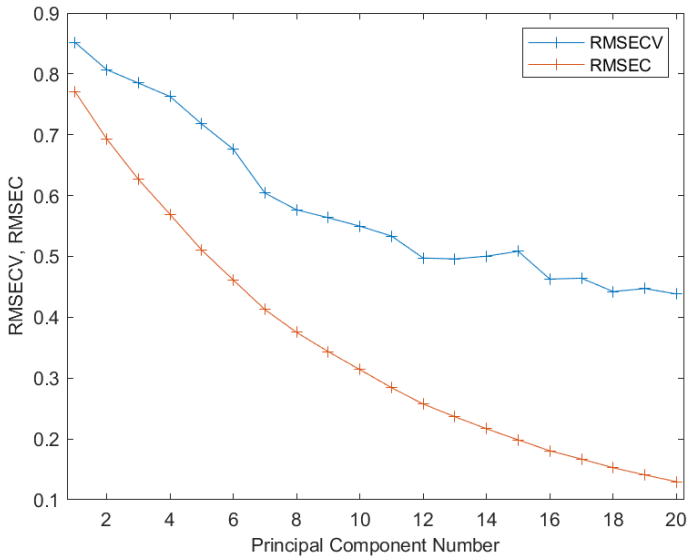


Figure E.27: The RMSEC and RMSECV curve of the putamen's 64-bin texture feature model. Both curves decreased constantly, until the RMSECV curve flattened out slightly.

E.3.1 Pattern Detection in the Score Plots

The score plots in Figures E.28-E.33 showed no signs of patterns or characteristics discriminating the treatment groups.

The Shape Feature-based PCA model

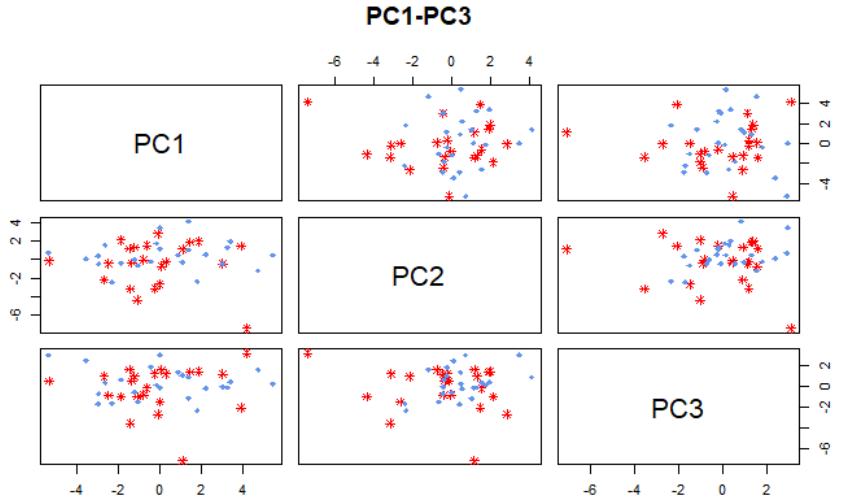


Figure E.28: The putamen's pair plot of the shape feature principal components nr 1-3. No principal components discriminating the treatment groups were detected.

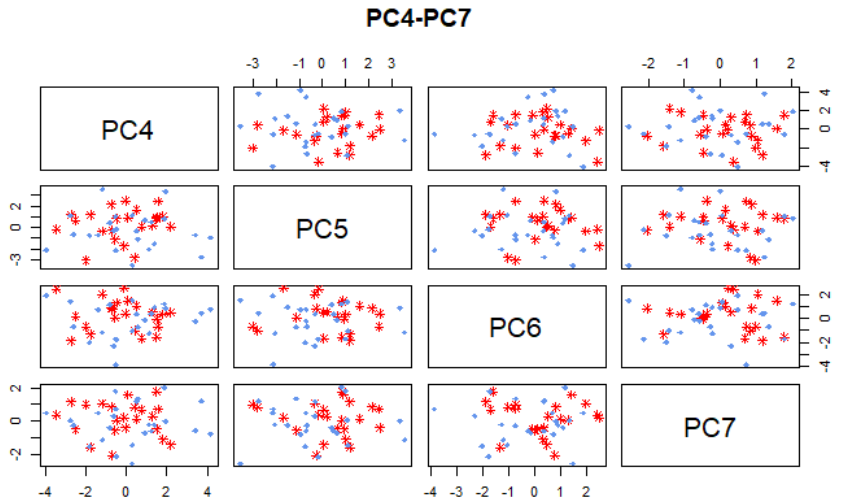


Figure E.29: The putamen's pair plot of the shape feature principal components nr 4-7. No principal components discriminating the treatment groups were detected.

The 128-bin Texture Feature-based PCA model

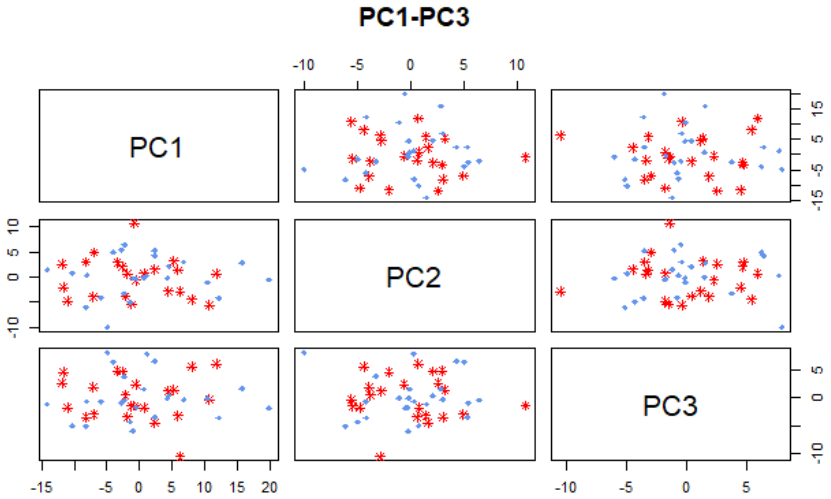


Figure E.30: The putamen's pair plot of the 128-bin texture feature principal components nr 1-3. No principal components discriminating the treatment groups were detected.

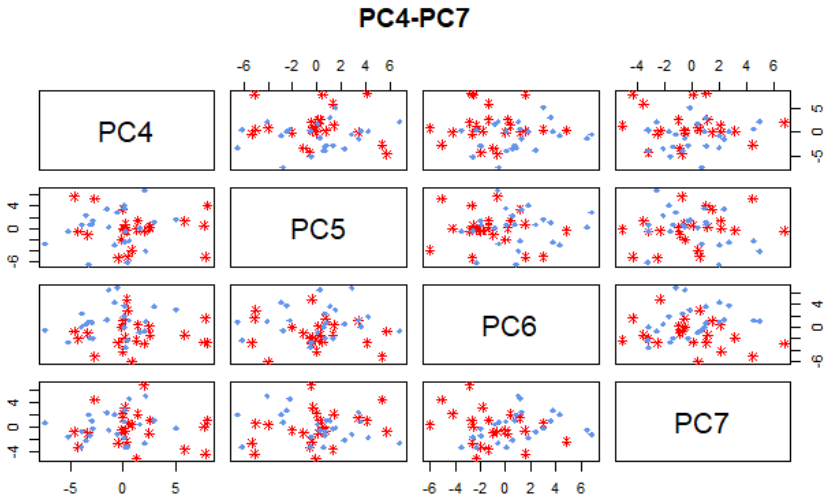


Figure E.31: The putamen's pair plot of the 128-bin texture feature principal components nr 4-7. No principal components discriminating the treatment groups were detected.

The 64-bin Texture Feature-based PCA model

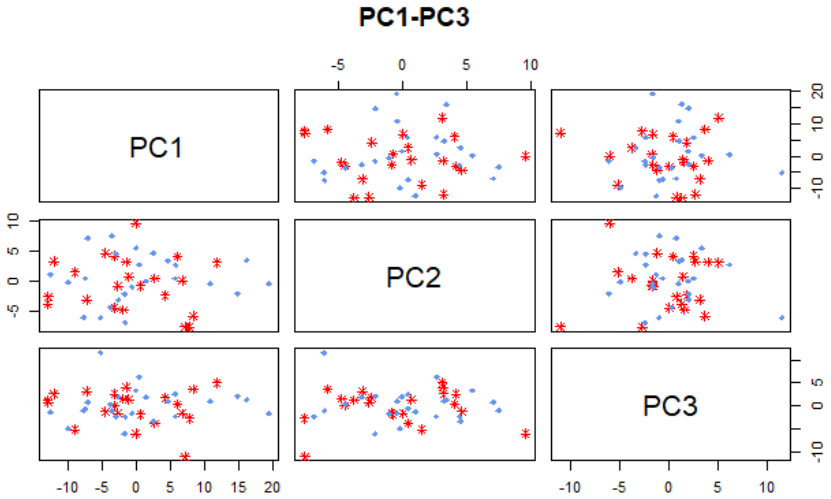


Figure E.32: The putamen's pair plot of the 64-bin texture feature principal components nr 1-3. No principal components discriminating the treatment groups were detected.

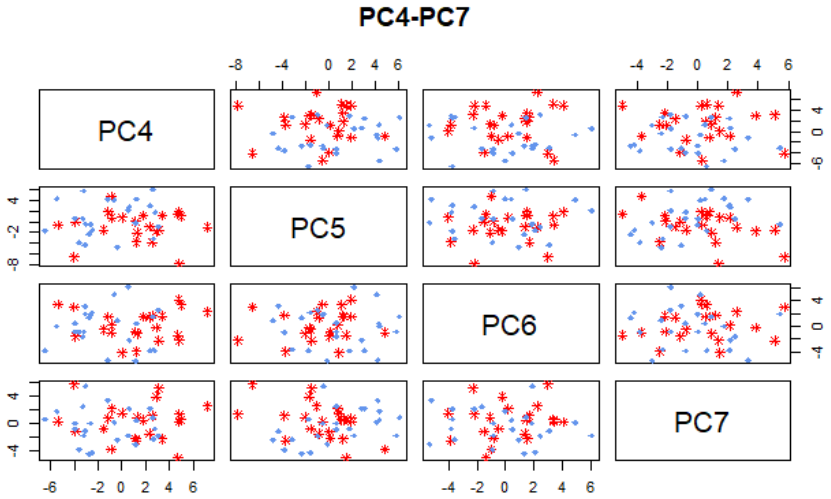


Figure E.33: The putamen's pair plot of the 64-bin texture feature principal components nr 4-7. No principal components discriminating the treatment groups were detected.

E.3.2 Identifying Possible Outliers in the Score and Q-T Plots

Based on the score plots, eight MPH-treated subjects and five placebo-treated subjects were identified as possible outliers in the putamen's shape feature model. Out of these, only subject nr. 13 was considered a possible outlier according to the Q - T plot in Figure E.34. Four and five participants were selected from the 64-bin texture feature and 128-bin texture feature model, respectively. Participant nr. 41 stood out in all three models but did not have a Q-residual or Hotelling T^2 value

above the 95% confidence limit in any of the corresponding Q - T plot in Figures E.34-E.36.

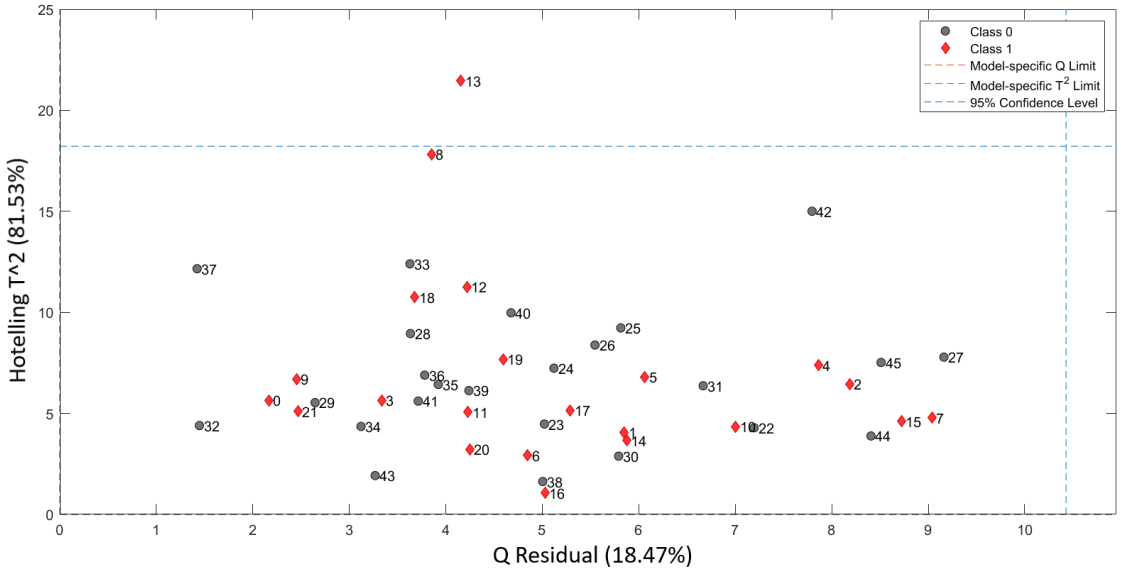


Figure E.34: The Q-T plot of the putamen's shape feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

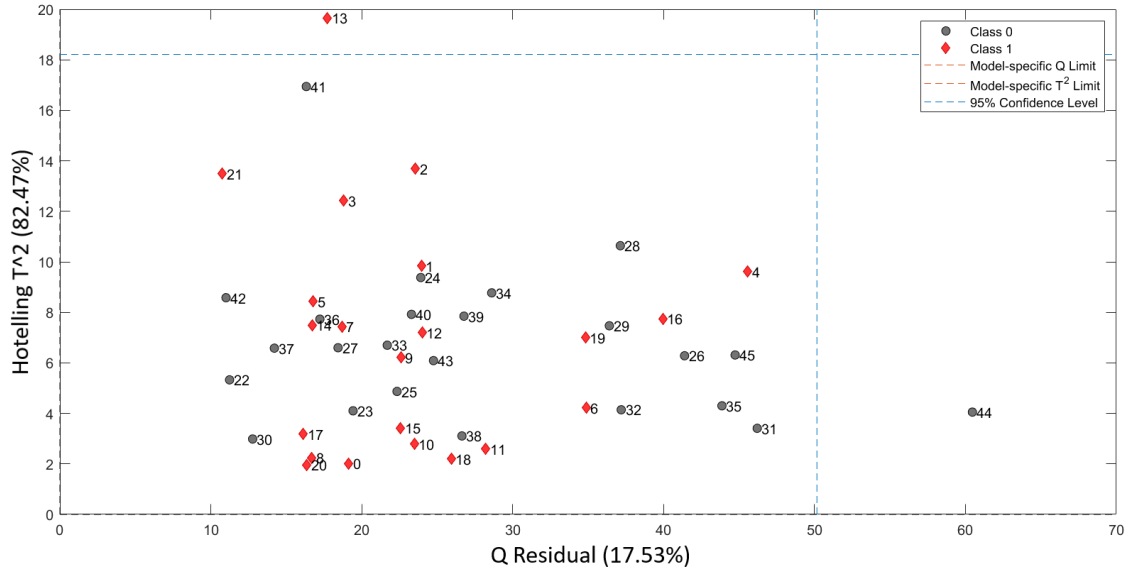


Figure E.35: The Q-T plot of the putamen's 128-bin texture feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

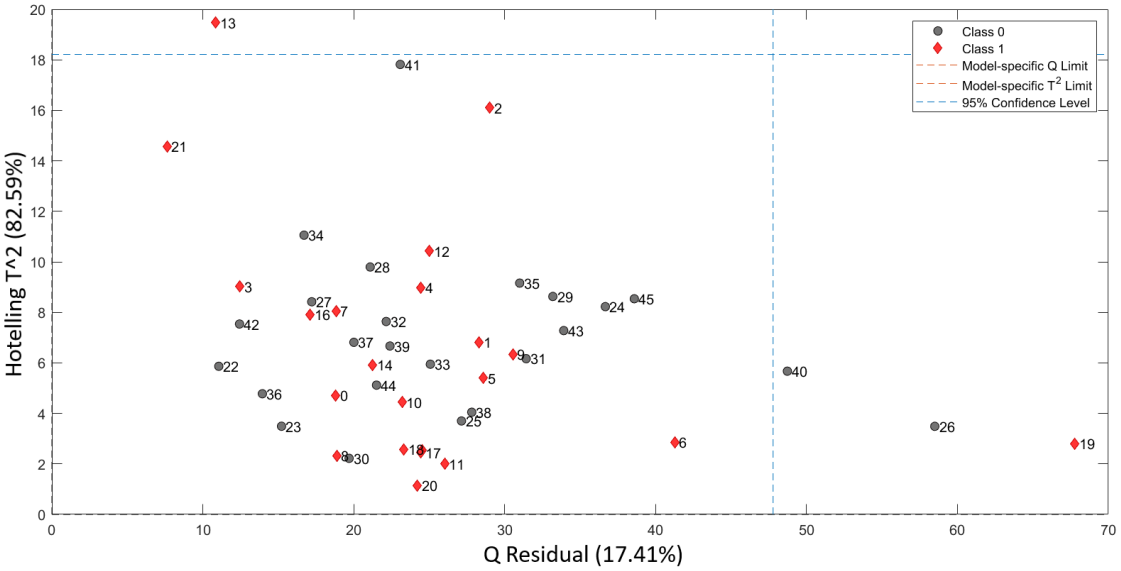


Figure E.36: The Q-T plot of the putamen's 64-bin texture feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

E.4 The Thalamus

Like the statistics of the putamen models, the statistics of the thalamus' shape feature model did not identify an explicit "cutoff"-component. The toolbox recommended using one component, covering 28.01% of the cumulative variance. Seven components were chosen, covering 80% according to Figure F.13. The RMSEC, as well as the RMSECV, of the model, decreased steeply, and flattened out upon reaching the third component, as illustrated in Figures E.38 and E.39. Due to this and the recommendation of PLS-Toolbox, the third component was chosen as a "cutoff" in the texture feature models.

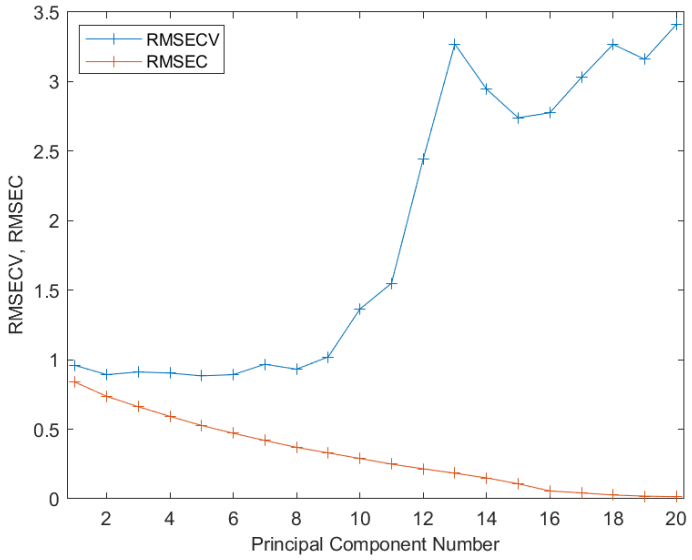


Figure E.37: The RMSEC and RMSECV curve of the thalamus' shape feature model. While the RMSEC curve decreased constantly, the RMSECV increased after six components, indicating the model overfited stronlgy.

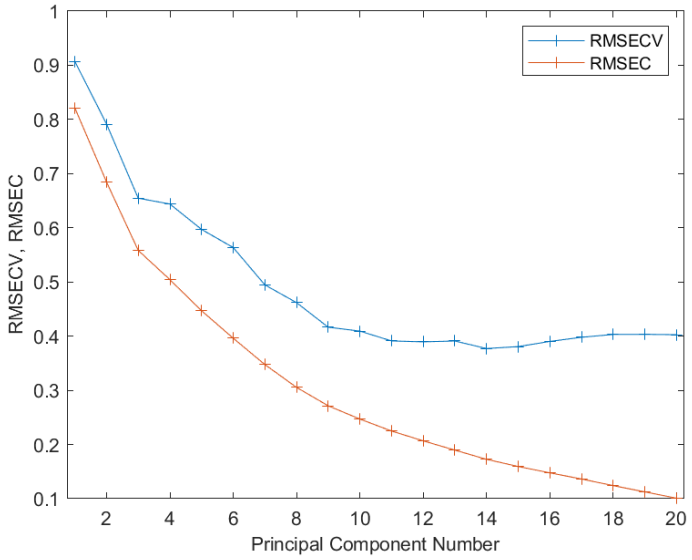


Figure E.38: The RMSEC and RMSECV curve of the thalamus' 128-bin texture feature model. Both curves decreased constantly, until the RMSECV curve flattened out slightly.

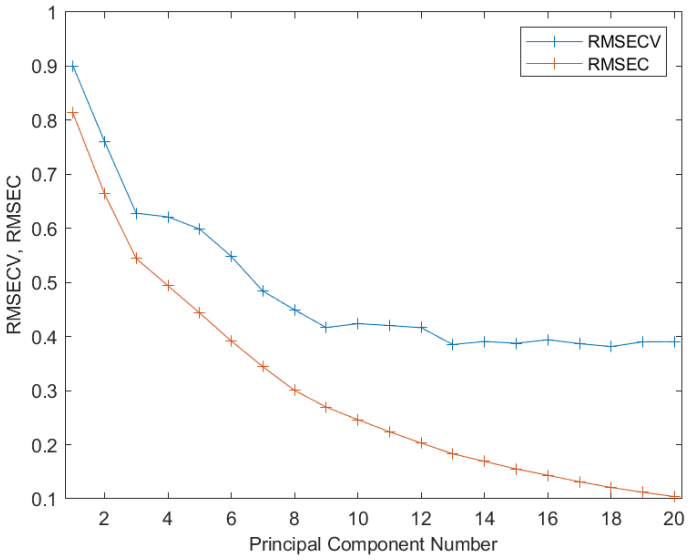


Figure E.39: The RMSEC and RMSECV curve of the thalamus' 64-bin texture feature model. Both curves decreased constantly, until the RMSECV curve flattened out slightly.

E.4.1 Pattern Detection in the Score Plots

Similar to the previous brain structures, no patterns distinguishing the treatment groups from each other were detectable through visual examinations of the score plots in Figures E.40 - E.43.

The Shape Feature-based PCA model

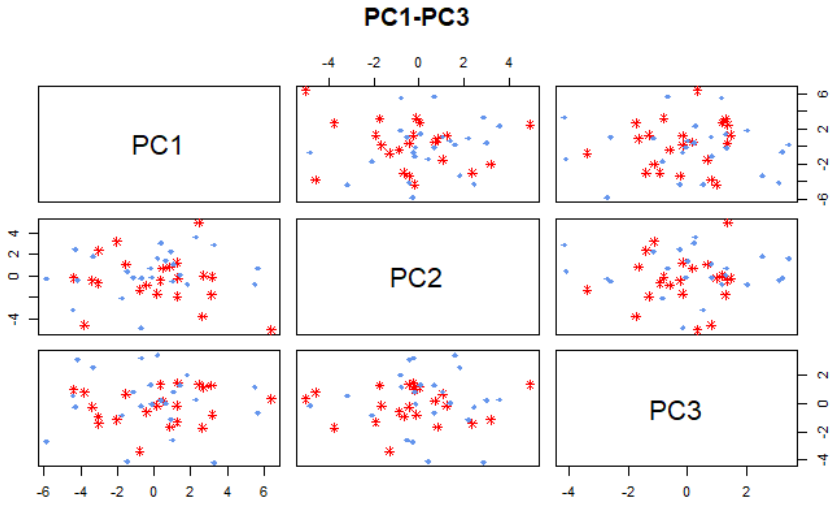


Figure E.40: The thalamus' pair plot of the shape feature principal components nr 1-3. No principal components discriminating the treatment groups were detected.

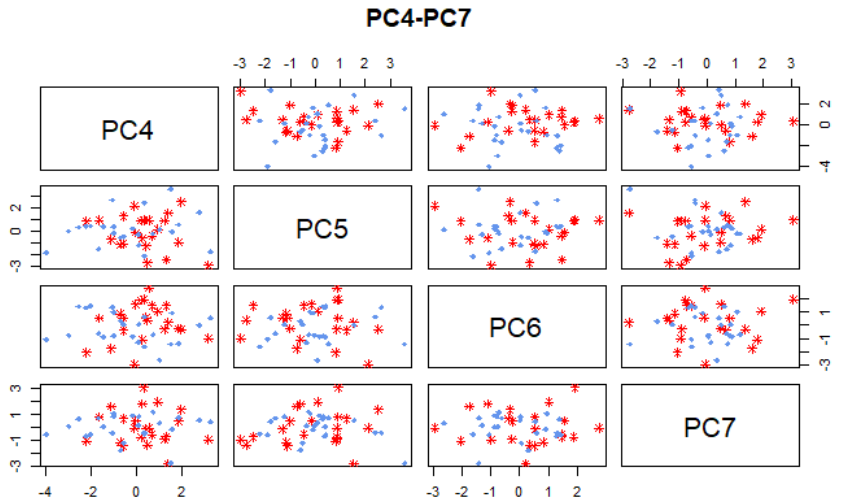


Figure E.41: The thalamus' pair plot of the shape feature principal components nr 4-7. No principal components discriminating the treatment groups were detected.

The 128-bin Texture Feature-based PCA model

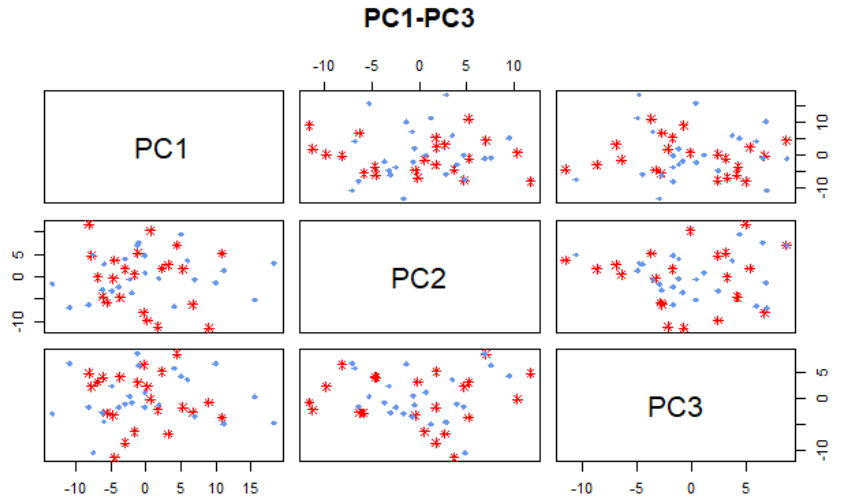


Figure E.42: The thalamus' pair plot of the 128-bin texture feature principal components nr 1-3. No principal components discriminating the treatment groups were detected.

The 64-bin Texture Feature-based PCA model

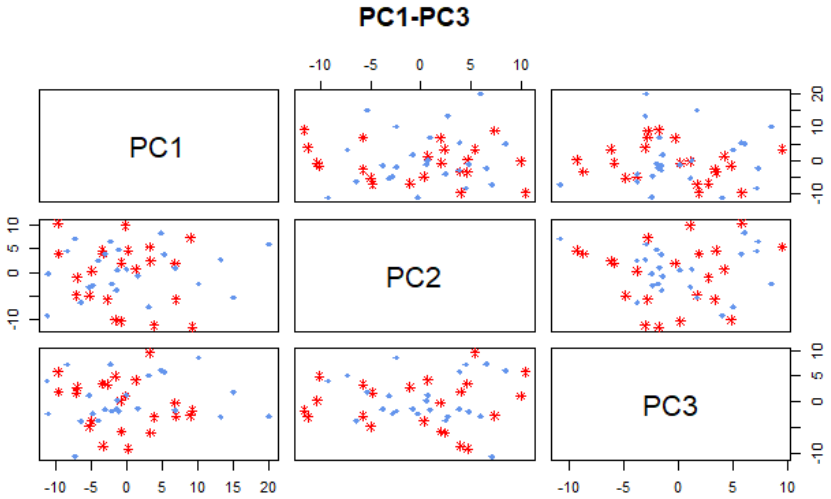


Figure E.43: The thalamus' pair plot of the 64-bin texture feature principal components nr 1-3. No principal components discriminating the treatment groups were detected.

E.4.2 Identifying Possible Outliers in the Score and Q-T Plots

No MPH-treated participants were categorised as noticeable observations or possible outliers in the texture feature models based in the same score plots and the corresponding Q - T plots in Figures E.45 and E.46. Two placebo-treated subjects were selected as possible outliers in the 64-bin texture feature model based on the models score plots. One of the subjects also stood out in the 128-bin texture feature model's score plots.

None of these subjects crossed the 95% confidence limit of the corresponding Q - T plots either. In the score plots of the shape feature model in Figure E.40 and E.41, eight MPH-treated and four placebo-treated subjects stood out. Similar to the texture feature models, none of the subjects were categorised as possible outliers according to the Q - T plot in Figure E.44.

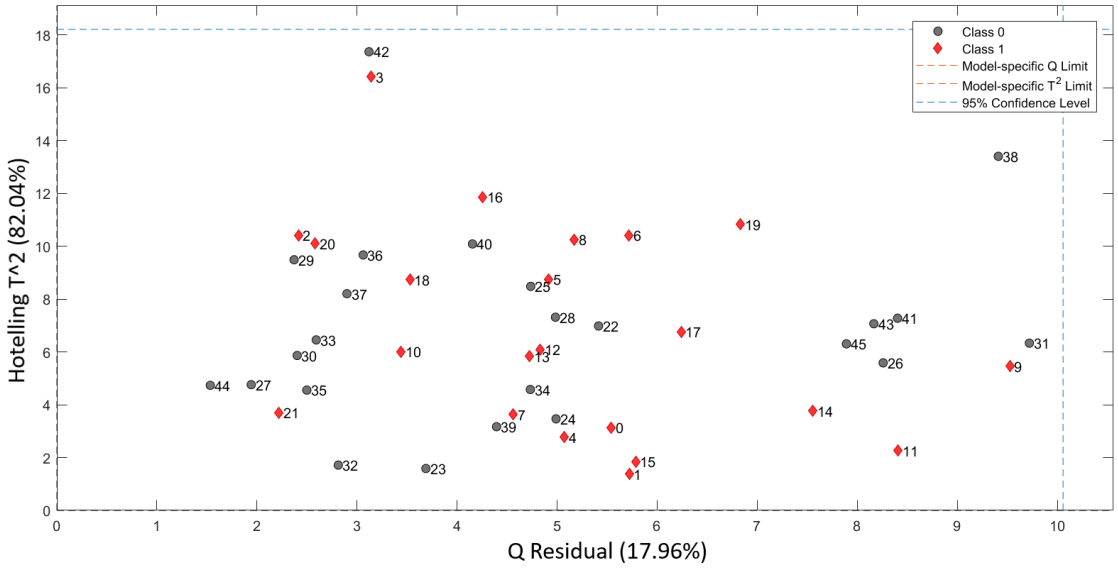


Figure E.44: The Q-T plot of the thalamus' shape feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

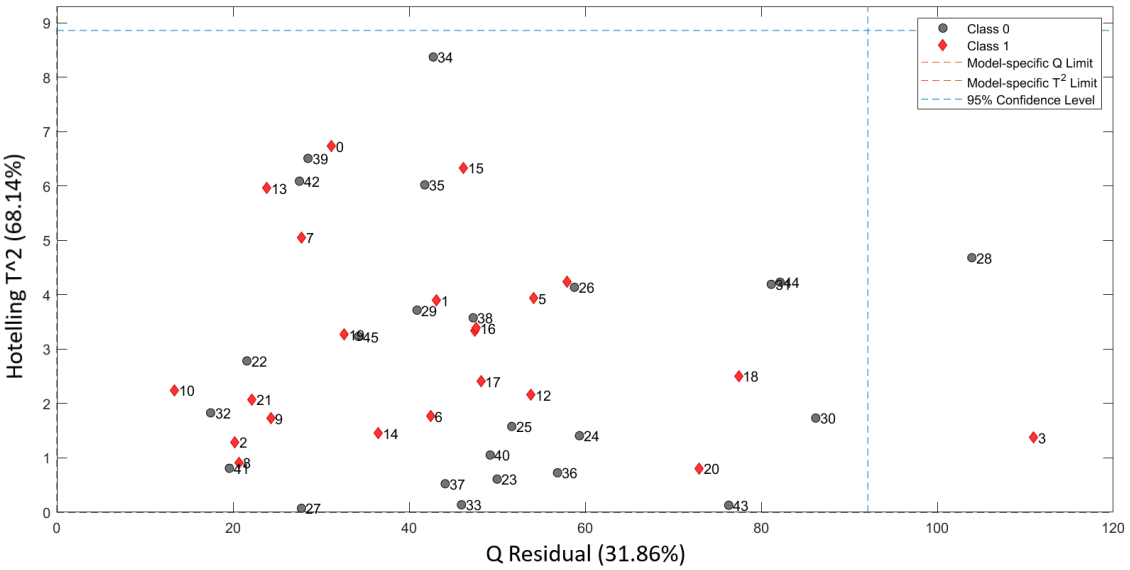


Figure E.45: The Q-T plot of the thalamus' 128-bin texture feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

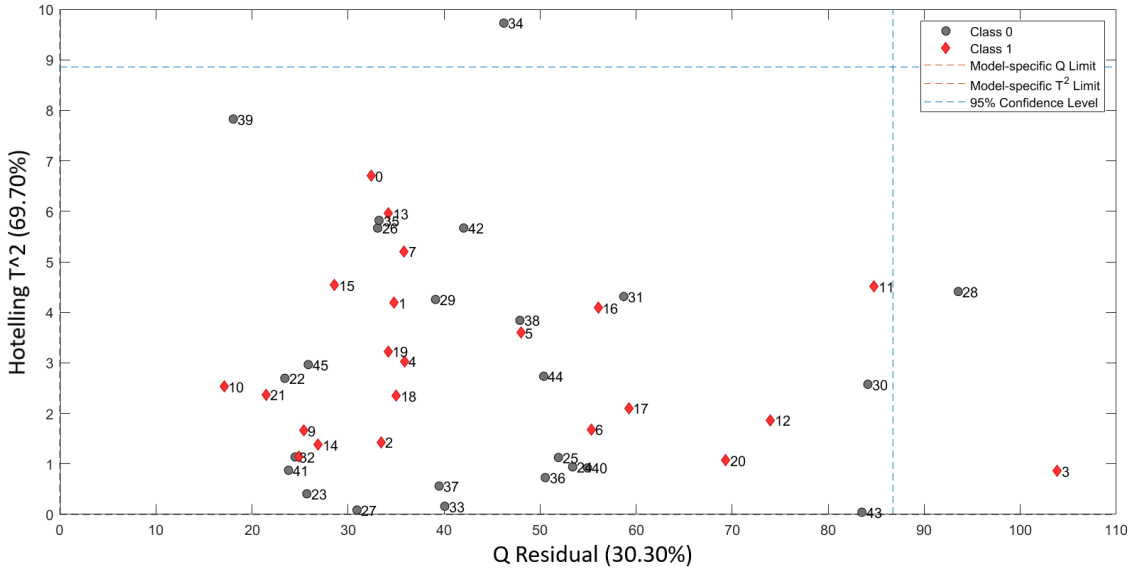


Figure E.46: The Q-T plot of the thalamus' 64-bin texture feature model.The blue dotted line represents the 95% confidence limit. MPH-treated (red subjects) and placebo-treated (grey subjects) with either q-residual or Hotelling's T^2 values exceeding the 95% confidence limit may qualify as outliers.

Appendix F

The PCA Model Statistics

The Caudate's PCA Model Statistics

In order to chose the number of principal components to include in the thesis' PCA models, model statistics generated by PLS-Toolbox [60] were examined. The model statistics included the eigenvalues, the individual explained variance and the cumulative explained variance, and are displayed in this Appendix.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	5.11e+00	18.25	18.25
2	4.41e+00	15.76	34.02
3	3.49e+00	12.47	46.48
4	2.65e+00	9.47	55.95
5	2.40e+00	8.56	64.52
6	1.97e+00	7.05	71.57
7	1.60e+00	5.70	77.27
8	1.46e+00	5.22	82.49
9	1.04e+00	3.72	86.20
10	8.20e-01	2.93	89.13
11	7.92e-01	2.83	91.96
12	6.10e-01	2.18	94.14
13	4.45e-01	1.59	95.73
14	3.65e-01	1.30	97.04
15	3.09e-01	1.11	98.14
16	2.06e-01	0.74	98.88
17	1.43e-01	0.51	99.39
18	9.57e-02	0.34	99.73
19	3.67e-02	0.13	99.86
20	2.06e-02	0.07	99.94

Figure F.1: The model statistics of the caudate's shape feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	5.84e+01	38.91	38.91
2	2.01e+01	13.37	52.28
3	1.22e+01	8.15	60.43
4	1.09e+01	7.24	67.68
5	9.63e+00	6.42	74.10
6	7.34e+00	4.89	78.99
7	6.39e+00	4.26	83.24
8	5.37e+00	3.58	86.83
9	3.61e+00	2.41	89.24
10	2.51e+00	1.67	90.91
11	2.42e+00	1.61	92.52
12	1.61e+00	1.07	93.60
13	1.43e+00	0.96	94.55
14	1.27e+00	0.84	95.40
15	1.07e+00	0.71	96.11
16	8.36e-01	0.56	96.67
17	7.22e-01	0.48	97.15
18	6.27e-01	0.42	97.57
19	5.24e-01	0.35	97.92
20	4.52e-01	0.30	98.22

Figure F.2: The model statistics of the caudate's 128-bin texture feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	5.80e+01	38.65	38.65
2	1.97e+01	13.15	51.80
3	1.20e+01	7.99	59.79
4	9.50e+00	6.33	66.13
5	9.29e+00	6.19	72.32
6	7.59e+00	5.06	77.38
7	6.60e+00	4.40	81.78
8	5.94e+00	3.96	85.73
9	4.58e+00	3.06	88.79
10	3.07e+00	2.04	90.83
11	2.43e+00	1.62	92.46
12	2.04e+00	1.36	93.81
13	1.64e+00	1.09	94.91
14	1.41e+00	0.94	95.85
15	9.78e-01	0.65	96.50
16	8.91e-01	0.59	97.09
17	6.38e-01	0.43	97.52
18	5.59e-01	0.37	97.89
19	4.91e-01	0.33	98.22
20	3.85e-01	0.26	98.47

Figure F.3: The model statistics of the caudate's 64-bin texture feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

The Hippocampus' PCA Model Statistics

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	5.72e+00	20.42	20.42
2	4.88e+00	17.43	37.85
3	3.26e+00	11.63	49.48
4	2.89e+00	10.30	59.78
5	1.97e+00	7.04	66.83
6	1.86e+00	6.65	73.48
7	1.65e+00	5.89	79.37
8	1.16e+00	4.16	83.52
9	9.86e-01	3.52	87.05
10	8.87e-01	3.17	90.21
11	6.88e-01	2.46	92.67
12	5.73e-01	2.05	94.72
13	3.57e-01	1.28	95.99
14	3.22e-01	1.15	97.14
15	2.71e-01	0.97	98.11
16	2.15e-01	0.77	98.88
17	1.13e-01	0.40	99.28
18	7.91e-02	0.28	99.56
19	6.95e-02	0.25	99.81
20	4.16e-02	0.15	99.96

Figure F.4: The model statistics of the hippocampus' shape feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	4.66e+01	31.06	31.06
2	2.16e+01	14.38	45.44
3	1.60e+01	10.67	56.11
4	1.24e+01	8.28	64.38
5	9.19e+00	6.13	70.51
6	8.19e+00	5.46	75.97
7	7.20e+00	4.80	80.77
8	5.16e+00	3.44	84.22
9	4.59e+00	3.06	87.28
10	3.93e+00	2.62	89.90
11	2.22e+00	1.48	91.38
12	2.02e+00	1.34	92.72
13	1.58e+00	1.05	93.78
14	1.48e+00	0.99	94.76
15	1.24e+00	0.82	95.59
16	1.09e+00	0.73	96.32
17	7.94e-01	0.53	96.85
18	7.02e-01	0.47	97.31
19	4.87e-01	0.32	97.64
20	4.63e-01	0.31	97.95

Figure F.5: The model statistics of the hippocampus' 128-bin texture feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	4.70e+01	31.33	31.33
2	2.30e+01	15.31	46.65
3	1.62e+01	10.78	57.43
4	1.27e+01	8.46	65.89
5	9.29e+00	6.19	72.08
6	7.27e+00	4.85	76.93
7	6.20e+00	4.13	81.06
8	4.96e+00	3.30	84.36
9	3.46e+00	2.31	86.67
10	3.19e+00	2.12	88.79
11	2.93e+00	1.95	90.75
12	2.42e+00	1.61	92.36
13	1.73e+00	1.15	93.51
14	1.64e+00	1.10	94.61
15	1.42e+00	0.94	95.55
16	1.05e+00	0.70	96.26
17	8.10e-01	0.54	96.80
18	6.49e-01	0.43	97.23
19	6.08e-01	0.41	97.63
20	5.47e-01	0.36	98.00

Figure F.6: The model statistics of the hippocampus' 64-bin texture feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

The Pallidum's PCA Model Statistics

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	8.04e+00	28.72	28.72
2	4.55e+00	16.26	44.99
3	3.94e+00	14.06	59.04
4	3.01e+00	10.74	69.78
5	2.52e+00	9.00	78.78
6	1.69e+00	6.03	84.81
7	9.31e-01	3.32	88.13
8	7.37e-01	2.63	90.76
9	5.90e-01	2.11	92.87
10	5.22e-01	1.86	94.74
11	3.84e-01	1.37	96.11
12	3.31e-01	1.18	97.29
13	3.00e-01	1.07	98.36
14	2.12e-01	0.76	99.12
15	9.78e-02	0.35	99.46
16	4.33e-02	0.15	99.62
17	3.28e-02	0.12	99.74
18	2.87e-02	0.10	99.84
19	1.63e-02	0.06	99.90
20	1.25e-02	0.04	99.94

Figure F.7: The model statistics of the palldum's shape feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	3.27e+01	21.83	21.83
2	2.18e+01	14.54	36.37
3	1.64e+01	10.91	47.28
4	1.28e+01	8.52	55.80
5	1.19e+01	7.96	63.75
6	9.37e+00	6.25	70.00
7	8.00e+00	5.33	75.34
8	7.29e+00	4.86	80.20
9	5.42e+00	3.61	83.81
10	4.65e+00	3.10	86.91
11	3.59e+00	2.40	89.31
12	2.63e+00	1.76	91.06
13	2.02e+00	1.35	92.41
14	1.81e+00	1.20	93.62
15	1.36e+00	0.91	94.52
16	1.09e+00	0.73	95.25
17	1.06e+00	0.70	95.95
18	1.01e+00	0.67	96.62
19	7.32e-01	0.49	97.11
20	6.72e-01	0.45	97.56

Figure F.8: The model statistics of the pallidumn's 128-bin texture feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	3.36e+01	22.39	22.39
2	2.90e+01	19.32	41.70
3	1.47e+01	9.78	51.48
4	1.25e+01	8.33	59.81
5	1.14e+01	7.58	67.39
6	9.03e+00	6.02	73.41
7	7.45e+00	4.97	78.38
8	5.61e+00	3.74	82.12
9	4.55e+00	3.03	85.15
10	4.32e+00	2.88	88.03
11	2.79e+00	1.86	89.89
12	2.36e+00	1.57	91.46
13	1.97e+00	1.31	92.77
14	1.55e+00	1.03	93.80
15	1.33e+00	0.89	94.69
16	1.08e+00	0.72	95.41
17	9.02e-01	0.60	96.01
18	8.24e-01	0.55	96.56
19	7.77e-01	0.52	97.08
20	7.15e-01	0.48	97.56

Figure F.9: The model statistics of the pallidum's 64-bin texture feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

The Putamen's PCA Model Statistics

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	6.13e+00	21.90	21.90
2	4.27e+00	15.24	37.14
3	3.44e+00	12.27	49.41
4	3.31e+00	11.81	61.23
5	2.61e+00	9.32	70.55
6	1.87e+00	6.67	77.22
7	1.21e+00	4.32	81.53
8	1.07e+00	3.84	85.37
9	1.03e+00	3.68	89.05
10	7.11e-01	2.54	91.59
11	5.76e-01	2.06	93.65
12	5.64e-01	2.01	95.67
13	3.69e-01	1.32	96.98
14	2.93e-01	1.05	98.03
15	1.89e-01	0.67	98.70
16	1.65e-01	0.59	99.29
17	9.33e-02	0.33	99.63
18	6.69e-02	0.24	99.87
19	1.93e-02	0.07	99.93
20	1.02e-02	0.04	99.97

Figure F.10: The model statistics of the putamen's shape feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	5.69e+01	37.95	37.95
2	1.58e+01	10.56	48.51
3	1.55e+01	10.36	58.88
4	1.14e+01	7.62	66.49
5	9.49e+00	6.33	72.82
6	7.98e+00	5.32	78.14
7	6.49e+00	4.33	82.47
8	5.88e+00	3.92	86.39
9	3.42e+00	2.28	88.67
10	2.92e+00	1.94	90.62
11	2.11e+00	1.41	92.02
12	1.99e+00	1.33	93.35
13	1.63e+00	1.09	94.44
14	1.37e+00	0.92	95.35
15	1.07e+00	0.71	96.07
16	9.48e-01	0.63	96.70
17	7.17e-01	0.48	97.18
18	6.28e-01	0.42	97.60
19	5.90e-01	0.39	97.99
20	4.54e-01	0.30	98.29

Figure F.11: The model statistics of the putamen's 128-bin texture feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	5.88e+01	39.18	39.18
2	1.75e+01	11.66	50.84
3	1.35e+01	8.97	59.81
4	1.06e+01	7.06	66.86
5	9.71e+00	6.47	73.34
6	7.35e+00	4.90	78.24
7	6.53e+00	4.36	82.59
8	4.52e+00	3.02	85.61
9	3.49e+00	2.33	87.94
10	2.98e+00	1.99	89.92
11	2.72e+00	1.81	91.74
12	2.23e+00	1.49	93.23
13	1.58e+00	1.06	94.28
14	1.37e+00	0.91	95.20
15	1.16e+00	0.78	95.97
16	1.05e+00	0.70	96.67
17	7.22e-01	0.48	97.15
18	6.98e-01	0.47	97.62
19	5.32e-01	0.35	97.97
20	4.60e-01	0.31	98.28

Figure F.12: The model statistics of the putamen's 64-bin texture feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

The Thalamus' PCA Model Statistics

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	7.84e+00	28.01	28.01
2	4.63e+00	16.53	44.54
3	2.98e+00	10.65	55.19
4	2.49e+00	8.91	64.09
5	2.10e+00	7.49	71.58
6	1.60e+00	5.72	77.30
7	1.33e+00	4.74	82.04
8	1.12e+00	3.99	86.02
9	8.02e-01	2.86	88.88
10	7.06e-01	2.52	91.41
11	6.30e-01	2.25	93.66
12	4.46e-01	1.59	95.25
13	3.63e-01	1.30	96.55
14	3.20e-01	1.14	97.69
15	3.13e-01	1.12	98.81
16	2.42e-01	0.87	99.67
17	4.03e-02	0.14	99.82
18	2.99e-02	0.11	99.93
19	1.05e-02	0.04	99.96
20	4.13e-03	0.01	99.98

Figure F.13: The model statistics of the thalamus' shape feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	4.66e+01	31.07	31.07
2	3.16e+01	21.06	52.13
3	2.40e+01	16.01	68.14
4	8.79e+00	5.86	74.00
5	8.36e+00	5.58	79.58
6	6.57e+00	4.38	83.96
7	5.54e+00	3.69	87.65
8	4.22e+00	2.81	90.46
9	3.01e+00	2.00	92.47
10	1.95e+00	1.30	93.77
11	1.60e+00	1.06	94.83
12	1.20e+00	0.80	95.63
13	1.04e+00	0.69	96.32
14	9.43e-01	0.63	96.95
15	6.74e-01	0.45	97.40
16	5.50e-01	0.37	97.77
17	4.99e-01	0.33	98.10
18	4.83e-01	0.32	98.42
19	4.24e-01	0.28	98.71
20	3.80e-01	0.25	98.96

Figure F.14: The model statistics of the thalamus' 128-bin texture feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

	Eigenvalue of Cov(X)	% Variance This PC	% Variance Cumulative
1	4.85e+01	32.34	32.34
2	3.40e+01	22.64	54.98
3	2.21e+01	14.72	69.70
4	8.02e+00	5.35	75.05
5	7.23e+00	4.82	79.87
6	6.63e+00	4.42	84.30
7	5.38e+00	3.59	87.89
8	4.33e+00	2.89	90.77
9	2.70e+00	1.80	92.57
10	1.84e+00	1.23	93.80
11	1.62e+00	1.08	94.88
12	1.37e+00	0.91	95.79
13	1.18e+00	0.79	96.57
14	7.53e-01	0.50	97.08
15	6.95e-01	0.46	97.54
16	5.36e-01	0.36	97.90
17	5.06e-01	0.34	98.23
18	4.08e-01	0.27	98.51
19	3.17e-01	0.21	98.72
20	2.64e-01	0.18	98.89

Figure F.15: The model statistics of the thalamus' 64-bin texture feature-based PCA model for the twenty first principal components, including the eigenvalues, the individual explained variance and the cumulative variance.

Appendix G

Number of Chosen PCs

When creating the PCA models in this thesis, the number of principal components analysed was limited in order to prevent inclusion of noise and unsystematic variance in the models. PLS-Toolbox [60] recommended the number of PCs to include in the models, but the models statistics were also considered when deciding the number of PCs to include further. Both the suggested and the chosen number of principal components are listed in Table G.1.

Table G.1: The number of PCs recommended by PLS-Toolbox included in the PCA models, and the number of PCs chosen included for further analyses.

Feature Matrix	PCs suggested by PLS toolbox	PCs decided upon
Caudate: Shape Features	1	8
Caudate: 128-bin Texture Features	9	9
Caudate: 64-bin Texture Features	2	7
Hippocampus: Shape Features	1	7
Hippocampus: 128-bin Texture Features	7	7
Hippocampus: 64-bin Texture Features	5	5
Pallidum: Shape Features	1	6
Pallidum: 128-bin Texture Features	11	11
Pallidum: 64-bin Texture Features	8	8
Putamen: Shape Features	1	7
Putamen: 128-bin Texture Features	1	7
Putamen: 64-bin Texture Features	2	7
Thalamus: Shape Features	1	7
Thalamus: 128-bin Texture Features	3	3
Thalamus: 64-bin Texture Features	3	3

Appendix H

Subjects Standing Out in Score Plots

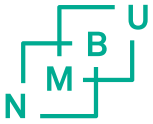
Subjects standing out in score plots generated as part of the principal component analyses presented in Chapter 4.3.1 and Appendix E were listed in Table H.1.

Table H.1: Subjects which stood out in the PCA models' generated score plots in Chapter 4.3.1 and Appendix E.

Model	IDs of class 1			IDs of class 0	
	1 PC	2 PCs	3 or 4 PCs	1 PC	2 PCs
Caudate, Shape	2, 3, 8, 14, 21	4, 13		25, 26, 27, 31, 34, 35, 42, 44	43
Caudate, 128-bin tex.	0, 4, 6, 13, 17, 18, 21	5, 16	2	27, 28, 31, 33, 40, 43	
Caudate, 64-bin tex.	0, 3, 4, 6, 7, 13, 18	2, 5, 16		26, 30, 31, 35, 37, 43	
Hippocampus, Shape	3, 4, 5, 11, 12, 17, 19	7		24, 25, 29, 34, 39	
Hippocampus, 128-bin tex.	3, 6, 7, 11, 13, 15, 20	5, 18	4	22, 25, 27, 29, 35, 38, 42	
Hippocampus, 64-bin tex.	2, 3, 7	13	4	31, 35, 40, 44	29, 41
Pallidum,					

Continuation of Table H.1

Model	IDs of class 1			IDs of class 0	
	1 PC	2 PCs	3 or 4 PCs	1 PC	2 PCs
Shape	2, 3, 8, 9, 16	5, 13		29, 33, 44	
Pallidum, 128-bin tex.	5, 6, 16, 18			29, 32, 45	44
Pallidum, 64-bin tex.	2, 4, 6, 8, 9, 10, 13, 17, 18	5		24, 25, 28, 29, 30, 32, 42, 43	
Putamen, Shape	2, 5, 8, 9, 12, 13, 17, 18			25, 37, 40, 41, 42	
Putamen, 128-bin tex.	3, 4	13, 21		41	
Putamen, 64-bin tex.	2, 12, 13			41	
Thalamus, Shape	2, 8, 10, 16, 17, 19	3, 20		25, 29, 33, 42	
Thalamus, 128-bin tex.					39
Thalamus, 64-bin tex.					34, 39



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitskapelege universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway